Diana Maria Patacas Viegas Vieira

MSc. Biotecnologia

Crystallographic studies of proteins involved in the mRNA localization mechanisms in *Drosophila melanogaster* and amidation of the peptidoglycan residues in *Staphylococcus aureus*

Dissertação para obtenção do Grau de Doutor em
Bioquímica, Especialidade Bioquímica Estrutural

Orientador: Doutor José Trincão
Senior Beamline Scientist, Diamond Light Source, UK
Co-orientador: Doutora Maria João Romão
Professora Catedrática, FCT/UNL
Doutor Thomas Edwards
Associate Professor, Leeds University, UK

Júri:

Presidente: Doutor José Paulo Barbosa Mota
Arguentes: Doutor Sérgio Raposo Filipe
Doutora Sandra de Macedo Ribeiro
Vogais: Doutora Maria Luísa Santos de Sousa Cyrne
Doutora Teresa Sacadura Santos Silva

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

16 de Dezembro de 2014

Diana Maria Patacas Viegas Vieira

Crystallographic studies of proteins involved in the mRNA localization mechanisms in *Drosophila melanogaster* and amidation of the peptidoglycan residues in *Staphylococcus aureus*

Dissertação apresentada para a obtenção do grau de Doutor em Bioquímica, Especialidade Bioquímica Estrutural, pela Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Monte da Caparica, 16 de Dezembro de 2014

Título: “Crystallographic studies of proteins involved in the mRNA localization mechanisms in *Drosophila melanogaster* and amidation of the peptidoglycan residues in *Staphylococcus aureus*”

O trabalho apresentado nesta Dissertação foi realizado com o apoio financeiro da Fundação para a Ciência e Tecnologia (FCT–MCTES), através da atribuição da Bolsa de Doutoramento individual SFRH/BD/62415/2009, e do projecto PTDC/BIA-PRO/80486/2006.

Agradecimentos
(Acknownledgements)

Agradeço ao meu orientador Dr. José Trincão por ter-me entrevistado em 2009 e ter confiado em mim desde então para levar a bom porto os seus projectos científicos e por te-los tornado meus, por ter-me dado liberdade para encontrar a minha personalidade no mundo da ciência, sem imposições, fazendo com que crescesse e aprendesse com os próprios erros.

À minha co-orientadora, Professora Maria João Romão
por ter-me aceite na sua família científica e
por ter-me dado todas as condições para concretizar os meus objectivos.

To my co-supervisor, Dr Thomas Edwards thank you for being present
even with the geographical distance and
thank you for helping me to overcome myself to always do better and know better
and specially for instigating my scientific curiosity.

Aos meus colegas no grupo XTAL.
Foi um prazer ter partilhado o 605 e o 621 convosco!
Um obrigado especial à Cecília, à Benedita, à Márcia e ao Filipe
por toda a ajuda ao longo destes anos de convivência.
Não esquecendo o David pela paciência nos jogos de ténis e,
the last but not the least, o inigualável Sr. Marino
pela sua imensa boa-disposição e por ser um exemplo de coerência e, já sabes, stand-up comedy!..

To Professor Paul Raithby for welcoming me in his workgroup and for all the support;
without it would have been impossible to fulfil my objectives.
My utmost sincere acknowledgments!

To OPPF and its entire staff: I had the time of my life working with you.

Às pessoas que conheci durante a melhor/pior experiência … ser emigrante!
Fizeram de 18 meses da minha existência algo leve de se suportar.
Obrigada por terem partilhado comigo o sorriso, o olhar, palavras e atitudes…

A todos os meus amigos… vocês são e serão o pilar da minha *in*sanidade mental.

Às pessoas que conheci pelo caminho, às que ficaram e às que se perderam…

Àquela pessoa brilhante que desde os meus 5 anos sempre
esteve ao meu lado - a minha melhor amiga Sandra.

Aos meus afilhados Martim e Beatriz por completarem uma parte da minha vida.

Aos *meus* miúdos pela fonte inesgotável de energia.
Pelas gargalhadas e por tudo o que as crianças nos ensinam,
desde que estejamos dispostos a aprender.

Ao meu Myke, o meu *baby*. O meu pequeno pedacinho de amor. A minha riqueza.

À fundação "Vieira" por todo o apoio logístico.
E nada disto, nem qualquer outra etapa da minha vida, teria sido possível sem a minha família. Somos apenas quatro mas para mim somos o mundo.

À minha irmã, por ser quem é e, por fazer-me sentir mais do que realmente sou.

Ao meu Pai por ter-me dado um porto seguro onde sempre pude retornar,
por nunca me limitar, por ter-me ensinado a integridade,
que a nossa palavra é o que nos define e que
"o capitão é o último a abandonar o navio".

E, um MUITO OBRIGADA à minha Mãe.
A pessoa mais importante!
Um exemplo de Mulher.
Uma Força da Natureza e uma Guerreira.
Se sou quem sou, a ti o devo!

Diana Vieira

Index

Agradecimentos vii
Index ix
Figures Index xi
Table Index xvii
Equations Index xix
Abbreviations xxi
Abstract 25
I. mRNA localization mechanisms 25
II. Amidation of *S. aureus* peptidoglycan residues 26
I. Chapter I 29
X-Ray Protein Crystallography 29
Overview of Protein Crystallography 30
Crystal Production 31
Data Collection and X-Ray Diffraction 33
Synchrotron radiation 33
Electron Density and Structure Determination 37
The Phase Problem 37
Refinement, Validation and Publication 38
II. Chapter II 41
II.1. mRNA localization mechanisms in *Drosophila melanogaster* 41
Drosophila melanogaster 41
mRNA localization mechanisms 43
II.2. mRNA localization mechanisms in *Drosophila melanogaster:* experimental procedure 49
Ypsilon-Schachtel 50
Exuperantia 54
Exu-Yps complex 58
II.3. mRNA localization mechanisms in *Drosophila melanogaster:* results and discussion ... 59
II.4. mRNA localization mechanisms in *Drosophila melanogaster*: conclusions and future work 65
III. Chapter III 69
III.1. Amidation of *S. aureus* peptidoglycan residues: general introduction 69
Bacterial Cell Wall 69
The Biochemistry of Peptidoglycan 70
Peptidoglycan amidation - genetic determinants and enzymes 74
III.2. Amidation of *S. aureus* peptidoglycan residues: experimental procedure 77
GatD 78
Source organism 78
MurT 84

MurT-GatD complex 86
Data collection and structure determination 91
III.3. Amidation of *S. aureus* peptidoglycan residues: results and discussion 95
Overall Structure Description 96
Comparison with homologous structures 99
Ligand binding prediction 101
Proposed mechanism for GatD-MurT 104
III.4. Amidation of *S. aureus* peptidoglycan residues: conclusion and future work 105
References 107
Appendix 115
A. DNA and protein sequences 115
A.1. Translation of exuperantia (*exu*) 115
A.2. Translation of ypsilon-schachtel (*yps*) 116
A.3. Translation of gatD 117
A.4. Translation of murT 118
B. Exuperantia and Ypsilon-Schachtel cloning and expression assays 119
C. OPPF construct design – Exuperantia and Ypsilon-Schachtel 133
D. OPPF construct design – GatD and MurT 189
E. Folding of proteins expressed as inclusion bodies 199
Protocol 1 199
Protocol 2 200
F. Thermofluor Screening Solutions 201

Figures Index

FIGURE I.1 - X-ray protein crystallography process overview. The description is detailed on the text above. 30

FIGURE I.2 – Vapour diffusion techniques: Hanging drop and sitting drop. Both methods require a protein solution drop with buffer and precipitant to equilibrate with a reservoir solution containing buffers and precipitants at higher concentrations. 31

FIGURE I.3 - Crystallization phase diagram. Schematic representation of a phase diagram illustrating the variation of protein concentration with the precipitating concentration. 32

FIGURE I.4 - Schematic representation of an X-ray experiment. The X-rays are generated by a source. The crystal mounted on the goniometer and when it's hit by the X-rays, diffracts them. The diffracted beam is then recorded on a detector. 33

FIGURE I.5 – On the left: Diamond Light Source is the UK national synchrotron. On the right: Diamond Macromolecular Crystallography (MX) village overview. MX village currently has five beamlines: three high brilliance MAD beamlines (I02, I03, I04), a fixed wavelength beamline (I04-1) and a microfocus MAD beamline (I24). Image from *http://www.diamond.ac.uk/Beamlines/Mx.html.* 33

FIGURE I.6 – Schematic representation of the Bragg's Law. In this model, each set of parallel planes is treated as an independent diffractor and produces a single reflection – constructive interference. Adapted from [7] 35

FIGURE I.7 – Elementary representation of an X-ray experiment. A crystal when submitted to an X-ray beam produces reflections which are recorded in the detector creating a diffraction pattern. The diffraction pattern is the result of a Fourier transformation of the crystal real space into reciprocal space and vice-versa. 36

FIGURE I.8 – Structure factor equation and schematic representation (F). F_{hkl} is a mathematical relation of the scattering factor and the positional terms of each atom; is represented as a vector which length is the intensity of the reflection ($F \propto I$) and α is the phase angle. Adapted from [7] 38

FIGURE II.1 – Representation of the axis that define *Drosophila* body pattern: Anterior-Posterior (A→P) and dorsal-ventral (D→V). Image adapted from *www.insectslimited.com.* 41

FIGURE II.2 – Maternal effect genes that establish polarity creating a protein gradient throughout the oocyte. The four genes: *hunchback* (*hbc*), *caudal* (*cdl*), *bicoid* (*bcd*) and *nanos* (*nos*) are transferred from the ovarian nurse cells into the oocyte: *bicoid* is trapped in the anterior pole while *nanos* is transported into the posterior pole. When translated, Bicoid protein creates a gradient A→P whilst Nanos protein creates an inverse gradient P→A. Caudal and Hunchback proteins gradient are responsible for the inhibition of the *bcd* and *nos* translation. Adapted from [32] 42

FIGURE II.3 – Body segmentation in *Drosophila*: comparison between the larval and adult fly. Image from [32] 42

FIGURE II.4 – Distribution of Exuperantia protein in *Drosophila* wild-type egg chambers. A-D: Representation of the developing oocyte where the nurse cells are on the left. Image adapted from [36] 44

FIGURE II.5 - Basic local alignment of Exuperantia amino acid sequence with its homologous based on NCBI Blast. [40], [41] and possible PEST (Proline-Glutamate-Serine-Threonine) sequence based on EMBOSS *epestfind*. [42] 45

FIGURE II.6 - GLOBPLOT (Intrinsic Protein Disorder, Domain & Globularity Prediction) for Exuperantia. This software (http:// globplot.embl.de) is a web service that allows plotting the tendency within the query protein for order/globularity and disorder. [43] 46

FIGURE II.7 – Basic local alignment of Yps primary sequence with four other proteins that share a high level of homology. In red are the highly conserved residues among species - the cold shock domain. The green rectangle highlights the residues that form the RNA recognition motifs - RNP1 and RNP2. [28][29] 47

FIGURE II.8 – Schematic representation of the typical topology of the RRM: four-stranded β-sheet with the position of the conserved RNP1 and RNP2 aromatic residues. Adapted from [51] 47

FIGURE II.9 – Representation of the three conserved aromatic residues which are responsible for the RNA binding surface in proteins that hold a cold shock domain. The model used is the cold shock domain from Y-box protein 1 from *Homo sapiens* (PDB entry: 1H95). [46].....48

FIGURE II.10 - Typical chromatographic profile of N-His_6-Sumo-Yps_CSD after the IMAC purification step. Injected sample volume of 50ml; Binding buffer: 20mM Tris HCl pH7.9, 500mM NaCl, 10mM imidazole, 3% glycerol. Elution buffer: 20mM Tris-HCl pH7.9, 500mM NaCl and 500mM imidazole, 3% glycerol. The fractions 1 to 6 were aliquoted and analysed by SDS-PAGE. Blue line represents the A_{280nm} and green line represents the elution gradient. 50

FIGURE II.11 - SDS-PAGE analysis of the purified protein – run at 40V for 30 minutes in Tris buffer and stained with a Coomassie based-solution. Lane 1 – molecular weight marker (kDa) and following lanes have the same nomenclature as the figure II.10 chromatogram. N-His_6-Sumo-Yps_CSD expected molecular weight: 25kDa. 51

FIGURE II.12 – Left: Typical chromatographic profile of Sumo-Yps_CSD after the SEC purification step. Injected sample volume of 0.5ml; Elution buffer: 20mM Tris HCl pH8, 100mM NaCl, 1mM DTT and 3% glycerol. The fractions 1 to 3 were aliquoted and analysed by SDS-PAGE. Right: SDS-PAGE analysis of the purified protein – run at 40V for 30 minutes in Tris buffer and stained with a Coomassie based-solution. Lane 1 – molecular weight marker (kDa) and following lanes have the same nomenclature as the corresponding chromatogram. Yps_CSD expected molecular weight: 9kDa. 51

FIGURE II.13 - Thermal shift assay result for Yps_CSD in the 36 buffer conditions listed on appendix F. The more stable the protein is in solution, the more likely it is to produce a crystal. In this sense, the buffers that yield the most promising results were used to solubilize Yps_CSD before proceeding to the crystallization trials. 52

FIGURE II.14 – Possible Yps_CSD crystal obtained for the crystallization trials of Yps_CSD, using $CaCl_2$ (left), PEG 8K (middle) and 12.5% w/v 12.5% PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v MPD (right) as precipitant. 53

FIGURE II.15 - Possible Exuperantia protein crystals, with dimensions 0,04x0,04x0,04mm^3, obtained in 24 hours by the method of vapour diffusion in sitting drop using A,B) PEG 8K and C) sodium/potassium phosphate solution as precipitant. 55

FIGURE II.16 – On the left: Typical chromatographic profile of N-TF-Exu2.210 after the second purification step, the size exclusion chromatography. Injected sample volume of 5ml;

Buffer: 20mM Tris-HCl pH7.5, 200mM NaCl, 1mM TCEP. Superdex 200 HiLoad 16/60 column (GE Healthcare). The fractions B6, B5, B4, B3, B2, B1, C1 and C2 were aliquoted and analysed by SDS-PAGE. The blue line corresponds to the A_{280nm} and the red line corresponds to the sample conductivity. On the right: HT-SDS-PAGE analysis of the purified protein – Invitrogen NuPAGE Gels – run at 200V for 40 minutes in BisTris buffer and stained with a Coomassie based-solution. Lane 1 – molecular weight marker (kDa) and following lanes have the same nomenclature as the chromatogram. N-TF-Exu2.210 expected molecular weight: 73kDa. 56

FIGURE II.17 – On the left: Typical chromatographic profile of N-TF-Exu2.193 after the second purification step, the size exclusion chromatography. Injected sample volume of 5ml; Buffer: 20mM Tris-HCl pH7.5, 200mM NaCl, 1mM TCEP. Superdex 200 HiLoad 16/60 column (GE Healthcare). The fractions D11, D10, D9, D8, D7, D6 and D5 were aliquoted and analysed by SDS-PAGE. The blue line corresponds to the A280nm and the red line corresponds to the sample conductivity. On the right: HT-SDS-PAGE analysis of the purified protein – Invitrogen NuPAGE Gels – run at 200V for 40 minutes in BisTris buffer and stained with a Coomassie based-solution. Lane 1 – molecular weight marker (kDa) and following lanes have the same nomenclature as the chromatogram. N-TF-Exu2.193 expected molecular weight: 72kDa. 56

FIGURE II.18 – Analysis of the globular and non-globular domains of Exuperantia by Globplot2. [43] The residues in bold correspond to the globular domains of the protein: 1-206 and 248-403 and secondary structure prediction by iTASSER. Helices highlighted in red, Strands highlighted in blue and Coils highlighted in black. [44] 59

FIGURE II.19 - Analysis of the globular and non-globular domains of Ypsilon-Schachtel by GlobPlot. [43] The residues in bold correspond to the Cold Shock Domain: residues 63-131 ... 60

FIGURE III.1 - Representation of the structural differences in Gram-positive and Gram-negative cell walls. A) The Gram-negative cell wall has three layers: the outer membrane (OM), the peptidoglycan and the cytoplasmic or inner membrane (IM). B) The Gram-positive cell wall has no OM and the peptidoglycan layer is thicker. 69

FIGURE III.2 – Chemical structure of the peptidoglycan pentapeptide. a) Typical pentapeptide sequence. Adapted from [74] b) *Staphylococcus aureus* peptidoglycan chemical structure. Adapted from [75]. L-Ala: L-alanine, D-*iso*-Glu: D-glutamic acid, *meso*-A_2pm: meso-diaminopimelic acid, D-Ala: D-alanine, L-Lys: L-lysine. Image created using ChemDraw®. .. 72

FIGURE III.3 – Schematic representation of the peptidoglycan biochemical pathway. In the overall process, precursors are synthesized in the cytoplasm, linked to the transport lipid and flipped across the inner membrane to finally be incorporated into the main chain. Adapted from [70][76]. 73

FIGURE III.4 – Schematic representation of *S. aureus* COL genome region that includes the two genes of interest: SACOL 1951 – *murT* and SACOL 1950 – *gatD* and the upstream and downstream regions. *murT* and *gatD* genes are transcribed in the same direction and no promoter was found between these two genes. Adapted from [5] 74

FIGURE III.5 – Comparison between the three domains characteristics of Mur ligases with MurT. a) This scheme represents the general three domains topology of Mur ligase proteins. N-terminal domain is responsible for the UDP-MurNAc-peptide binding, the C-terminal domain is responsible for the binding of the incoming amino acid and the central domain is where the ATP and co-factor Mg^{2+} bind. b) MurT topology lacks the flanking N- and C-terminal sequences and the C-terminal has a domain of unknown function that is conserved in more than 900 prokaryotic proteins – DUF1727. Adapted from [5] 75

FIGURE III.6 – Comparison of glutamine-dependent amidotransferase structure with GatD. a) glutamine-dependent amidotransferases modular structure representation: a substrate binding domain (synthase domain) and a GAT domain with its conserved residues for glutaminase activity. b) GatD shares the GAT domain and the glutaminase activity motifs. Adapted from [5] 75

FIGURE III.7 – On the left: Typical chromatographic profile of N-His_6-GatD after the second purification step, the size exclusion chromatography. Injected sample volume of 6ml; Buffer: 20mM Tris-HCl pH7.5, 200mM NaCl, 1mM TCEP. Superdex 200 HiLoad 16/60 column (GE Healthcare). The fractions A12, C1 to C12 (*) were aliquoted and analysed by SDS-PAGE. The blue line corresponds to the A280nm and the red line corresponds to the sample conductivity. On the right: HT-SDS-PAGE analysis of the purified protein – Invitrogen NuPAGE Gels – run at 200V for 40 minutes in BisTris buffer and stained with a Coomassie based-solution. Lane 1 – molecular weight marker (kDa) and following lanes have the same nomenclature as the chromatogram. N-His_6-GatD expected MW: 27kDa. 79

FIGURE III.8 – On the left: Typical chromatographic profile of N-His_6-SeGatD after the second purification step, the size exclusion chromatography. Injected sample volume of 5ml; Buffer: 20mM Tris-HCl pH7.5, 200mM NaCl, 1mM TCEP. Superdex 200 HiLoad 16/60 column (GE Healthcare). The fractions A8, A11, B12, C4 to C12 and E2 (*) were aliquoted and analysed by SDS-PAGE. The blue line corresponds to the A280nm and the red line corresponds to the sample conductivity. On the right: HT-SDS-PAGE analysis of the purified protein – Invitrogen NuPAGE Gels – run at 200V for 40 minutes in BisTris buffer and stained with a Coomassie based-solution. First and last lanes – molecular weight marker (kDa) and following lanes have the same nomenclature as the chromatogram. N-His_6-GatD expected MW: 27kDa. 80

FIGURE III.9 – Mass spectra acquired for native (on the left) and labelled (on the right) GatD, by manual drop analysis and direct injection MS in a LC-ESI-MS. Sample concentration: 20μM. Service available at OPPF-UK and kindly performed by Dr. Joanne Nettleship. 80

FIGURE III.10 – GatD protein crystals. Representative selection of different morphology obtained for GatD protein crystals and their conditions. Common to all crystallization solutions is the use of polyethylene glycol (PEG) as precipitant. 82

FIGURE III.11 - SeGatD protein crystals. Representative selection of different morphology obtained for SeGatD protein crystals and their conditions. Common to all crystallization solutions is the use of polyethylene glycol (PEG) as precipitant. 83

FIGURE III.12 – On the left: Chromatographic profile of N-His_6-GatD co-purified with N-His_6-Sumo-MurT after the second purification step, the size exclusion chromatography. Injected sample volume of 5ml; Buffer: 20mM Tris-HCl pH7.5, 200mM NaCl, 1mM TCEP. Superdex 200 HiLoad 16/60 column (GE Healthcare). The fractions A11, B9 B1, C5, C8, C9, C11, C12, D12, D9, D6 and D4 (*) were aliquoted and analysed by SDS-PAGE. The blue line corresponds to the A_{280nm} and the red line corresponds to the sample conductivity. On the right: HT-SDS-PAGE analysis of the purified protein – Invitrogen NuPAGE Gels – run at 200V for 40 minutes in BisTris buffer and stained with a Coomassie based-solution. First and last lanes – molecular weight marker (kDa) and following lanes have the same nomenclature as the chromatogram. N-His_6-GatD expected MW: 27kDa and N-His_6-Sumo-MurT MW: 61kDa. 86

FIGURE III.13 – On the left: Chromatographic profile of co-expression of N-His_6-GatD with N-His_6-Sumo-MurT after the second purification step, the size exclusion chromatography. Injected sample volume of 5ml; Buffer: 20mM Tris-HCl pH7.5, 200mM NaCl, 1mM TCEP. Superdex 75 HiLoad 16/60 column (GE Healthcare). The fractions A11, A12, B12, B11, B10, B9, B8, B7, B6, B5, B4, B3, B2, B1, C1, C2, C3, C5, C6, C8, C9, C10, C11 and C12 (*) were aliquoted and analysed by SDS-PAGE. The blue line corresponds to the A_{280nm} and the red line

corresponds to the sample conductivity. On the right: HT-SDS-PAGE analysis of the purified protein – Invitrogen NuPAGE Gels – run at 200V for 40 minutes in BisTris buffer and stained with a Coomassie based-solution. First and last lanes – molecular weight marker (kDa) and following lanes have the same nomenclature as the chromatogram. N-His_6-GatD expected MW: 27kDa and N-His_6-Sumo-MurT MW: 61kDa. 88

FIGURE III.14 – Possible MurT-GatD protein crystals. Representative selection of different morphology obtained Common to all crystallization solutions is the use of polyethylene glycol (PEG) as precipitant. 90

FIGURE III.15 – Representative diffraction patterns of the crystals. Left: native; Right: SeMet derivative. The resolution at the edge of the detector is 2.1 and 2.5Å, respectively. 91

FIGURE III.16 – GatD model in a light blue cartoon representation with the anomalous electron density (in orange at a 3σ contour) for the two (out of three) atoms of selenium in the SeGatD derivative. This figure and other structure-related figures reported in this thesis were prepared using WinPymol. [60] 92

FIGURE III.17 - The experimental phases to determine the 3D structure of GatD were obtained by single-wavelength anomalous diffraction (SAD) using data collected from the SeMet derivative at the Se edge peak. The model was refined with Refmac5 from CCP4. The 2mFo-DFc electron density maps are contoured at 1σ (blue). 92

FIGURE III.18 – Sequence alignment of four proteins with conserved glutamine amide transfer domains. The highlighted residues correspond to the conserved triad of residues responsible for the catalytic function. gi_23465027: cobyric acid synthase CobQ from *Bifidobacterium longum*; gi_18144850: probable cobyric acid synthase from *Clostridium perfringens* str. 13; gi_17131811: alr2718 from *Nostoc sp.*; gi_23003197: hypothetical protein Lgas_03000924 from *Lactobacillus gasseri.* [47].....95

FIGURE III.19 – Crystallographic dimer of GatD in a cartoon representation and coloured according to the secondary structure: *α*-helices in blue, *β*-sheets in red and loops in purple. This figure and other structure-related figures reported in this thesis were prepared using WinPymol. [60].....96

FIGURE III.20 – GatD secondary structure analysis. The catalytic triad has a reactive cysteine at position 94 in the turn between a β-strand and α-helix designated as the 'nucleophile elbow'. The residues highlighted in a green box are GatD catalytic triad. Image generated by PDBSum. [94].....97

FIGURE III.21 - GatD topology. GatD has a mixed topology, instead of a α-β-α fold characteristic of glutaminase structures. Image generated by PDBSum. [94].....97

FIGURE III.22 – On the left: Overall representation of GatD and the position of the residues that compose the triad within the molecule, located at positions 94, 149 and 151. On the right: closer view of the catalytic triad residues. The cartoon representation is from the monomeric GatD. The residues C_{94}-His_{149}-Gly_{151} are represented as sticks and coloured by atoms: carbons in white, nitrogen in blue, oxygen in red and sulfur in yellow. 98

FIGURE III.23 - Distances (Å) between the catalytic triad residues. The mean donor-acceptor distances in proteins structure elements are approximately 3Å. In this structure, the distance between this two residues is >3Å, which suggest no interaction between histidine and cysteine residues and consequently, no activation of the nucleophile. The residues are coloured by atoms: carbons in white, nitrogen in blue, oxygen in red and sulfur in yellow. 98

FIGURE III.24 – Superimposition of GatD structure with three other type 1 GATases [96]. Glutamine Amidotransferase from *Thermotoga maritima* (PDB ID: 1KXJ); Imidazole glycerol phosphate synthase subunit from *Thermotoga maritima* (PDB ID: 3ZR4); HISF Protein from *Thermotoga maritima* (PDB ID: 1GPW) and Amidotransferase HisH from *Thermotoga maritima* (PDB ID: 1K9V) in grey. The proteins are represented as cartoon and GatD is coloured in red. RMSD: 5.238 (1GPW, 1K9V); 5.250 (1KXJ); 5.220 (3ZR4).....99

FIGURE III.25 – Superimposition of the catalytic triad Cys_{94}-His_{149}-Gly_{151} from GatD with the catalytic triad Cys_{84}-His_{178}-Glu_{180} of the homologous structures: Glutamine Amidotransferase from *Thermotoga maritima* (PDB ID: 1KXJ); HISF Protein from *Thermotoga maritima* (PDB ID: 1GPW) in grey. GatD is represented in red and the catalytic triad's residues are shown as ribbon and coloured by atom type: carbon in white, nitrogen in blue, oxygen in red and sulfur in yellow.....100

FIGURE III.26 - Comparison between glutamine (left) and its analogue 6-diazo-5-oxo-L-norlucine (DON).....101

FIGURE III.27 – Cluster predicted by AutoDock Vina in a GatD:glutamine interaction at the active site and a closer look at the active site. The residues are coloured by atom type: carbon in white, nitrogen in blue, oxygen in red and sulfur in yellow. [99].....102

FIGURE III.28 - Cluster predicted by AutoDock Vina in a DON:GatD interaction at the active site and a closer look at the active site. The residues are coloured by atom type: carbon in white, nitrogen in blue, oxygen in red and sulfur in yellow. [99].....103

FIGURE III.29 – Proposed mechanism for GatD-MurT combined action. Glutamine (Gln) is used as a nitrogen donor and is conversed at the glutaminase site in Glutamate (Glu) and ammonia (NH_3). Ammonia is then transferred from GatD to the synthase domain, possibly present in MurT, across an ammonia channel, to finally amidate the lipid-II precursor. [6] 104

Table Index

TABLE I.1 - Crystal systems and protein space groups. The space group is characterised by the number of lattice points within the unit cell. P, C, I and F represent the four types of unit cell: primitive, end-face centered, body centered and face-centered. Adapted from [13] 34

TABLE II.1 – Summary table of the gene sequences selected, vectors used and *E. coli* expression strains in the first part of this experimental work. 49

TABLE II.2 – List of the 96-well Greiner plates' barcodes, the commercial crystallization screens, concentrations and temperatures trialled in order to crystalize the tagged protein sample: N-TF-Exu2.210 and N-TF-Exu2.193. 57

TABLE II.3 – Growth temperatures and times tested for Exu expression in *E.coli*. 61

TABLE III.1 – Constructs designed at Oxford Protein Production Facility. Further details on appendix C.....77

TABLE III.2 – Macromolecule production information. GatD construct 12143. Further details on appendix C.....78

TABLE III.3 – List of the 96-well Greiner plate's barcodes, the commercial crystallization screens, and concentrations and temperatures trialled in order to crystalize the native protein sample: GatD. 81

TABLE III.4 – List of the 96-well Greiner plate's barcodes, the commercial crystallization screens, concentrations and temperatures trialled in order to crystalize the labelled protein SeGatD. 83

TABLE III.5 - Macromolecule production information. MurT construct 12139. Further information on appendix C. 84

TABLE III.6 – List of the 96-well Greiner plate's barcodes, the commercial crystallization screens, concentrations and temperatures trialled in order to crystalize the protein N-His_6-Sumo-MurT. 85

TABLE III.7 - List of the 96-well Greiner plate's barcodes, the commercial crystallization screens, concentrations and temperatures trialled in order to crystalize the protein complex MurT-GatD.....89

TABLE III.8 – List of the 96-well Greiner plate's barcodes, the commercial crystallization screens, concentration, temperature and protein:reservoir drop ratio (P:R) trialled in order to optimize the hits obtained for the protein complex MurT-GatD. 90

TABLE III.9 - X-ray crystallography data-collection, processing and refinement statistics. (Values in parenthesis correspond to the outer shell) [91] 93

TABLE III.10 – Affinity binding energies from the 9 modes of glutamine predicted for the glutamine:GatD complex predicted by AutoDock Vina. [97].....101

TABLE III.11 - Affinity binding energies from the 9 modes of DON predicted for the GatD:DON complex predicted by AutoDock Vina. [97].....102

Equations Index

EQUATION I.1 - Bragg's law diffraction equation. 35

EQUATION I.2 - Calculation of V_M (Matthew's coefficient). Units: $Å^3Da^{-1}$. In this equation, V represents the volume of the unit cell ($Å^3$), Z is the number of asymmetric units and M is the molecular weight of the asymmetric unit contents (Da). [17] 36

EQUATION I.3 - Electron density equation $\rho x, y, z$ is given by a Fourier transform of the sum functions describing the atoms positions (x,y,z) in the crystal. V is the unit cell volume; $Fhkl$ corresponds to the structure factor amplitude and is experimentally obtained; αhkl is the phase angle which is not obtained directly from the diffraction experiment – phase problem. [8].....37

EQUATION I.4 – R-*factor* equation. The R-*factor* (or R_{work}) is a statistical measurement that relates the observed reflection amplitudes (F_{obs}) and those calculated from the model (F_{calc}), in order to measure how well the refined data fits to the observed data. 38

Abbreviations

3D	Three-Dimensional
A.U.	Asymmetric Unit
AP	Anterior-Posterior
AS	Anomalous Scattering
ATP	Adenosine Triphosphate
bcd	*bicoid*
BLAST	Basic Local Alignment Search Tool
CDD	Conserved Domain Database
cDNA	Complementary DNA
CSD	Cold Shock Domain
Da	Dalton
DGRC	*Drosophila* Genomics Resource Center
D-*iso*-Glu	D-*iso*-Glutamic Acid
DLS	Diamond Light Source
DNA	Deoxyribonucleic Acid
DSF	Differential Scanning Fluorimetry
DUF	Domain Of Unknown Function
DV	Dorsal-Ventral
E. coli	*Escherichia coli*
EM	Electron Microscopy
Exu	Exuperantia
GAT	Glutamine-dependent Amidotransferase
GF	Gel Filtration
GlcNAc	N-Acetyl-Glucosamine
Gln	Glutamine
Glu	Glutamate
hbc	*hunchback*
HTP	High Throughput
IM	Inner Membrane
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Isomorphous Replacement
L-Ala	L-Alanine
LC-ESI-MS	Liquid Chromatography-Electrospray Ionisation-Mass Spectrometry
L-Lys	L-Lysine
LPS	Lipopolysaccharide
MCS	Multiple Cloning Sites

***meso*-A_2pm**	Meso-Diaminopimelic Acid
MR	Molecular Replacement
mRNA	messenger RNA
MRSA	Methicillin-Resistant *Staphylococcus Aureus*
MurNAc	N-Acetylmuramic Acid
MW	Molecular Weight
NCBI	National Center for Biotechnology Information
NMR	Nuclear Magnetic Resonance
nos	*nanos*
OB	oligonucleotides and/or oligosaccharides
OD	Optical Density
OM	Outer Membrane
OPPF	Oxford Protein Production Facility
osk	*oskar*
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PCT	Pre-Crystallization Test
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PG	Peptidoglycan
RNA	Ribonucleic Acid
RNP	Ribonucleoprotein
Rpim	Precision-Indicating Merging Factor R
RRM	RNA recognition motif
S. aureus	*Staphylococcus aureus*
SAD	Single-Wavelength Anomalous Diffraction
SCOP	Structural Classification of Proteins
SDS	Sodium Dodecyl Sulfate
SEC	Size Exclusion Chromatography
SeMet	Selenomethionine
TCEP	tris(2-carboxyethyl)phosphine
TF	Tissue Factor
UDP	Uridine Diphosphate
Yps	Ypsilon-Schachtel

"Estuda sempre.
O estudo liberta da ignorância e
favorece a criatura com discernimento.
O estudo e o trabalho são as asas
que facilitam a evolução do ser.
O conhecimento é um bem que,
por mais que seja armazenado,
jamais toma qualquer espaço.
Pelo contrário,
faculta a mais ampla facilidade
para novas aquisições.
O homem que ignora, caminha às escuras."

Joanna de Ângelis

Abstract

I. mRNA localization mechanisms

Asymmetry is critical in higher organisms. mRNA is always found in complex with many proteins, forming ribonucleoprotein particles (RNP). These particles travel along cytoskeletal filaments with the help of motor proteins. mRNA transcripts contain *cis*-acting elements which determine the timing of expression and final destination of the RNA and therefore of the encoded protein. These also determine which molecular motor will be used to reach the proper destination. Early genetic studies indicated that the proteins Exuperantia (Exu) and Swallow (Swa) are important for the anterior localization of *bicoid* mRNA. Recent studies showed that Exu is also present in RNP particles containing *oskar* mRNA, and is important for its posterior localization. It requires the presence of Ypsilon-Schachtel (Yps), a member of the cold shock family of RNA binding proteins. Exu and Yps were shown to co-purify, and interact *in vitro* even in the absence of RNA. Exu seems to be a core component of the transport complex.

Although many proteins have been implicated in this mechanism, there is still very little biochemical and structural information about the process. We intend to study the structural determinants which control mRNA transport within the cell. [1][2][3]

Structural studies of these proteins, in particular Exuperantia and Ypsilon-Schachtel, will provide a wealth of information that will help us to understand the mechanisms involved in mRNA sorting, loading onto the correct carrier, anchoring and translation regulation.

In order to determine protein structure of Exu and Yps, the protein encoding region was cloned into a pGEX and pET vector systems and expressed in several *E. coli* expression strains and well as in a high-throughput facility where both genes were cloned into the pOPIN plasmids. The protein of interest was purified using the most common chromatographic methods: Immobilized metal ion affinity chromatography (IMAC), affinity chromatography and size-exclusion chromatography and the protein purity was estimated by polyacrylamide gel electrophoresis (SDS-PAGE).

With a pure protein sample was obtained, crystallization trials were performed and several crystallization screens were used and promising hits were obtained. Although from all the crystals obtained, the size of the crystal was the limiting step and no data was collected as well as all the crystal optimizations were unsuccessful.

II. Amidation of *S. aureus* peptidoglycan residues

The basic structure of *S. aureus* peptidoglycan and its synthesis pathway are well characterized. Nevertheless, the mechanisms by which glutamic acid residues in bacterial cell walls undergo modification are still poorly understood. Inhibition of this step of bacterial cell wall synthesis reduces growth rate, resistance to β-lactam antibiotics and increased sensitivity to lysozyme. Two genetic determinants, *murT* and *gatD*, have been identified and were shown to be required and sufficient to perform this step. The *murTgatD* operon emerged as a syntenic block that seems to be widespread among bacteria. The genome co-localization of the two determinants, together with data from sequence analysis, suggests a coordinated function of MurT and GatD proteins in the peptidoglycan glutamate amidation. Both proteins together harbour all domain functions required for amidation of peptidoglycan precursor: MurT may be responsible for the recognition of the reaction substrates, the lipid linked peptidoglycan precursor and ATP, while GatD could be the catalytic subunit involved in the transfer of the amino group from free glutamine to the peptidoglycan precursor. The GatD sequence lacks an ATP binding motif which is common to all members of the GnAT family suggesting an activity that depends on the MurT protein which exhibits a typical Mur ligase central domain including the ATP binding motif. The structures of these two proteins by themselves and in complex will help confirm this model and understand the last missing genetic determinant to account for the structural variation in the *S. aureus* peptidoglycan. [4][5][6]

In order to determine the *S. aureus* GatD protein structure, the protein encoding region was cloned into a pOPIN plasmid and expressed in *E. coli* Lemo21(DE3) as a N-terminal His-tag fusion. The protein of interest was purified using the most common chromatographic methods: Immobilized metal ion affinity chromatography (IMAC) and Size-exclusion chromatography. The protein purity was estimated by polyacrylamide gel electrophoresis (SDS-PAGE), which showed a single band corresponding to the molecular weight of GatD. With a pure protein sample, crystallization trials were performed and several crystallization screens were used and positive hits were obtained using the Emerald Wizard I and II screen (Rigaku Reagents). Diffraction data were collected at Diamond Light Source – beamlines I02 and I04 – to a resolution beyond 1,9Å. Initial phases were obtained by single-wavelength anomalous diffraction (SAD) using data collected from SeMet derivatives at the Se edge peak. The crystals belong to the space group $P2_12_12_1$ with unit-cell dimensions: a=48,29Å; b=93,00Å and c=109,31Å. The preliminary structural analysis confirms the similarity of GatD to others glutamine amidotransferases, already known, as well as a superimposition of the active site confirming the existence of the catalytic triad characteristic of this type of enzymes.

Part of the work described in this chapter, was the subject of the following publication:

D. Vieira, T. a. Figueiredo, A. Verma, R. G. Sobral, A. M. Ludovice, H. de Lencastre, and J. Trincao, "Purification, crystallization and preliminary X-ray diffraction analysis of GatD, a glutamine amidotransferase-like protein from *Staphylococcus aureus* peptidoglycan," *Acta Crystallogr. Sect. F Struct. Biol. Commun.*, vol. 70, no. 5, pp. 1–4, Apr. 2014.

Chapter I

X-Ray Protein Crystallography

"... the chemist of the future who is interested in the structure of proteins, nucleic acids, polysaccharides, and other complex substances with higher molecular weights will come to rely upon a new structural chemistry, involving precise geometrical relationships among the atoms in the molecules ..."

Linus Pauling, Nobel Lecture, 1954

Crystals have long been admired for their orderliness and their beauty. Their properties were scientifically explored from the 17th century. It took over one century of extensive research to validate their inner structure, which was made possible only by the development of X-ray crystallography in the late 19th century.

X-rays were discovered in 1895 by Wilhelm Conrad Röntgen. Since then, X-rays have become an invaluable tool for the study of the atomic structures and properties of molecules.

The discovery of the structures of important biological molecules began in the late 1950s, with the structure of sperm whale myoglobin by John Cowdery Kendrew, followed by the DNA structure by J. Watson, F. Crick, R. Franklin and M. Wilkins, and Dorothy Hodgkin, who solved the structures of penicillin, vitamin B12 and insulin.

Since J. Kendrew's success, over 80000 crystal structures of proteins, nucleic acids and other biological molecules have been determined by X-ray crystallography and nowadays (September 2014) there are 103354 structures, including proteins, nucleic acids and protein-nucleic acid complexes and others, deposited in the Protein Data Bank (PDB). Some of these structures were obtained by other structural techniques, such as nuclear magnetic resonance (NMR) and electron microscopy (EM), although the vast majority (~89%) was solved by X-ray crystallography.

Overview of Protein Crystallography

Protein crystallography allows us to visualize protein structures at an atomic level, improving our understanding of protein function. The basic principle is the interaction of X-rays with the electrons in the atoms. In order to see molecules in atomic detail, electromagnetic radiation with a wavelength of the same order of magnitude of the chemical bond (around 0.1 nm or 1Å) needs to be used, in other words, X-rays. [7]

The bottleneck in this technique is the absolute need of a single crystal. The diffraction from a single molecule is still too weak to be measurable. To overcome this problem and amplify the signal to be measured an ordered three-dimensional array of molecules - a crystal – is required. The X-rays are then scattered by the electrons in the structure and, consequently, the result is a three-dimensional map representing the electron density of the molecule. To calculate electron densities from a diffraction experiment we need to gather information regarding **i)** the indices of a reflection (*h*,*k*,*l*); **ii)** the intensity of the reflection (I_{hkl}); **iii)** the phase angles of the reflection (α_{hkl}). The first two are obtained directly from the experiment. The indices are determined by the crystal symmetry, the intensities are measured from the photons that reach the detector, whereas the phase angles depend on the atoms distance between the Bragg's planes and the interference of the radiation diffracted by those planes. In this sense, the phases have to be further determined in order to calculate the electron density model. This missing piece of information and the way of recovering it is called "the phase problem" and a long part of crystallography is dedicated to solving it. [8][9]

There are several ways to recover the lost phases and they all involve acquisition of new data either from anomalous scattering by heavy atoms present in the structure or from the incorporation of new heavy atoms or by using a structure of a similar protein as a starting model. Once the phases are obtained, the preliminary model can be built into an electron density map which is little by little completed and refined. After the refinement is completed, the structure is validated and the coordinates are deposited in the PBD.

The global process is summarized on figure I.1.

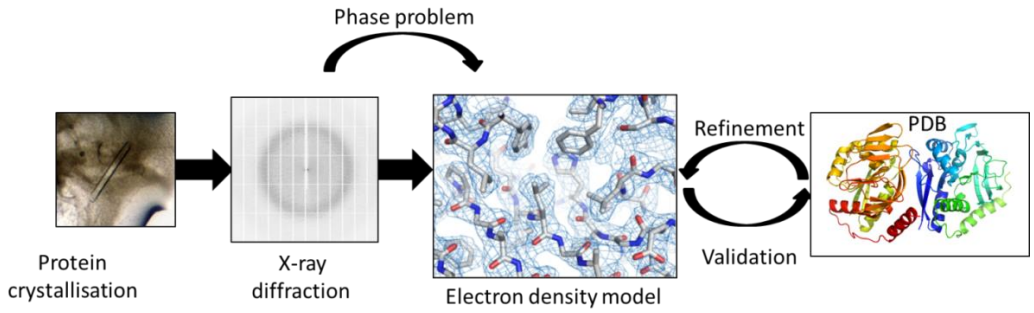


FIGURE I.1 - X-ray protein crystallography process overview. The description is detailed on the text above.

Crystal Production

The first step in order to produce a protein crystal is the preparation of large enough quantities of protein (generally in the milligram range) in a highly purified form. This starting material can either be obtained from its natural source or expressed heterologously by gene cloning, followed by purification that includes one or more chromatographic steps. The key to a successful and reproducible crystallization is a highly soluble, pure and monodispersed (of a single multimerisation state) starting sample.

The most common method to evaluate the purification yield is by running a polyacrylamide gel electrophoresis (SDS-PAGE) and by the light absorbance at 280nm.

The production of a protein crystal is the most critical and, sometimes, the most time consuming stage in protein crystallography since it follows a trial and error principle. Due to the difficulty in predicting the ideal crystallization conditions, initial crystallization trials must be performed with different precipitants, concentrations, pH and temperatures. Two of the most used methods for protein crystallization are vapour diffusion - hanging drop and vapour diffusion - sitting drop methods (Figure I.2); both methods require a protein solution drop with buffer and precipitant to equilibrate with a reservoir solution containing buffers and precipitants at higher concentrations. As time goes by water evaporates from the drop in to the reservoir increasing the precipitant concentration to an optimal level for crystallization. With the system in equilibrium, conditions are kept until crystals appear in the drop.

A conventional explanation of crystal formation and growth is given by the crystallization phase diagram (Figure I.3). In a vapour diffusion experiment, water evaporates from the protein solution which will start to concentrate from the unsaturated zone to reach a supersaturated zone. After the first crystals appear, the protein concentration decreases and the crystals will grow until it reaches the solubility curve.

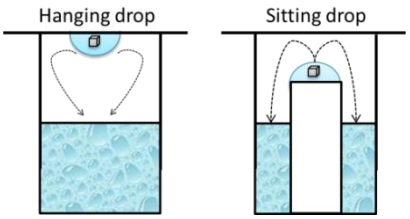


FIGURE I.2 – Vapour diffusion techniques: Hanging drop and sitting drop. Both methods require a protein solution drop with buffer and precipitant to equilibrate with a reservoir solution containing buffers and precipitants at higher concentrations.

Once the crystal is produced and prior to data collection, it's necessary to harvest the crystal from the drop and protect it. Protein crystals are quite fragile since they have high solvent content (20-80%). To overcome this issue and to prevent crystal dissolution they need to be

harvested with mother liquor with higher concentration of precipitant and then transferred to a cryo-protecting solution to prevent ice formation. To protect the crystals from radiation damage, these are cryo-cooled to ~100K and maintained in a stream of cold nitrogen during data collection. [10]

Once the crystal is obtained, harvested and cryoprotected, the next step is to mount it on a focused x-ray beam in order to be diffracted creating a reflection pattern that is recorded in the detector.

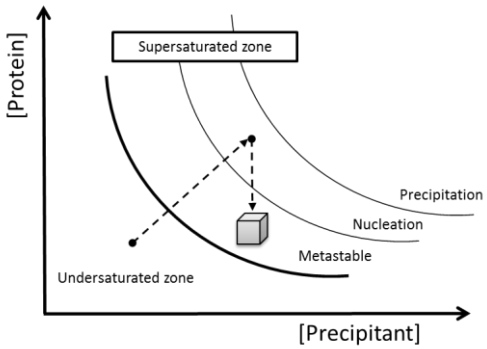


FIGURE I.3 - Crystallization phase diagram. Schematic representation of a phase diagram illustrating the variation of protein concentration with the precipitating concentration.

Why do we need a crystal?

By definition, a crystal is a solid material whose constituents (atoms, molecules or ions) are organised in a well-ordered arrangement covering all three spatial dimensions that occur due to the intrinsic nature of molecules to form symmetric patterns. To build the crystal lattice we start with the asymmetric unit - the smallest possible unit cell spatial occupation. Applying crystallographic symmetry operations we obtain the unit cell - identical blocks that are repeated throughout the lattice and are characterised by the lengths of the cell edges (*a*,*b* and *c*) and the angles between them (α, β and γ). The crystal is obtained therefore by the translation of the unit cell in all three spatial dimensions. [11]

The set of symmetry operations that generate the unit cell from the asymmetric unit is called the space group. The space group is characterised by the number of lattice points within the unit cell, the symmetry operations and the geometry of the unit cell. Combining the seven crystal systems with the four different unit cells we obtain 14 Bravais lattices. By adding translation symmetry operations we obtain a total of 230 different space groups. Proteins are chiral molecules which mean that mirror planes or inversion centres are not allowed, which reduces the number of possible protein space groups to 65, as represented in table I.1. [11]

Data Collection and X-Ray Diffraction

When a suitable macromolecular crystal is selected, the next step is to check if it diffracts X-rays and, when it does, if it is good enough to be used for structure determination.
The crystal structure described in this thesis was obtained from data collected at Diamond Light Source, Harwell Science and Innovation Campus, Didcot, United Kingdom (Figure I.5 on the left).

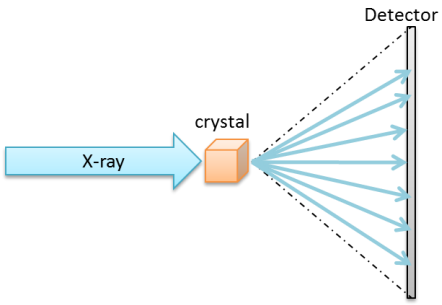


FIGURE I.4 - Schematic representation of an X-ray experiment. The X-rays are generated by a source. The crystal mounted on the goniometer and when it's hit by the X-rays, diffracts them. The diffracted beam is then recorded on a detector.

The scattered X-ray beam once it hits a crystal is a result from the interactions between the electric component of the beam and the electrons within the crystal structure. A unit cell contains a large number of electrons and the waves scattered by these electrons interfere with each other. [12]

Synchrotron radiation

A synchrotron works as a storage ring, where electrons move around in a circle, accelerated to nearly the speed of light wiggling through a set of magnetic fields, generating intense X-rays. When electron's beam route is bent by magnets, the electrons lose energy in the form of light. This light is then channelled into the experimental stations, the beamlines, where users carry out their diffraction experiments (Figure I.5 on the right).

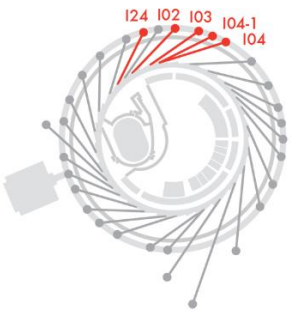
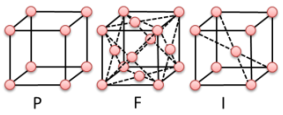
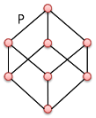
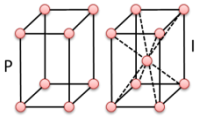
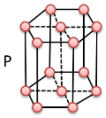
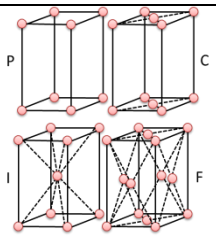
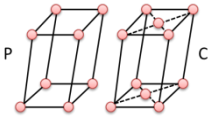
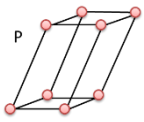


FIGURE I.5 – On the left: Diamond Light Source is the UK national synchrotron. **On the right:** Diamond Macromolecular Crystallography (MX) village overview. MX village currently has five beamlines: three high brilliance MAD beamlines (I02, I03, I04), a fixed wavelength beamline (I04-1) and a microfocus MAD beamline (I24). Image from *http://www.diamond.ac.uk/Beamlines/Mx.html*.

TABLE I.1 - Crystal systems and protein space groups. The space group is characterised by the number of lattice points within the unit cell. P, C, I and F represent the four types of unit cell: primitive, end-face centered, body centered and face-centered. Adapted from [13]

Crystal system	Cell length	Cell angles	Bravais lattices	Space group
Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90°$		P23, F23, I23, $P2_13$, $I2_13$, P432, $P4_232$, F432, $F4_132$, I432, $P4_332$, $P4_132$, $I4_132$
Trigonal (or rhombohedral)		$\alpha = \beta = \gamma \neq 90°$		P3, $P3_1$, $P3_2$, P312, P321, $P3_112$, $P3_121$, $P3_212$, $P3_221$
Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90°$		P4, $P4_1$, $P4_2$, $P4_3$, I4, $I4_1$, P422, $P42_12$, $P4_122$, $P4_12_12$, $P4_222$, $P4_22_12$, $P4_322$, $P4_32_12$, I422, $I4_122$
Hexagonal		$\alpha = \beta = 90°$ $\gamma = 120°$		P6, $P6_5$, $P6_4$, $P6_3$, $P6_2$, $P6_1$, P622, $P6_122$, $P6_522$, $P6_222$, $P6_322$, $P6_422$
Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90°$		P222, $P222_1$, $P2_12_12$, $P2_12_12_1$, $C222_1$, C222, F222, I222, $I2_12_12_1$,
Monoclinic		$\alpha = \gamma = 90°$ $\beta \neq 90°$		P2, $P2_1$, C2
Triclinic		$\alpha \neq \beta \neq \gamma \neq 90°$		P1

The highly accessible use of synchrotrons nowadays, explains the increasing number of structures deposited each day in the PDB. The major advantages are the speed of data collection, the data quality, and especially the intensity and collimation of the beam making possible to use even smaller crystals in the experiments, which is highly important for crystal that are extremely difficult to optimize.

Once an X-ray beam hits a crystal there are interactions between the electric component of the beam and the electrons within the crystal structure, resulting in a unit cell that contains a large number of electrons and waves that are scattered by these electrons and that interfere with each other. [12]

Bragg's Law

In 1913, two physicists, William Lawrence Bragg and his son, William Henry Bragg, postulated a physical model to explain conditions where diffraction was observed. They realised that the arrangement of atoms within a crystal could be determined by the observation of the X-ray beams reflected by planes of atoms in that same crystal. [14]

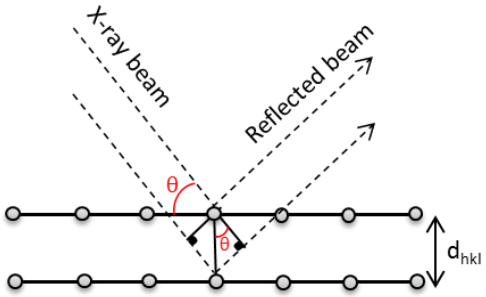


FIGURE I.6 – Schematic representation of the Bragg's Law. In this model, each set of parallel planes is treated as an independent diffractor and produces a single reflection – constructive interference. Adapted from [7]

If λ represents X-rays' wavelength, d_{hkl} the perpendicular distance between planes, θ the angle between the planes of the incident or reflected X-ray, and n is an integer, when the equation I.1 is obeyed then a diffraction pattern is obtained which means the waves emanating from this set of planes are in phase with each other – constructive interference. If the Bragg's law isn't obeyed, the waves aren't in phase and no diffraction pattern is obtained – destructive interference. [15]

$$n\lambda = 2d_{hkl}\sin\theta$$ **EQUATION I.1 -** Bragg's law diffraction equation.

Nonetheless, d_{hkl} is a misleading element; it is not related to the atomic distances in the real space. The subscript indices are the Miller indices, where each index represents an orthogonal

plane to a direction (*h, k, ℓ*) in the basis of the reciprocal lattice vectors and there is an inverse relationship between the crystal lattice (real space) and the reciprocal lattice – the spacing of reflections on the detector. [15]

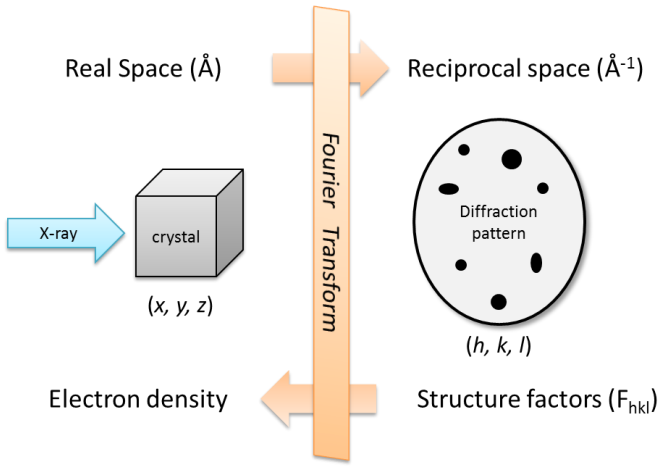


FIGURE I.7 – Elementary representation of an X-ray experiment. A crystal when submitted to an X-ray beam produces reflections which are recorded in the detector creating a diffraction pattern. The diffraction pattern is the result of a Fourier transformation of the crystal real space into reciprocal space and vice-versa.

In order to understand the relationship between the crystal (real space) and its diffraction pattern (reciprocal space), the crystallographer uses a computer that correlates the electron density within the unit cell with the list of reflection's intensities. This mathematical relationship is described as the Fourier transform (Figure I.7). [7]

The data collection experiment allows the crystallographer to begin with information regarding the space group, the crystal unit cell measurements as well as the number of molecules in the asymmetric unit and consequently its volume can be calculated. Protein crystals contain solvent channels. The Matthews Coefficient allows predicting the number of molecules in the unit cell and the solvent content of the crystal. It is calculated using the unit cell parameters and the molecular weight of the molecules present in the unit cell, according to the following equation: [16][17]

$$V_M = \frac{V}{Z.M}$$

EQUATION I.2 - Calculation of V_M (Matthew's coefficient). Units: $Å^3Da^{-1}$. In this equation, V represents the volume of the unit cell ($Å^3$), Z is the number of asymmetric units and M is the molecular weight of the asymmetric unit contents (Da). [17]

This is important information although the real goal of the crystallographer is to answer the question: where are the atoms located within the crystal?

Electron Density and Structure Determination

In order to determine a three-dimensional structure we first need to obtain the best possible electron density map. We then have to interpret the map to build the atomic model.
The diffraction pattern corresponds to the square root of the measured intensities of the structure factors (F_{hkl}). By definition, the structure factor is a wave created by the superimposition of many single waves described as Fourier series.
In order to calculate an electron position (*x,y,z*) in the real space we need to sum all the *hkl* planes' contributions to that particular point as well as the phases (Equation I.3). [8]

$$\rho(x,y,z) = \frac{1}{V}\sum |F_{hkl}|\, e^{(i\alpha_{hkl})}\, e^{(-2\pi i hx+ky+lz)}$$

EQUATION I.3 - Electron density equation $\rho(x,y,z)$ is given by a Fourier transform of the sum functions describing the atoms positions (x,y,z) in the crystal. V is the unit cell volume; F_{hkl} corresponds to the structure factor amplitude and is experimentally obtained; α_{hkl} is the phase angle which is not obtained directly from the diffraction experiment – phase problem. [8]

And since the phase information cannot be measured directly, the Fourier transform cannot be simply applied in order to get the electron density. This fact is known as the phase problem.

The Phase Problem

When a crystallographer says he has determined a structure, he is actually saying that he solved the phase problem, meaning he has obtained enough phase information to calculate an interpretable electron density map.
The phase problem has to be solved in order to determine a structure from its diffraction data and to do so, the used techniques provide estimated phases that are obtained through additional experimental information: **i)** heavy atoms derivative crystals (Isomorphous Replacement), **ii)** anomalous diffractors (Anomalous Scattering) or **iii)** using homologous structural molecular models (Molecular Replacement). [18]
In the first two methods, a normal experiment is conducted, where the electrons in the crystal vibrate in concordance with the incident beam; however, if the incident beam has photons which can take the electrons to a transition state, the vibrational energy changes and re-irradiates in a different phase from the incident beam. [8]
In practice, a crystallographer collects a complete data set with native crystals obtaining the native structure factors – F_P. Next, a second data set is collected from the derivative crystal

giving the reflections in the heavy-atom data – F_{PH} and the difference between the derivative and native structure factors is calculated and the phase obtained.

$$F_{hkl} = \sum_{i=l}^{n} f^i . e^{2\pi i(hx_i + ky_i + lz_i)}$$

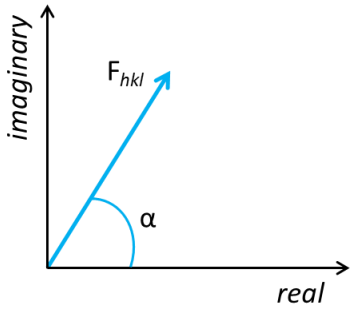


FIGURE I.8 – Structure factor equation and schematic representation (F). F_{hkl} is a mathematical relation of the scattering factor and the positional terms of each atom; is represented as a vector which length is the intensity of the reflection ($F \propto \sqrt{I}$) and α is the phase angle. Adapted from [7]

The choice of wavelengths above and below the absorption edges of the scatterer is possible due to the availability of tuneable beams at synchrotrons, where measurements can be made precisely at a chosen wavelength.[15]

The most common method to obtain the phases from anomalous scattering, and the one used in this thesis, is the replacement of the amino acid methionine for selenomethionine and the overexpression of the protein of interest in a specific system.[19]

Once the initial phasing is complete, the electron density is interpreted by fitting amino acids into it, forming a preliminary model of the protein. This initial protein model is generally very crude and mostly incorrect. It still needs to be refined to be validated as a final model. The refinement methods are very important in crystallography as they allow the improvement of the phases and, consequently, a more precise interpretation of the electron density map.[20]

Refinement, Validation and Publication

After model building, the atomic coordinates are refined in order to fit the experimental diffraction data as best as possible. In order to statistically measure the adjustment between the observed reflection amplitudes and those calculated from the model, crystallographers take into account the R-*factor* (Equation I.4). [21]

$$\boldsymbol{R} = \frac{\sum \left| |\boldsymbol{F_{obs}}| - |\boldsymbol{F_{calc}}| \right|}{\sum |\boldsymbol{F_{obs}}|}$$

EQUATION I.4 – R-*factor* equation. The R-*factor* (or R_{work}) is a statistical measurement that relates the observed reflection amplitudes (F_{obs}) and those calculated from the model (F_{calc}), in order to measure how well the refined data fits to the observed data.

To avoid data bias, a cross-validation scheme is used. This scheme represents the R_{free}, which is calculated with a small portion of the observed reflections that are not used in the refinement, commonly 5 to 10% of the observed data. [21]

Whilst the R factors judge how well the model fits the experimental data, there are other parameters that evaluate the quality of the model. The B-factor, which relates the thermal vibration of an atom in its position, the signal-to-noise ratio and the data completeness are important to crystallographers. [22]

In addition, to validate a newly solved protein structure, the chemistry behind the secondary structure needs to be accurate. For that reason, the Ramachandran plot is a good approach to evaluate the presence of outliers. This bioinformatics tool checks the ψ and φ angles around the C_α of each amino acid, the atom responsible for the protein fold. [23]

The last step is the publication of the protein structure with its atomic coordinates and electron density map, in the protein databank – PDB – and posterior publication in a peer-reviewed journal. [24]

Chapter II

II.1. mRNA localization mechanisms in *Drosophila melanogaster*

"I immediately loved working with flies. They fascinated me, and followed me around in my dreams."

Christiane Nüsslein-Volhard, Nobel Prize, 1995

Drosophila melanogaster

Drosophila melanogaster, also known as the common fruit fly, is an insect that lives in a wide range of habitats. It is a valuable organism for biological research, especially in genetics and developmental biology. It is a small animal with a short lifespan of only two weeks, cheap and easy to breed in captivity. In this sense, *Drosophila* has been used as a model organism for almost a century. In addition, nowadays it's a well understood model since its entire genome has already been sequenced. [25]

Drosophila developmental process

Fertilization is the initial step in the development process. It takes about 10 days since the egg fertilization until the establishment of an adult fly. However, between the events of fertilization and organ differentiation, two important steps occur: gastrulation and cleavage. During these two critical stages, the major axes of the embryo are determined and the cells begin to migrate to their final destinations. [26]

There are two axes that need to be specified: the anterior-posterior (AP) and the dorsal-ventral (DV). The AP axis is responsible for the correct position of the head and tail, and the DV axis is responsible for back and abdomen location (Figure II.1). [27]

The polarity of the fly has its origin in the egg. It is due to two mRNA products of maternal effect genes, the *bicoid* (*bcd*) and *oskar* (*osk*) mRNAs that are placed in opposite regions of the egg creating a gradient (Figure II.2). [2][3]

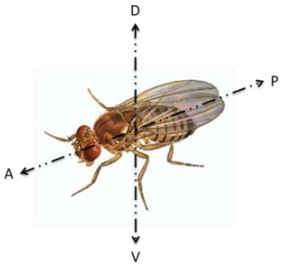


FIGURE II.1 – Representation of the axis that define *Drosophila* body pattern: Anterior-Posterior (A→P) and dorsal-ventral (D→V). Image adapted from *www.insectslimited.com.*

Parallel interaction at the anterior and posterior poles of the egg with several other determinants establishes the patterning of the *Drosophila* embryo. At the anterior end *bicoid* mRNA is translated and produces a protein gradient that is responsible for the outline of the embryo's head and thorax. At the posterior end *oskar* mRNA is translated into Oskar protein, which recruits the abdominal determinant, *nanos* mRNA. [28][29] Despite the fact that several factors required for *oskar* and *bicoid* mRNA localization have been identified, this is still a poorly understood mechanism, whereas the proposed model encompasses the formation of ribonucleoprotein (RNP) particles, transport along the cytoskeletal components and posterior anchoring of the mRNA on its final destination. [2][30][31]

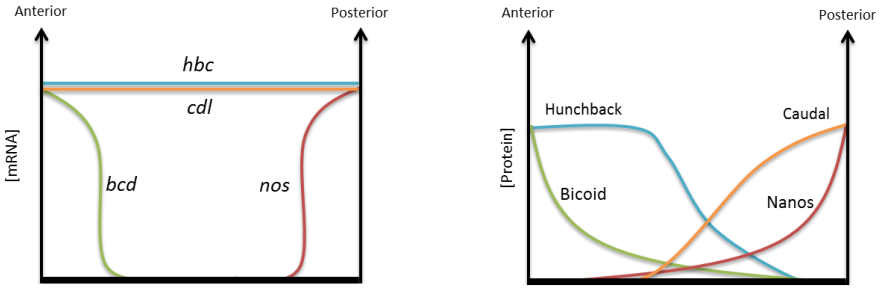


FIGURE II.2 – Maternal effect genes that establish polarity creating a protein gradient throughout the oocyte. The four genes: *hunchback* (*hbc*), *caudal* (*cdl*), *bicoid* (*bcd*) and *nanos* (*nos*) are transferred from the ovarian nurse cells into the oocyte: *bicoid* is trapped in the anterior pole while *nanos* is transported into the posterior pole. When translated, Bicoid protein creates a gradient A→P whilst Nanos protein creates an inverse gradient P→A. Caudal and Hunchback proteins gradient are responsible for the inhibition of the *bcd* and *nos* translation. Adapted from [32]

The general body plan is the same in the embryo and in the adult fly and each segment has its own identity. For instance, the thoracic segment has only legs, the second thoracic segment has legs and wings, and the third thoracic segment has legs and halteres (Figure II.3).

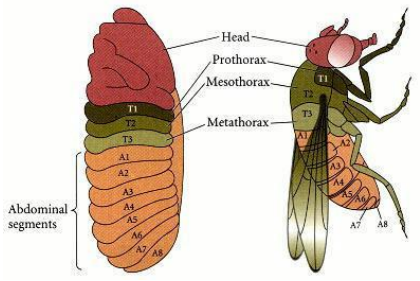


FIGURE II.3 – Body segmentation in *Drosophila*: comparison between larval and adult fly. Image from [32]

mRNA localization mechanisms

mRNA localization and translational control are coupled processes that cooperate to target proteins to specific locations within the cell. Localizing proteins through their mRNA has several advantages; the main advantage is the expression of a targeted protein in restricted areas of the cytoplasm thus preventing it from being present elsewhere. This aspect is extremely important in the case of cytoplasmic determinants which may alter the developing pattern of the embryo if expressed in the wrong place. [33]

In *Drosophila*, mRNA is transported from the nurse cells into the oocyte and, once in the cytoplasm, the highly polarized system of motor transports is responsible for the proper localization of the maternal mRNAs throughout the oocyte. The pathway by which all this transport happens and how the mRNAs reach their destinations is poorly understood even though it is believed that these RNAs are recognized by diverse proteins and use different transport mechanisms. Numerous genetic characterization experiments have been performed but only few biochemical studies of the proteins are available. [29][34][35]

The latest and most extensive genetic study is more than 10 years old and stated undoubtedly that the transport or *bcd* and *osk* mRNA involves a large RNAse-sensitive complex with at least seven proteins. [31]

In this thesis, we aimed to determine the protein structure of two of the proteins identified in this RNP particle. The proteins of interest are Exuperantia (Exu) and Ypsilon-Schachtel (Yps).

Exuperantia and Ypsilon-Schachtel

EXUPERANTIA is a novel protein which is a core component of the protein complex involved in the localization of mRNA within the nurse cells and oocyte, however, all of the genetic studies involving Exu have not determined if it is directly involved in the transport of the RNAs or if it has an indirect role. [35]

Temporal and spatial distribution of the *Drosophila* maternal effect gene - *exu* - shows that Exuperantia is needed for the proper localization of the *bcd* RNA during the formation of oocytes but not for its maintenance. *bcd* mRNA is synthesized in the nurse cells and transported to the anterior pole of the oocyte by a microtubule-dependent mechanism mediated by Exuperantia. [36]

Analysing the distribution of Exu within the developing oocyte it is possible to observe the increase in the concentration from very low levels in the nurse cells of early egg chambers until higher levels at the anterior ends of the oocyte (Figure II.4). These results obtained by the

injection of polyclonal anti-Exu serum allow inferring that Exu is deposited at the anterior pole of the oocyte by the nurse cells and this is followed by diffusion within the oocyte. [36]
Besides its role in the AP polarity of the developing oocyte, *exu* is required for *Drosophila* spermatogenesis encoding sex-specific transcripts. [37][38]

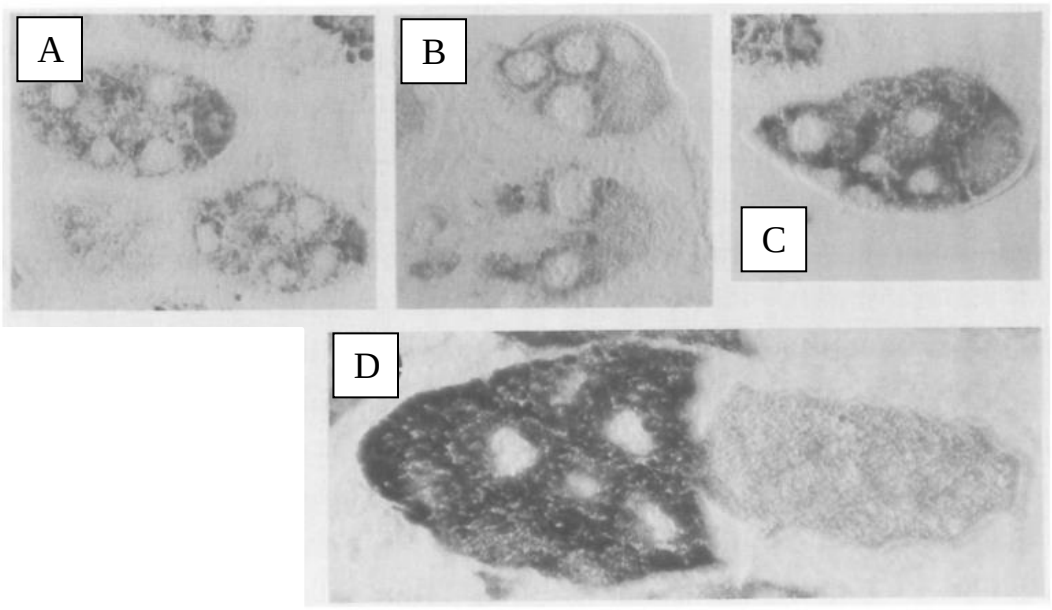


FIGURE II.4 – Distribution of Exuperantia protein in *Drosophila* wild-type egg chambers. A-D: Representation of the developing oocyte where the nurse cells are on the left. Image adapted from [36].

Nowadays, the Exuperantia role in the mRNA transport and localization is still a mystery. It has been proposed that Exu may modify a component that binds *bcd* mRNA or *bcd* itself or it may be directly involved in docking *bcd* message at its site of localization in the developing oocyte. [1][33][36][39]
Further studies of cloning, expression, purification and protein crystallization will allow to obtain the 3D structure of Exuperantia which will be a crucial step in answering some questions concerning the role of Exu in the oocyte.
Figure II.5 represents a basic local alignment showed ~30% sequence homology with three proteins: Integrator complex subunit 1 from *Mus musculus* (INT1_MOUSE) and *Homo sapiens* (INT1_HUMAN) and with Lipid II:glycine glycyltransferase from *S.aureus* (FEMX_STAAB). The multiple sequence alignment analysis shows the existence of 5 conserved residues among prokaryotic and eukaryotic species even though they are not related to any identified conserved domain.

FIGURE II.5 - Basic local alignment of Exuperantia amino acid sequence with its homologous based on NCBI Blast [40], [41] and possible PEST (Proline-Glutamate-Serine-Threonine) sequence based on EMBOSS *epestfind.* [42]

If Exu binds *bcd* mRNA, the recognition may be specific for *bcd* itself and so the lack of significant similarity to previously identified RNA-binding domains. Exu is a basic protein (theoretical pI=9.55) and due to its basic nature the potential interaction is with negatively charged nucleic acids and it remains possible that Exu contains an unknown RNA recognition domain. [36]

Further primary structure analysis has identified a possible PEST sequence located at amino acids 435-450. This specific sequence is rich in proline (P), glutamic acid (E), serine (S) and threonine (T) and has been associated with proteins with a short life. Hence, it's hypothesized that the PEST sequence acts as a signal peptide for protein degradation. [42]

The existence of a PEST sequence within Exu primary structure may explain the high concentrations of Exu at the anterior pole of the oocyte followed by its disappearance in mature oocytes (Exu may be rapidly degraded). The existence of higher levels in the nurse cells could be due to protection from degradation of rapid re-synthesis of protein (high levels of *exu* RNA present in the nurse cells but not in the oocyte). [36]

Additional bioinformatics tools have been used with the objective of obtaining more information about the biophysical and biochemical properties of Exuperantia. The first approach refers to the prediction of the globularity (hence the presence of disordered regions) of the protein using GLOBPLOT (Figure II.6). In this case, the aim was to predict the secondary structure regularity or the lack of it. [26]

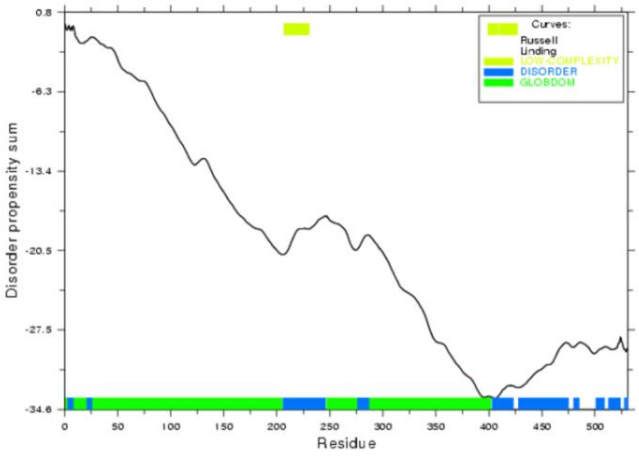


FIGURE II.6 - GLOBPLOT (Intrinsic Protein Disorder, Domain & Globularity Prediction) for Exuperantia. This software (http:// globplot.embl.de) is a web service that allows plotting the tendency within the query protein for order/globularity and disorder. [43]

The DNA fragment as well as the protein sequence were analyzed by several structural bioinformatics tools such as iTASSER - an Internet service for protein structure and function predictions, GlobProt 2.3 – Prediction of intrinsic protein disorder and globularity and XTALPred - a web server for prediction of protein crystalizability. [43][44][45]

Combining the predicted secondary structure by iTASSER, the ordered sequence by GlobProt and the crystallization class by XTALPred several different constructs were designed in order to maximize the probability of obtaining stable protein with preservation of its physiological activity.

YPSILON-SCHACHTEL is part of a protein family known to be responsible for both transcriptional and translational control, the Y-box proteins or Y-box binding proteins. One of the roles that are attributed to Y-box proteins is its location in the cytoplasm and its association with mRNA as components of messenger ribonucleoprotein particles. [46]

In a sequence database search (BLAST) [28][29] with the *yps* product sequence, an RNA-binding motif can be identified: Cold-Shock Domain (CSD) that is widely spread and highly conserved in eukaryotes, prokaryotes, and archaea. [47][48]

According to SCOP (Structural Classification of Proteins database) [49], the cold-shock proteins belong to the family of β–barrel that bind oligonucleotides/oligosaccharides (OB fold) by their nucleic acid binding motifs. Typically, an RNA recognition motif can be recognized as a 90 amino acids long domain containing two conserved sequences - RNP1 and RNP2 (Figure II.7). [50][51]

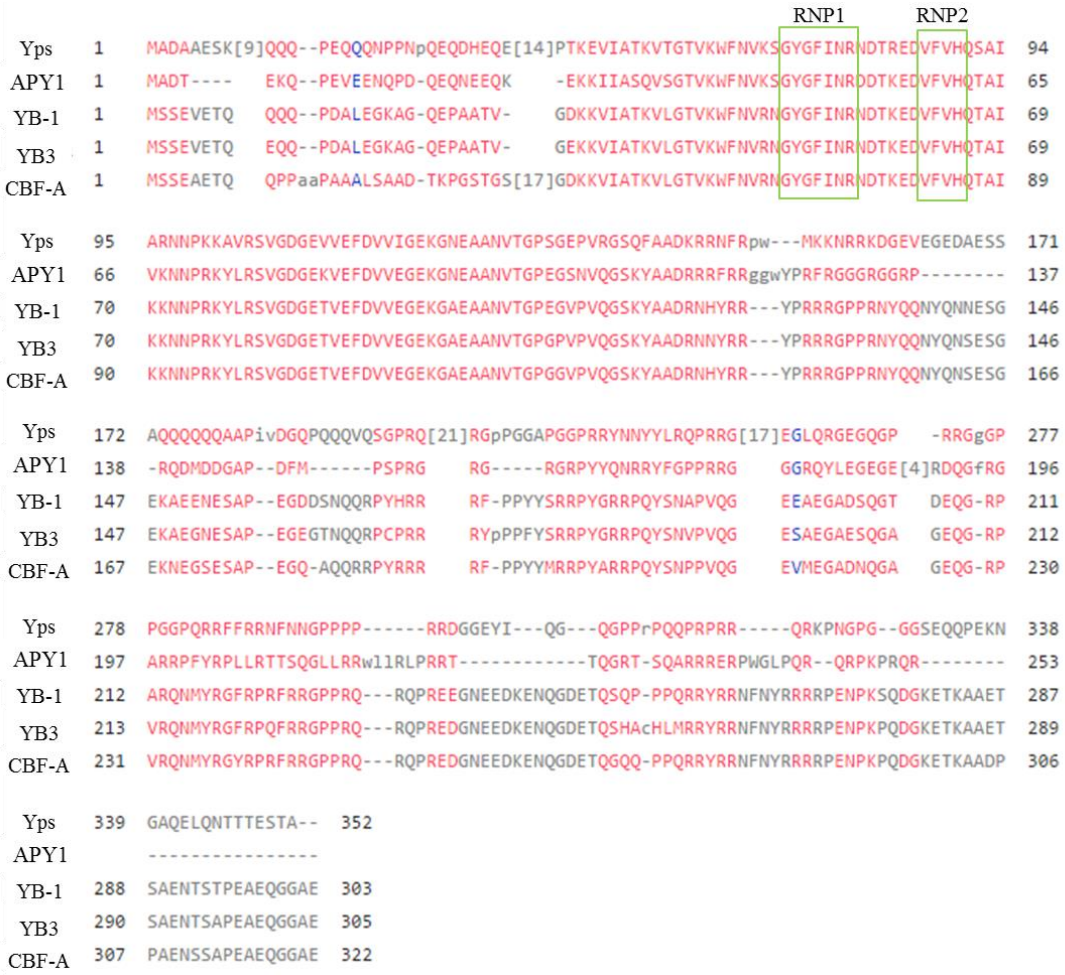


FIGURE II.7 – Basic local alignment of Yps primary sequence with four other proteins that share a high level of homology. In red are the highly conserved residues among species - the cold shock domain. The green rectangle highlights the residues that form the RNA recognition motifs - RNP1 and RNP2. [28][29]

This type of RRM is the most abundant in higher vertebrates and adopts a typical topology of β1-α1-β2-β3-α2-β4. RNP1 and RNP2 are located in the two central β-strands exposing three conserved aromatic residues which are responsible for the RNA binding surface. [51]

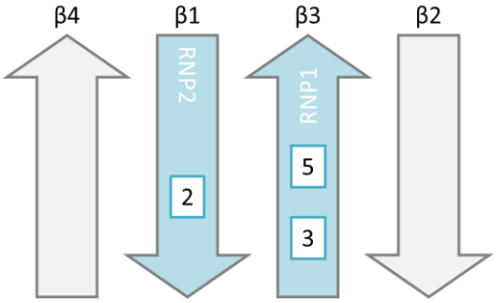


FIGURE II.8 – Schematic representation of the typical topology of the RRM: four-stranded β-sheet, and the conserved RNP1 and RNP2 aromatic residues position. Adapted from [51]

The proposed binding mechanism is by the three conserved aromatic side-chains that accommodate two nucleotides. The bases of the 5' and of the 3' nucleotides stack on the aromatic ring at β1 (position 2 of RNP2) and at β3 (position 5 of RNP1), respectively, while the third aromatic ring located in β3 (position 3 of RNP1) is frequently found between the two sugar rings of the dinucleotide (Figure II.8). [51]

Analysing the primary sequence of YPS and its RNP1 and RNP2 motifs, the aromatic residues that are involved in the RNA binding mechanism are phenylalanine (RNP2), tyrosine and phenylalanine (RNP1) (Figure II.9).

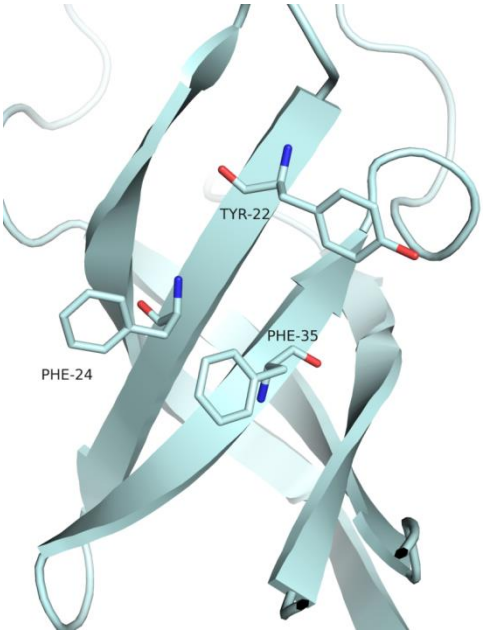


FIGURE II.9 – Representation of the three conserved aromatic residues which are responsible for the RNA binding surface in proteins that hold a cold shock domain. The model used is the cold shock domain from Y-box protein 1 from *Homo sapiens* (PDB entry: 1H95). [46]

Structural studies of these proteins will provide a wealth of information that will help to understand the mechanism involved in mRNA sorting, loading onto the correct carrier, anchoring and translation regulation.

II.2. mRNA localization mechanisms in *Drosophila melanogaster:* experimental procedure

The objective of this project was to determine the 3D structure of both proteins – Exu and Yps – and to further explore the protein-protein interaction.

The first part in this experiment description corresponds to the cloning, expression, purification and crystallization of Ypsilon-Schachtel whereas the second part is related to the cloning, expression, purification and crystallization of Exuperantia.

In addition, results will be presented for both conventional protocols executed at the Macromolecular Crystallography Laboratory, Faculty of Sciences and Technology, *Universidade NOVA de Lisboa*, at the Astbury Centre for Structural Molecular Biology, University of Leeds, and high-throughput protocols at Oxford Protein Production Facility, Research Complex at Harwell, University of Oxford.

Gene sequences, its products and oligonucleotide sequences are detailed on appendices A and B. The obtained results regarding the expression of both genes and its several constructs are summarized on appendix B.

TABLE II.1 – Summary table of the gene sequences selected, vectors used and *E. coli* expression strains in the first part of this experimental work.

Genes	Vectors	Expression strains
• Exuperantia	• pGEX	• *E.coli*
Exu205C	pGEX-6P	BL21
Exu210C	• pET	BL21 Star(DE3)pLysS
Exu399C	pET-Sumo-28a	BL21 Gold(DE3)pLysS
Exu531C	pET-MAL-28b-Prescission	Rosetta2(DE3)pLysS
• Ypsilon-Schachtel	pET-MAL-29b-Prescission	
Yps_CSD	pET-GFP-19b-TEV	
	pET-14b	
	pET-15b	
	pET-28b	
	pET-Duet	

Ypsilon-Schachtel

The cDNA encoding for the Cold Shock Domain (residues 54-132) from *Drosophila melanogaster* was amplified from the *Drosophila* Genomics Resource Center (DGRC) Gold cDNA Collection (LD37574; FlyBase ID: FBcl0179037), by PCR.

Yps_CSD sequence was then cloned into pET-Sumo-28a, pET-MAL-28b-Prescission, pET-MAL-29b-Prescission, pET-GFP-19b-TEV, pET-15b and pET-28b and transformed into *E. coli* BL21 Star(DE3)pLysS, BL21 Gold(DE3)pLysS and BL21 Rosetta2(DE3)pLysS (Table III.1).

Expression of Yps-CDS as a His-tag fusion protein in the pET system produced the best results. The clone that yielded the highest amount of protein was N-His_6-Sumo-Yps_CDS, which expressed a fusion of His-tagged Sumo with Yps in *E.coli* BL21 Star(DE3)pLysS. Expression was induced with 0.4mM IPTG at 18ºC. Cells were resuspended in 20mL of PBS 1x, 0.1% Triton X-100 supplemented with lysozyme (300µg/ml), DNAseI (1µg/ml) and 5mM $MgCl_2$ and kept at -80ºC until cells disruption with an ultrasonic homogenizer UP200S (Hielscher Ultrasonics) in 10 cycles of 30 seconds with a 2mm probe.

The lysate was centrifuged for 30min at 18000x*g*. The supernatant was loaded onto a 5ml HisTrap FF column previously equilibrated with 20mM Tris HCl pH7.9, 500mM NaCl, 10mM imidazole, 3% glycerol (binding buffer). Once the supernatant was loaded onto the column, an extensive (20 column volumes) wash step was performed with wash buffer and the protein eluted in a gradient of 20mM Tris-HCl pH7.9, 500mM NaCl and 500mM imidazole, 3% glycerol (elution buffer) - Figure II.10.

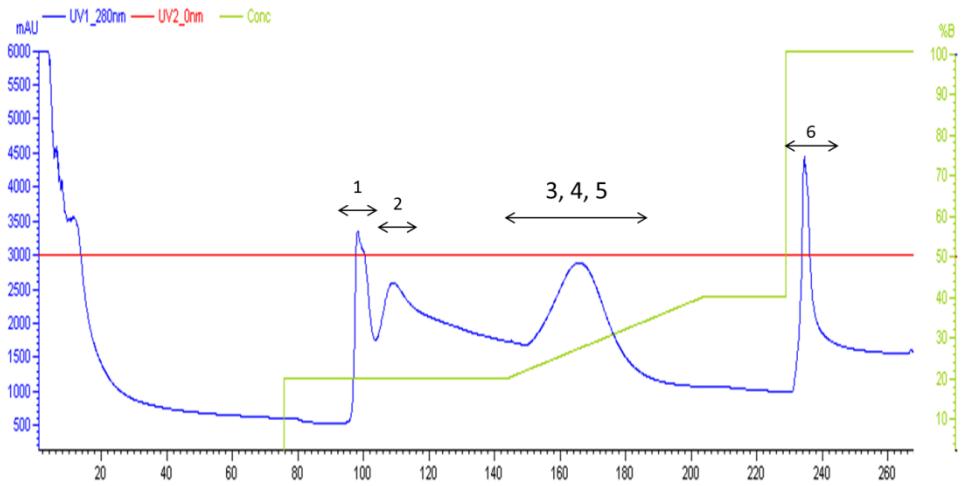


FIGURE II.10 - Typical chromatographic profile of N-His_6-Sumo-Yps_CSD after the IMAC purification step. Injected sample volume of 50ml; Binding buffer: 20mM Tris HCl pH7.9, 500mM NaCl, 10mM imidazole, 3% glycerol. Elution buffer: 20mM Tris-HCl pH7.9, 500mM NaCl and 500mM imidazole, 3% glycerol. The fractions 1 to 6 were aliquoted and analysed by SDS-PAGE. Blue line represents the A_{280nm} and green line represents the elution gradient.

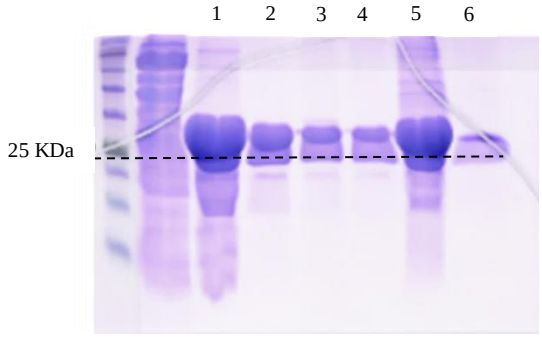


FIGURE II.11 - SDS-PAGE analysis of the purified protein – run at 40V for 30 minutes in Tris buffer and stained with a Coomassie based-solution. Lane 1 – molecular weight marker (kDa) and following lanes have the same nomenclature as the figure II.10 chromatogram. N-His_6-Sumo-Yps_CSD expected molecular weight: 25kDa.

The peaks of the chromatogram that correspond to the purified protein (fractions 1-6 in Figure II.10) were collected as several different fractions which were then analysed on an SDS-PAGE. Fractions 1 to 6 (Figure II.11) were pooled together, incubated with sumo protease (*overnight* at 4ºC) and dialysed against 20mM TrisHCl pH8, 100mM NaCl, 1mM DTT and 3% glycerol. The fusion-tag hydrolysis was followed by a reverse-nickel-affinity chromatography and size exclusion purification on a Superdex 75 10/300 GL (GE Healthcare) and the protein eluted with 20mM TrisHCl pH8, 100mM NaCl, 1mM DTT and 3% glycerol and analysed by SDS-PAGE (Figure II.12).

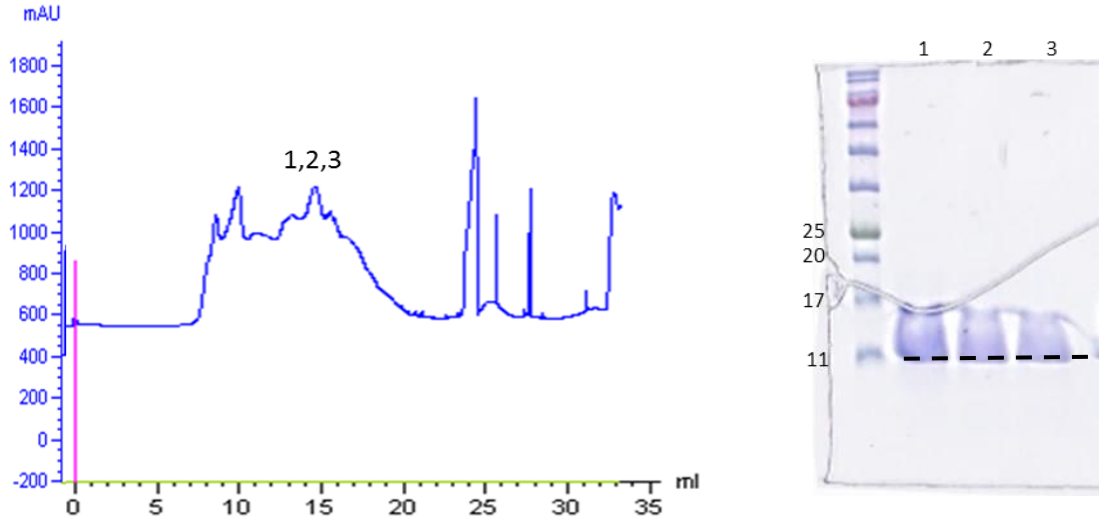


FIGURE II.12 – Left: Typical chromatographic profile of Sumo-Yps_CSD after the SEC purification step. Injected sample volume of 0.5ml; Elution buffer: 20mM Tris HCl pH8, 100mM NaCl, 1mM DTT and 3% glycerol. The fractions 1 to 3 were aliquoted and analysed by SDS-PAGE. Right: SDS-PAGE analysis of the purified protein – run at 40V for 30 minutes in Tris buffer and stained with a Coomassie based-solution. Lane 1 – molecular weight marker (kDa) and following lanes have the same nomenclature as the corresponding chromatogram. Yps_CSD expected molecular weight: 9kDa.

After obtaining a pure sample of Yps-CSD, an aliquot was analysed by differential scanning fluorimetry (DSF) also known as fluorescence-based thermal-shift assay but commonly called thermofluor. [52]

This method is based on the presupposition that folded and unfolded proteins can be distinguished when undergo a thermally induced unfolding in the presence of a hydrophobic fluoroprobe.

The fluorophore – *Sypro Orange* (Molecular Probes) – binds to the exposed hydrophobic residues originating a decrease in the quench. Fluorescence emission is detected and plotted as a function of the temperature. In the process, the protein changes from the folded to the unfolded state and the midpoint temperature (melting temperature) – T_M. This value correlated with the conformational homogeneity of the sample, since it corresponds to the temperature at which the hydrophobic residues become exposed. [52][53][54]

Thermofluor method is often used to screen for optimized buffer conditions by changing the pH, buffer molecules and small-molecule additives. The screening solutions are listed on appendix F and the screen used is based on the published protocol [54].

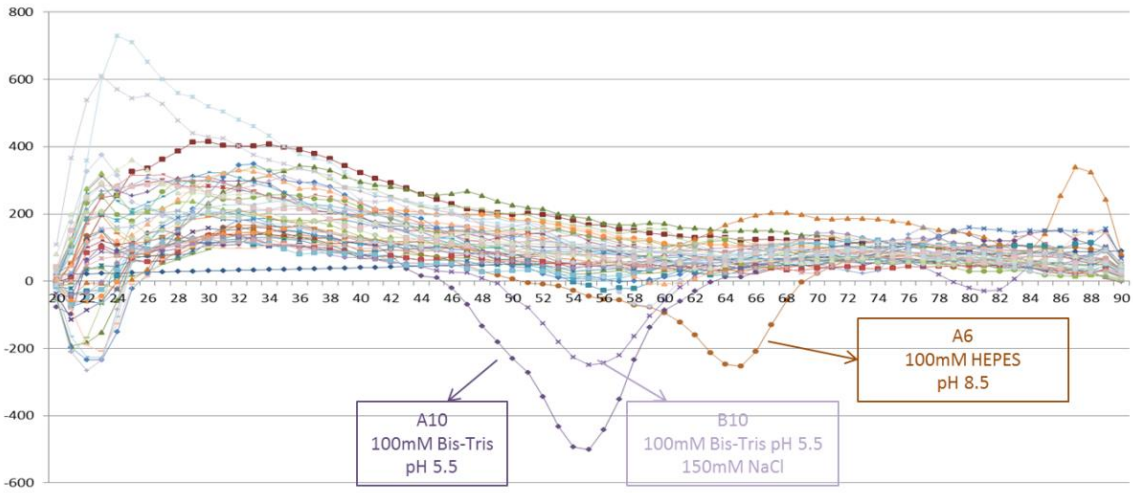


FIGURE II.13 - Thermal shift assay result for Yps_CSD in the 36 buffer conditions listed on appendix F. The more stable the protein is in solution, the more likely it is to produce a crystal. In this sense, the buffers that yield the most promising results were used to solubilize Yps_CSD before proceeding to the crystallization trials.

Several commercially available crystallization screens were used, such as SaltRx (Hampton Research), TACSIMATE (Hampton Research), JBScreen Classic (Jenna BioSciences), Emerald Wizard Screens I + II (Jenna BioSciences), MIDAS (Molecular Dimensions), MD1-01 (Molecular Dimensions), Morpheus (Molecular Dimensions), and 80! (adapted from [55]). Three protein concentrations were tested: 10, 15 and 25mg.ml^{-1} in combination with four

different buffers: B1 - 20mM TrisHCl pH8, 100mM NaCl, 1mM DTT and 3% glycerol, B2 – 100mM BisTris pH5.5, B3 - 100mM BisTris pH5.5, B4 – 100mM HEPES pH8.5.

All the screens were performed on an Oryx8 crystallization robot (Douglas Instrument) using the sitting-drop vapour-diffusion method. 0.4ul of the protein were mixed with the same volume of mother liquor and equilibrated over a 50µl reservoir solution. The 96 well plates were incubated at 20ºC and 4ºC.

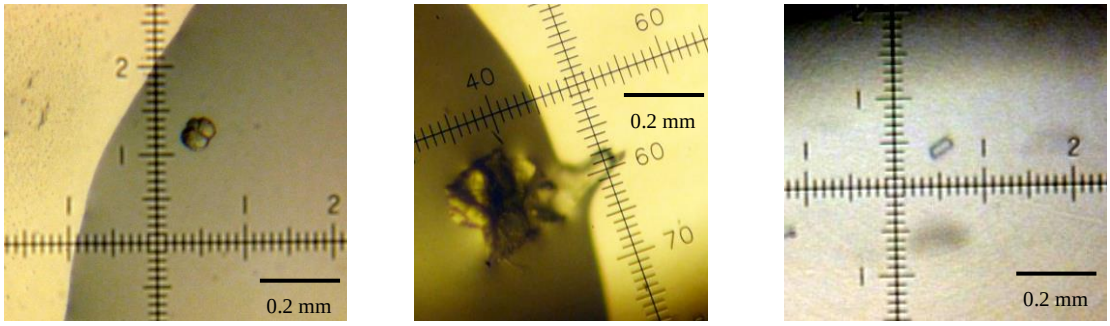


FIGURE II.14 – Possible Yps_CSD crystal obtained for the crystallization trials of Yps_CSD, using $CaCl_2$ (left), PEG 8K (middle) and 12.5% w/v 12.5% PEG 1000, 12.5% w/v PEG 3350, 12.5% v/v MPD (right) as precipitant.

Ypsilon-Schachtel high-throughput cloning, expression, purification and crystallization

Four different constructs were designed and amplified by PCR from the cDNA. Each was cloned into the pOPIN vectors at OPPF (Appendix C) and the protein expression was tested using two expression strains *E. coli* Lemo21(DE3) and *E. coli* Rosetta2 (DE3) according to the protocols of IPTG induction and auto-induction. [56]

From the 4 constructs of *ypsilon-schachtel* gene, the highest expression level was obtained for the construct number 12515, where Yps-CSD was expressed as a fusion protein with an N-terminal GST tag. The purification protocol was the standard OPPF which consists of two steps: immobilized nickel affinity chromatography followed by size exclusion chromatography.

The yield of pure protein obtained with this protocol is lower than the one obtained with the conventional protocol described above.

The HTP protocol was adapted from the optimized protocol described above and the protein was expressed and purified in the ÄKTAxpress Twin system but even in this mode, the result was worse than the obtained previously.

Exuperantia

The cDNA template of *exuperantia* was acquired from *Drosophila* Genomics Resource Center (DGRC) – Gold cDNA collection LD26657 with FlyBase Id FBcl0168744.
Four different constructs have been designed (Exu205C, Exu210C, Exu399C, Exu531C) and cloned into pET vectors (pET-Sumo-28a, pET-MAL-28b-Prescission, pET-MAL-29b-Prescission, pET-GFP-19b-TEV, pET-15b and pET-28b) expressing his-tagged fusion proteins. Resulting expression vectors were transformed into *E. coli* expression strains such as BL21, BL21 Star(DE3)pLysS, BL21 Gold(DE3)pLysS and BL21 Rosetta2(DE3)pLysS (Table II.1).
First of all, the expression levels of the four Exu constructs were tested, using either IPTG-induction or auto-induction [56] in different expression strains, induction temperatures and times. All tests resulted in the protein being recovered in the insoluble fraction, which leads to the possibility of inclusion bodies formation. In this sense, two refolding protocols (Appendix E) were tested but with no satisfactory outcome. Even though, the best result was obtained when Exu210C was expressed as N-His_6-Sumo-Exu210C (pET-Sumo-28a) in *E.coli* BL21 Star(DE3)pLysS and auto-induced overnight at 20ºC. The purification step included an affinity chromatography (IMAC) and some protein was obtained when eluted with 100mM of imidazole. The second step was the size exclusion chromatography, but all sample precipitated right after the elution. Several buffers with variable glycerol concentrations, pHs and ionic strength were tested in order to stabilize the protein but all yielded the same result.
One explanation for this instability might be the existence of 24 *E. coli* forbidden codons. In this sense, an optimized synthetic gene (sExu210C) for *E. coli* expression was ordered from NZYTech, Lisbon, Portugal.

sExu210C was cloned into pET-Sumo-28a, and the conditions described above were repeated, but the same result was obtained, low amount of soluble protein and unstable when in solution. Due to its instability, the next approach was to express as a His_6-sExu201C to decrease the downstream handling. The purification protocol involved two chromatographic steps: affinity followed by an ionic exchange. By the end of the second purification step, we were able to obtain pure soluble protein in a yield of 0.5mg from 8L of cell culture, concentrated using Amicon Ultra Centrifugal Filters (EMDMilipore) to 10mg.ml^{-1}. Cyclically following this expression and purification protocol, we were able to proceed to crystallization trials using a crystallization robot (Oryx8 from Douglas Instruments) in order to maximize the number of crystallization conditions using low quantities of protein. The drops were 0.8µl in a 1:1 ratio in a sitting-drop, vapour-diffusion method. From all the conditions tested for His_6-sExu210C, three of them resulted in small crystals, represented on Figure II.15.

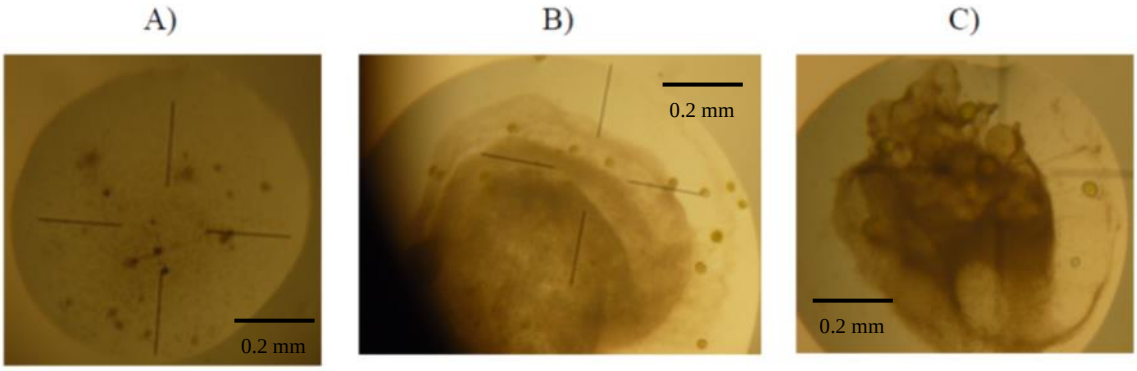


FIGURE II.15 - Possible Exuperantia protein crystals, with dimensions 0.04x0.04x0.04mm^3, obtained in 24 hours by the method of vapour diffusion in sitting drop using A,B) PEG 8K and C) sodium/potassium phosphate solution as precipitant.

Due to their small size further optimization was necessary in order to increase the crystal size so it could be harvested and analysed by X-ray crystallography. To optimize the crystals obtained, several factors were changed, such as the drop size, protein:well solution ratio, protein concentration, incubation temperatures - but no crystals grew in the conditions tested.

Exuperantia high-throughput cloning, expression, purification and crystallization

The next approach was to follow a high-throughput protocol for cloning, expression, purification and crystallization. For this, the cDNA was cloned into the pOPIN vectors at OPPF (Appendix C) and the protein expression was tested using two expression strains *E. coli* Lemo21(DE3) and *E. coli* Rosetta2(DE3) according to the protocols of IPTG induction and auto-induction. [56]

From the 52 constructs of *exuperantia* gene, the highest expression level was obtained for the construct numbers 12498 and 12488, where Exu2.210 and Exu2.193 (corresponding to the N-terminal domain of the protein as predicted by GLOBPLOT (Figure II.6) are expressed as fusion proteins with N-terminal TF tag. The purification protocol was the standard OPPF which has two steps: immobilized nickel affinity chromatography followed by size exclusion chromatography (Figures II.16).

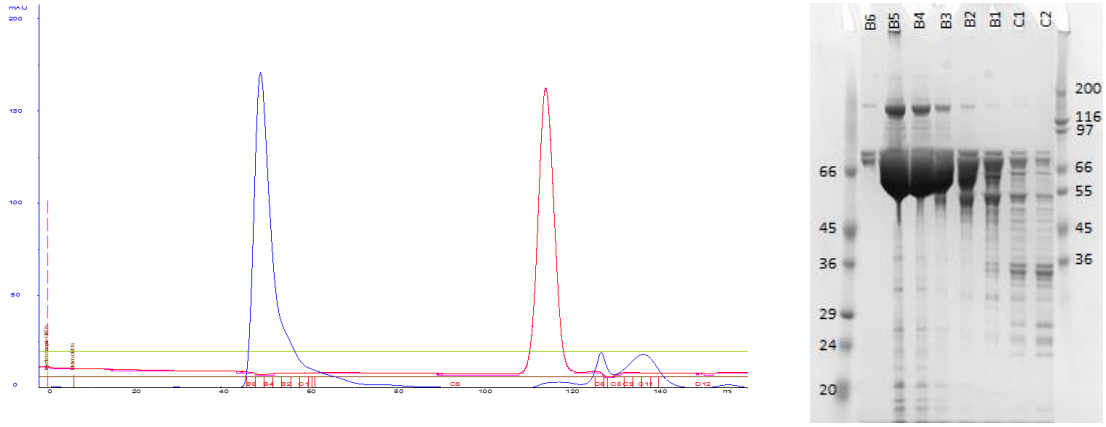


FIGURE II.16 – On the left: Typical chromatographic profile of N-TF-Exu2.210 after the second purification step, the size exclusion chromatography. Injected sample volume of 5ml; Buffer: 20mM Tris-HCl pH7.5, 200mM NaCl, 1mM TCEP. Superdex 200 HiLoad 16/60 column (GE Healthcare). The fractions B6, B5, B4, B3, B2, B1, C1 and C2 were aliquoted and analysed by SDS-PAGE. The blue line corresponds to the A_{280nm} and the red line corresponds to the sample conductivity. **On the right:** HT-SDS-PAGE analysis of the purified protein – Invitrogen NuPAGE Gels – run at 200V for 40 minutes in BisTris buffer and stained with a Coomassie based-solution. Lane 1 – molecular weight marker (kDa) and following lanes have the same nomenclature as the chromatogram. N-TF-Exu2.210 expected molecular weight: 73kDa.

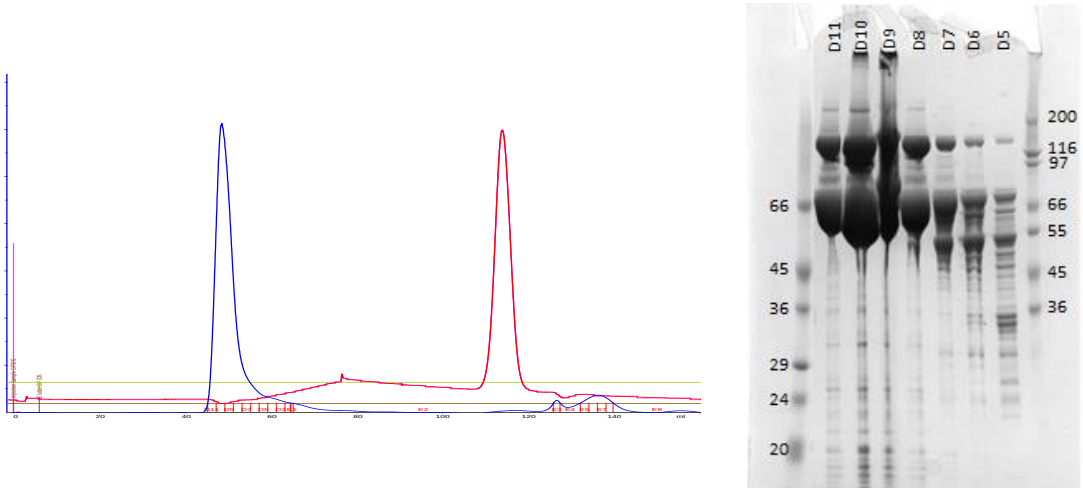


FIGURE II.17 – On the left: Typical chromatographic profile of N-TF-Exu2.193 after the second purification step, the size exclusion chromatography. Injected sample volume of 5ml; Buffer: 20mM Tris-HCl pH7.5, 200mM NaCl, 1mM TCEP. Superdex 200 HiLoad 16/60 column (GE Healthcare). The fractions D11, D10, D9, D8, D7, D6 and D5 were aliquoted and analysed by SDS-PAGE. The blue line corresponds to the A280nm and the red line corresponds to the sample conductivity. **On the right:** HT-SDS-PAGE analysis of the purified protein – Invitrogen NuPAGE Gels – run at 200V for 40 minutes in BisTris buffer and stained with a Coomassie based-solution. Lane 1 – molecular weight marker (kDa) and following lanes have the same nomenclature as the chromatogram. N-TF-Exu2.193 expected molecular weight: 72kDa.

For each sample the yield achieved was 8mg and 15mg per 1L of cell culture, respectively. Fractions B3 and B4 (Exu2.120) and D11 (Exu2.193) were sent directly to crystallization (Table II.2) while the rest of the protein sample was incubated with protease to cleave the TF-tag.

The protein concentration was chosen according to the PCT™ Pre-Crystallization Test from Hampton Research. Several concentrations were tested although the most appropriate sample concentration was between 22 and 28mg.ml^{-1}. Thus, 25mg.ml^{-1} was chosen.

The other half of the sample that was incubated overnight with 3C protease to cleave the His_6-TF tag, precipitated so it didn't follow to crystallization.

TABLE II.2 – List of the 96-well Greiner plates' barcodes, the commercial crystallization screens, concentrations and temperatures trialled in order to crystalize the tagged protein sample: N-TF-Exu2.210 and N-TF-Exu2.193.

Plate Barcode	Commercial Screen	Protein sample and concentration	Temperature (ºC)
441305013768	JCSG+	N-TF-Exu2.210 25mg.ml^{-1}	21
441305013775			6.5
441305013782	Index		21
441305013591			6.5
441305013607	Morpheus		21
441305013614			6.5
441305013645	PACTpremier		21
441305013652			6.5
441305013669	PACTpremier	N-TF-Exu2.193 25mg.ml^{-1}	21
441305013676			6.5
441305013683	Morpheus		21
441305013898			6.5
441305013904	JCSG+		21
441305013911			6.5
441305013928	Index		21
441305013935			6.5

Over 1500 conditions were tested without success. No crystals or conditions that could suggest improvement were observed.

Exu-Yps complex

Exu revealed to be an unstable protein *in vitro* while Yps_CSD, after the optimization, is quite easy to obtain pure in solution. Since one of the aims of this experimental work is to obtain the three-dimensional structures of both proteins independently as well as in complex, the next stage was to co-express Yps and Exu in an attempt to stabilize Exu.

The strategy was to clone both genes in pET-Duet-1, a plasmid for bacterial expression, designed for the co-expression of two target genes. The vector contains two multiple cloning sites (MC1 and MC2). The cloning in pET-Duet-1 has two scenarios: **i)** Yps-CDS in MCS1 and sExu210C in MCS2 and **ii)** sExu210C in MCS1 and Yps_CSD in MCS2.

Both products, N-His_6-Yps_CSD-sExu210C and N-His_6-sExu210C-Yps_CSD were expressed in a small scale but there was no observable overexpression of the fusion protein, in any of the *E. coli* expression strains using either of the two induction methods.

II.3. mRNA localization mechanisms in *Drosophila melanogaster:* results and discussion

In structural biology, depending on the aim of the experiment, there are some aspects that need to be carefully analysed before heading to the wet lab and start the cloning, expression and purification process.

The aim of this project was to structurally analyse two proteins that were proven to be involved in the mRNA localization mechanism in the oocyte of *Drosophila melanogaster* [36] and, therefore, it was important to select the best possible sequence to clone and express.
First, the protein sequence was analysed to predict which the class of the protein of interest is: globular, non-globular, membrane or multi-domain protein.
To do so, several bioinformatics tools available that predict the potential globular domains were used. [43] Exuperantia primary sequence was analysed for globular domains and two possible regions of interest were identified, from residues 1 to 206 and 248 to 403 (Figure II.18), separated by a section of unstructured polypeptide chain.

```
1   MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPT   60
61  DHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLE  120
121 QLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIG  180
181 DANIKNYSLRKLSKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNG  240
241 DDKQSSKNAVQGKRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPF  300
301 AKLVVSDVLELDIQIENLERQNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLS  360
361 AIWAEKNIEGLDIALQSIGRLKSKDKAELLELLDSYFDPKKTTVKPVVKGNSNNNNNYRR  420
421 RNRRGGRQSVKDARPSSSPSASTEFGAGGDKSRSVSSLPDSTTKTPSPNKPRMHRKRNSR  480
481 QSLGATPNGLKVAAEISSSGVSELNNSAPPAVTISPVVAQPSPTPVAITA             532
```

FIGURE II.18 – Analysis of the globular and non-globular domains of Exuperantia by Globplot2. [43] The residues in bold correspond to the globular domains of the protein: 1-206 and 248-403 and secondary structure prediction by iTASSER. Helices highlighted in red, strands highlighted in blue and coils highlighted in black. [44]

With this prediction, it is possible to assume that Exu is a multi-domain protein, where well-ordered regions are mixed with unstructured linkers that are usually not important for the protein's function. The presence of random coiled areas can undermine all the crystallization

attempts. For this reason, sometimes it's necessary to solve individual domains separately and therefore to set correct domain boundaries.
Crossing data from the globular domain prediction with the secondary structure prediction, the constructs were defined according to the structured areas of the protein. In this case, from the residues 1-205, 1-210, 1-399, 1-431 as well as the full length.

The same bioinformatics analysis was followed for Ypsilon-Schachtel. First, the Yps primary sequence was analysed for the presence of globular domains, using the GlobPlot web server and one globular domain was identified from the residues 36-123. And within this globular domain, a specific cold shock protein domain was identified from residues 63 to 131 (Figure II.19).
The basic local alignment search tool (BLAST) for Yps complete primary sequence found, as expected, a conserved domain that is found in eukaryotes, prokaryotes, and archaea – cold shock domain (CSD).

1 MADAAESKPLAAEQQQAQQQPEQQQNPPNPQEQDHEQEPLDELQGQQGQPAPPTKEVIAT 60

61 KV**TGTVKWFNVKSGYGFINRNDTREDVFVHQSAIARNNPKKAVRSVGDGEVVEFDVVIGE** 120

121 **KGNEAANVTGP**SGEPVRGSQFAADKRRNFRPWMKKNRRKDGEVEGEDAESSAQQQQQQAA 180

181 PIVDGQPQQQVQSGPRQPRQNFRRGPPGGPPGGPRGGPRGPPGGAPGGPRRYNNYYLRQP 240

241 RRGLGGGDGSAEPGVHDQNPEGLQRGEGQGPRRGGGPPGGPQRRFFRRNFNNGPPPPRRD 300

301 GGEYIQGQGPPRPQQPRPRRQRKPNGPGGGSEQQPEKNGAQELQNTTTESTA 352

FIGURE II.19 - Analysis of the globular and non-globular domains of Ypsilon-Schachtel by GlobPlot**.** [43] The residues in bold correspond to the Cold Shock Domain: residues 63-131

According to these results, the sequence chosen to be cloned and expressed is the one that corresponds to the cold shock domain including the neighbouring residues to ensure that all relevant residues are present – 55 to 171.
After defining the regions to clone and express, the next step is the choice of the expression system. To perform large scale protein production the most used systems are *E. coli*, yeast, baculovirus infected insect cells and mammalian cells.
E. coli is the cheapest and the easiest expression system and to do so there are many expression vectors available with different N- and C-terminal tags as well as many different strains. When there is no indication otherwise, this is the first system to be tested in order to express the protein of interest.
In this project several *E. coli* constructs were designed and expressed. The poor results obtained for Exu, in which the low level of protein obtained was insoluble or inexistent, may be due to

the fact that we are trying to express a eukaryotic protein in *E. coli*. In this case, there are refolding protocols, like the ones that were used to refold Exu, where it is possible to solubilize the protein from the inclusion bodies. The refolding protocol involves the use of high concentrations of denaturants, with a reducing agent (DTT). The solubilized (denatured) protein is then refolded by slowly decreasing the denaturant concentration along with an oxidizing agent. This is a very expensive and time-consuming protocol and in the end, the amount of protein retrieved from the refolding protocol is rarely enough to follow through. [57]

Another option is to investigate the protein overexpression with different strains, cell growth media and incubation temperatures.

Exu was cloned into plasmids with *T7lac* promoter. The resulting vectors were transformed into *E. coli* and the expression was induced by the addition of IPTG ranging from 0.05 - 2.0 mM. [58] The expression was also tested using auto-induction as well. [56] The best results were obtained with auto induction, although the yield was still very low. The advantage of auto induction over IPTG induction is that it allows growth and induction of recombinant proteins with low input from the experimenter; there is no need to monitor cell density or add an inducer. With a low amount of protein being produced, several strategies were used to recover a larger quantity of soluble Exu. These strategies involved decreasing the growth temperatures and times of induction are described on Table II.3.

TABLE II.3 – Growth temperatures and times tested for Exu expression in *E.coli*.

Incubation temperature	Incubation time
18°C	*overnight*
20°C	*overnight*
25°C	*overnight*
30°C	5-6 h
37°C	3-4 h

The best results were obtained at 18ºC. All these results suggest that Exu is toxic to *E.coli* (unpredictable levels of expression, low cell culture density, better results at lower temperatures and in auto-induction media).

To overcome this protein toxicity issue, two strategies were applied and they both involved the modification of the auto induction media composition: **i)** decrease the lactose quantity; **ii)** no lactose at all. *E.coli* primarily uses glucose as carbon source. Glucose serves as an inhibitor of the *lac* operon – in its presence, *E. coli* is not able to use lactose as a carbon source, with the added advantage that any genes under the control of the *lac* operator will not be expressed while

glucose is present. This allows the cells to grow for longer without expressing any cloned proteins. As the glucose is depleted, overexpression of the target protein can start. As the cell density is high, the yield of the protein consequently increases, even though the cells stop growing due to its toxicity. In the case of Exu expression, even when no lactose was added to the growth media, there was no protein expression, besides the fact that the measured OD in the first 6h of cell growth was slightly higher when compared to the growth curve in the presence of the full quantity of lactose (6mM) and 3mM.

Proteins have been empirically classified due to their toxicity on a scale of 1 to 6 [54], where 1 is the most toxic and 6 the non-toxic. According to this classification, Exu is a class 2 toxic protein. To overcome its toxicity in *E. coli* expression, the plan is to use detoxification cell strains, such as DetoxE^{TM} Competent Cell (Expression Technologies Inc.), high density growth media and regulated vectors with multiple repressor binding sites. (See Section II.4 for Future Work Plan)

Yps is classified as a non-toxic protein, which is supported by the data presented in this thesis.

Once the expression and purification steps were optimized and a pure protein sample was obtained, the next stage is the crystallization. Crystallization is the main bottleneck in X-ray crystallography.

The first approach in order to obtain Yps_CSD crystals was to screen the maximum number of conditions using the minimum standard protein concentration (10mg.ml^{-1}) in a 1:1 ratio of protein-precipitant. To do so, we used a crystallization robot for sitting drops and large scale crystallization plates for hanging drops. The basic theory is to create a supersaturation state and as a result, initiate the nucleation process (Figure I.3).

In this initial screen no crystals were obtained. From the careful examination of the drops, we were able to redraw the strategy to crystallize Yps_CSD.

The drops were examined with a stereomicroscope immediately after setup, each day for the first week, and once a week for several weeks. Due to storage issues, plates over 1 year have been discarded. The conditions that had clear drops or precipitate have been identified.

In this initial screen, the majority were clear drops (>95%) that persisted for several weeks. Clear drops indicate the need to increase protein or precipitant concentration. A second batch was performed with a higher protein concentration (15mg.ml^{-1}). The same drop observation protocol was used and, in this case, 65% of the drops remained clear and in the remaining drops was visible a light precipitate. Since more than half of the drops still remained clear, the next batch of crystallization experiments was performed with Yps_CSD at 25mg.ml^{-1}. In this case, several scenarios took place: clear drops, light precipitate, dark precipitate and the most interesting one, gelatinous protein precipitate. To improve this condition a gradient of precipitant concentration and pH was made and drops set up with the same Yps_CSD

concentration, drop ratio and temperature. In this case, a crystal was generated (Figure II.14) although its small size and morphology need further optimization.

Further crystallization experiments were performed, such as directly mixing the macromolecule with excess precipitant, change the plate's incubation temperatures, increase the salt concentration, change the pH, and further concentrate the protein beyond 25mg.ml^{-1}. We were not able to concentrate Yps_CSD above 25mg.ml^{-1} in the 20mM Tris-HCl pH7.9, 500mM NaCl and 500mM imidazole, 3% glycerol buffer as it started precipitating. The same outcome happened when the protein buffer was replaced for those suggested by the thermofluor assay.

So far, the only possible hits obtained for Yps_CSD are those on Figure II.14 that grew after a 2 month incubation period at 4ºC. These crystals have not been tested for diffraction due to their small size.

II.4. mRNA localization mechanisms in *Drosophila melanogaster*: conclusions and future work

Conclusions

The specific mRNA localization mechanism in cells enables a tight regulation of protein expression spatially and temporally. In *Drosophila melanogaster* oocyte such mechanisms are crucial in the proper embryonic development and in the formation of the body patterns in the adult fly. [3]

This project intended to structurally study two proteins involved in the mRNA localization mechanism of *bicoid* mRNA in *Drosophila* oocyte: Exuperantia and Ypsilon-Schachtel. To accomplish this we combined Molecular Biology techniques which involved the cloning of the target genes in plasmids of the pET expression systems, which add histidine tails to the proteins facilitating the purification process. After the proper expression of proteins in *E.coli*, biochemical procedures, such as affinity chromatography and ionic-exchange chromatography, were performed in order to purify the protein.

We were able to obtain several clones of the Exu protein although the major problems occurred in the expression process with Exuperantia proving to be a class 2 toxic protein, with no detectable expression. In the expression system using *E. coli* we weren't able to obtain enough amount of any of the cloned constructs to pursue crystallization experiments.

A highly conserved domain was identified in Yps: Cold Shock Domain. The sequence corresponding to this domain was successfully cloned and expressed in *E.coli*. The purification protocol was optimized and crystallization experiments were followed. Only three crystals were generated after an incubation period of 2 months but due to their size and morphology we weren't able to perform any diffraction experiments.

For the time being, this project has been suspended but we aim to return to it as soon as possible. At this point, crucial information has been obtained and has allowed redrawing the experimental approach.

I believe this work may provide a further step in the understanding of the embryonic development of *Drosophila melanogaster*, providing ultimately a better comprehension of the mechanisms of eukaryotic cells differentiation.

Future work plan

The future work plan for this project can be divided into 4 main tasks.

1. Exuperantia expression, purification and crystallization
2. Ypsilon-Schachtel crystallization and crystal optimization
3. Complex Yps-Exu expression, purification and crystallization
4. Structural and biochemical characterization

The aim of this project was to characterise these proteins using X-ray crystallography. To do so, was mandatory to obtain a soluble and pure protein solution in order to proceed to crystallization experiments. As described in the previous chapter, Exu has been a tricky protein to express. This difficulty is thought to be related to its toxicity. In biological systems, proteins are expressed with a specific function, meaning that they have a specific location within the cell or tissue, for a specific period of time and at a specific amount. When a fusion protein is generated and overexpressed in a different biological system, there is always a possibility that the recombinant protein may interfere with the host cellular machinery causing, in the worst case scenario, cell death.

At this point there are at least two direct methods to follow. The first one involves optimization of the expression in *E. coli*. To do so, we can use detoxification cell strains. These *E.coli* strains are genetically engineered to over-express the *lacI* repressors. However, the best approach is to change the expression system to a eukaryotic expression system like yeast, baculovirus infected insect cells or mammalian cells. Comparison of these three systems and taking into account the cell growth rate, the growth media complexity, the expression level and since Exu is a eukaryotic protein, post-translational modifications are very important; the best option is to express Exu in yeast. [59]

Yeast expression system has a rapid cell growth, a low media complexity, generates a high level of protein expression, and provides for almost all of the post translational modifications. This aspect is fundamental for the proper function of Exu as it has two identified phosphorylation sites and this post-translational modification is proven to be required for the correct localization of *bcd* mRNA to the posterior pole of the oocyte. [28]

In the case that the expression in yeast does not generate satisfactory results, the following approach would be to express Exu in insect cells, for instance, in Schneider cells. These types of cells have been isolated from late-stage *Drosophila melanogaster* embryos and provide a good system to study *Drosophila* cell biology. [60][61] Exu has been cloned into pOPIN vectors,

which allow expression of proteins from multiple hosts, in this sense, there are several Exu constructs ready to follow expression in yeast or insect cells.
Once the expression issue is overcome, the next step is the protein purification and crystallization. Whereas the expression is done in yeast cells or in insect cells, the protein is secreted to the media. The purification follows the same pattern of chromatographic methods [62], with the aim of obtaining soluble, stable and homogeneous protein solution to follow to crystallization experiments. In order to obtain high quality diffracting crystals, we need to find the condition where a thermodynamically stable nucleus is formed leading to the association of molecules in the three dimensions, forming a crystal. The process of obtaining a crystal and all its implications has been described in detail on chapter I.

At the time being, Yps has been cloned, expressed and purified. The hold-up in attaining the three-dimensional structure of this protein, and specifically the Cold Shock Domain (Yps_CSD) which is the biologically relevant domain, is the production of suitable protein crystals. In order to fine-tune hit conditions and to grow larger or better diffracting crystals, there are several strategies: experiment different protein and precipitant concentrations, buffer composition, incubation temperature or droplet size. If none of these strategies produce better crystals, the other option is to go back to the protein and rework its environment such as by the addition of possible ligands, use mutant constructs and to use protein from a different organism. In Yps_CSD case, it is a good option to overexpress the protein in insect cells as proposed for Exu. Yps is a eukaryotic protein and even though it has a high expression level in *E. coli* it may not be with the proper folding or lacking post-translational modifications, creating an unstable sample and thus becoming difficult its crystallization. [63]

The protein-protein complex Exu-Yps will follow the same approach as Exu by itself, changing the expression system from *E. coli* to yeast or insect cells. The characterization of protein-protein interactions involves kinetics and binding specificity and affinity: search for binding partners and/or inhibitors (which in the case of Exu-Yps are ribonucleic acids), confirm activity after protein purification; which is the rate of the complex formation and dissociation and the ligand binding affinity. In parallel, by knowing the three-dimensional structure of this protein-protein interaction and its interaction with the substrate, it is possible to suggest mechanisms for the mRNA transport along the cell, how Exu and Yps interact with each other, how they interact with the nucleic acids, which residues are responsible for this interaction, and in the future introduce the other proteins that form the RNP where Exu and Yps have been identified, to finally understand how *bicoid* mRNA is transported and localized in the posterior pole of the oocyte and which protein is the direct responsible for the protein-mRNA binding.

Chapter III

III.1. Amidation of *S. aureus* peptidoglycan residues: general introduction

"What you see is that the most outstanding feature of life's history is a constant domination by bacteria."

Stephen Jay Gould,
American paleontologist, evolutionary biologist and historian of science,
1941-2002

Bacterial Cell Wall

The cell wall is the essential structure responsible for the stress-bearing and shape-maintenance in bacteria and is composed of unique components that turn this structure into the most important site for antibiotics attack. Despite its structural role, the bacterial cell wall is involved in a wide range of biological processes, like growth and division, interaction between bacteria and the environment and pathogenesis. The bacterial cell wall also provides immunological variation among bacteria strains. [64] [65] [66]
Cell wall containing bacteria are classified in two groups, according to their cell envelope structure: Gram-negative and Gram-positive. [67]

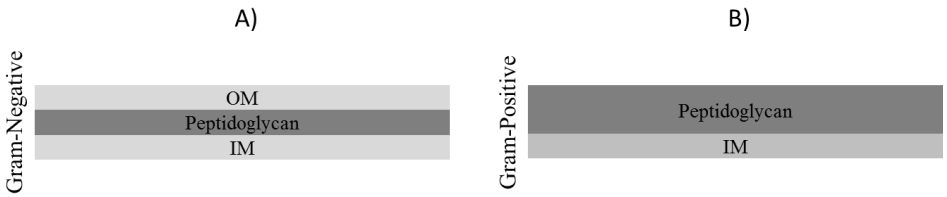


FIGURE III.1 - Representation of the structural differences in Gram-positive and Gram-negative cell walls. A) The Gram-negative cell wall has three layers: the outer membrane (OM), the peptidoglycan and the cytoplasmic or inner membrane (IM). B) The Gram-positive cell wall has no OM and the peptidoglycan layer is thicker.

The Gram-negative cell envelope

In the Gram-negative bacteria, the cell wall is composed of a thin layer of peptidoglycan, 7-8nm thick, surrounded by an outer membrane that contains a negatively charged component, lipopolysaccharide (LPS or endotoxin) and the cytoplasmic or inner membrane (Figure III.1, left). [67]

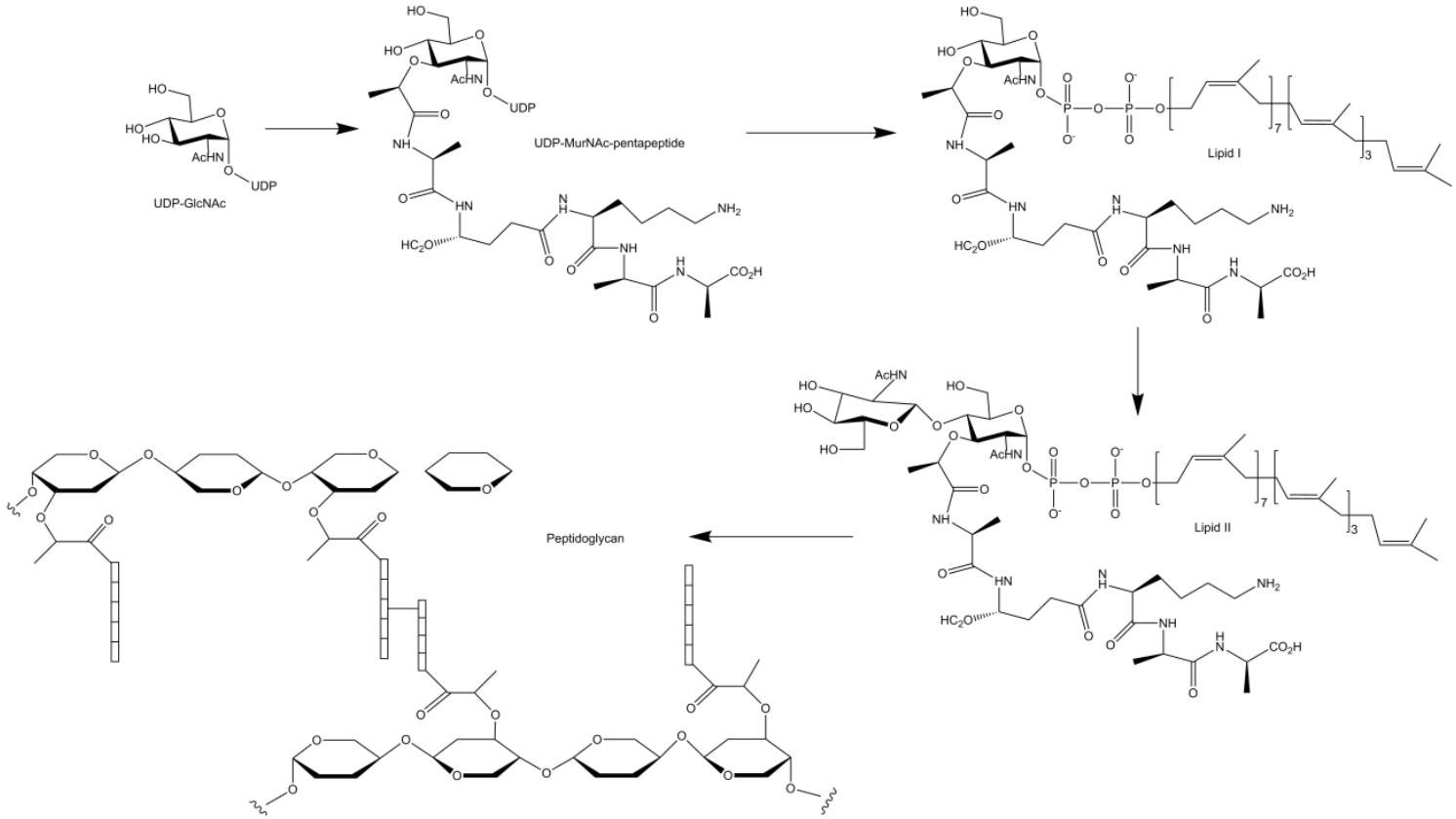
The Gram-positive cell envelope

In the Gram-positive bacteria, the outer membrane is absent (Figure III.1, right). To resist the turgor pressure on the plasma membrane, the surrounding layer of peptidoglycan is thicker (20-80nm) than the Gram-negative. [67]

The composition and structure of the peptidoglycan seem to be rather constant among Gram-negatives, but there is great variation among Gram-positives bacteria.
In both bacterial groups, the peptidoglycan layer functions as a platform for anionic polymers called teichoic acids. In addition to teichoic acids, in Gram-positive microorganisms, the surface is populated with proteins that can vary depending on the growth conditions if adaptation to the environment is necessary. [68]

The Biochemistry of Peptidoglycan

Peptidoglycan is a macromolecule composed of polysaccharide strands cross-linked with peptides. Its biosynthetic pathway is complex, with consecutive enzymatic steps occurring in the cytoplasm and on the inner and outer membrane surfaces and can be grouped into three phases. In the first stage, the synthesis of the alternating units of N-acetyl-glucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) precursors occurs. [69] [5] [70] [71]
The second step concerns the synthesis of the lipid intermediate precursor and its transfer to the membrane acceptor, resulting in a structure called lipid I (undecaprenylpyrophosphoryl-MurNAc-pentapeptide). Then the GlcNAc is transferred to the pentapeptide resulting in lipid II, the final intermediate in this biosynthetic cascade. [70]
The last step is when the lipid II undergoes translocation, transglycosylation and transpeptidation to become a single macromolecule of multiple layers of β-1,4-linked carbohydrate polymers cross-linked with peptide chains. This final structure is the peptidoglycan (Scheme V.1). [71]



SCHEME III.I - Schematic representation of the peptidoglycan synthesis pathway. The peptidoglycan synthetic pathway can be grouped in three phases: in the first stage occurs the synthesis of the precursors of the alternating units UDP-GlcNAc and UDP-MurNAc; in the second stage occurs the synthesis of the precursor lipid intermediate – lipid I – and its transfer to the membrane acceptor to form lipid II. The last step is the translocation, transglycosylation and transpeptidation of lipid II to become the mature molecule of peptidoglycan. Adapted from [72] Image created using ChemDraw®. [73]

Glycan strand and peptide moiety

The glycan strands are composed of alternating units of N-acetyl-glucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) linked by β-1,4-glycosidic bonds, catalysed by transglycosylases.

The pentapeptide is bound to the glycan strand through the N-terminus lactyl group of MurNAc and contains alternating L- and D- residues, which the presence of D- amino acids is a characteristic feature of the peptidoglycan and is thought to help protect against attacks by most peptidases (Scheme III.1).

The sequence L-Alanine (L-Ala), D-glutamic acid (D-*iso*-Glu), meso-diaminopimelic acid (*meso*-A_2pm), D-Alanine and D-Alanine is the typical composition, although it is variable according to the organism. In the case of *S. aureus* and other Gram-positive bacteria, the pentapeptide structure contains L-Lysine at the third position in replacement of the *meso*-A_2pm (Figure III.2).

Also, in Gram-positive bacteria, the thick layer of peptidoglycan is additionally altered by the presence of anionic polymers – teichoic acids and by the variation of the crosslinking degree between glycan chains.

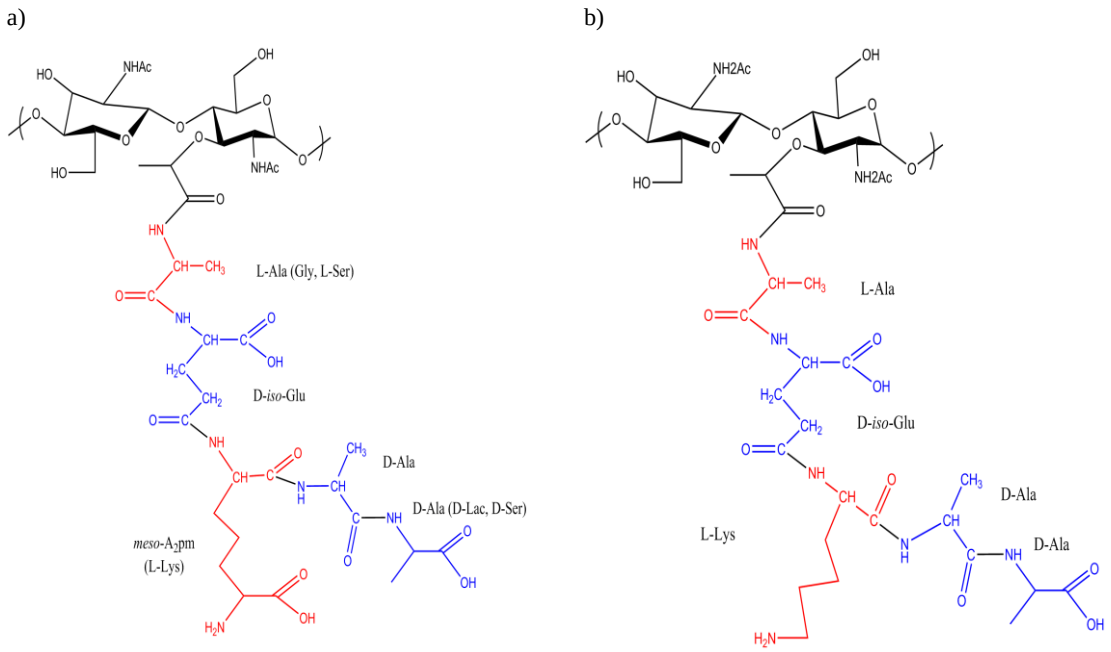


FIGURE III.2 – Chemical structure of the peptidoglycan pentapeptide. a) Typical pentapeptide sequence. Adapted from [74] b) *Staphylococcus aureus* peptidoglycan chemical structure. Adapted from [75]. L-Ala: L-alanine, D-*iso*-Glu: D-glutamic acid, *meso*-A_2pm: meso-diaminopimelic acid, D-Ala: D-alanine, L-Lys: L-lysine. Image created using ChemDraw®.

Biosynthetic pathway

The synthesis and incorporation of the new peptidoglycan strand to the cell wall involves a cascade of biochemical reactions catalysed by different proteins. In the general process, precursors are synthesized in the cytoplasm, linked to the transport lipid and flipped across the inner membrane to finally be incorporated into the main chain (Figure III.3). [76]

The pathogenic *Staphylococcus aureus* has a highly modified peptidoglycan and these modifications confer resistance and could also represent the mechanism how bacteria overcome the immune system of the host. [77]

Chemical analysis of the peptidoglycan of a *S. aureus* resistant strain showed that its structure suffered a small number of secondary modifications: some hydroxyl groups in the glycan chain were modified and also the second residue of the pentapeptide appeared to be D-*iso*-glutamine instead of D-*iso*-glutamic acid, although its modification mechanism remains a mystery. [5]

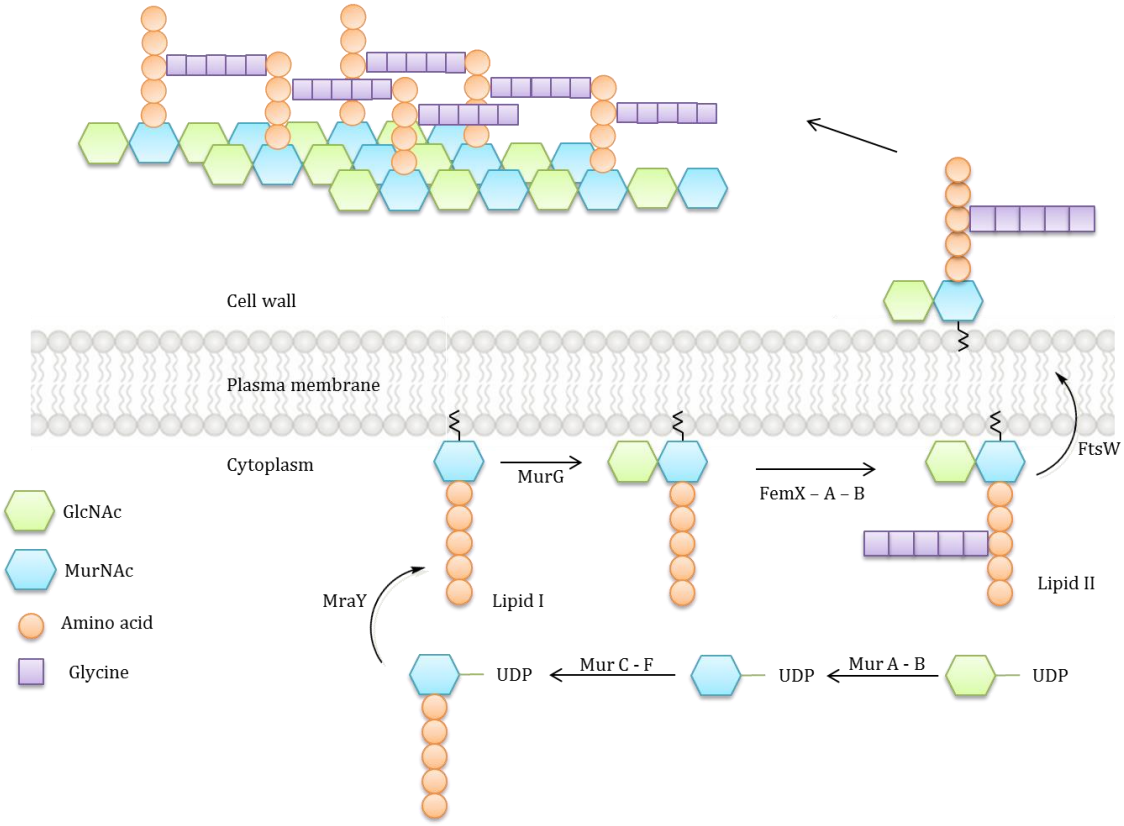


FIGURE III.3 – Schematic representation of the peptidoglycan biochemical pathway. In the overall process, precursors are synthesized in the cytoplasm, linked to the transport lipid and flipped across the inner membrane to finally be incorporated into the main chain. Adapted from [70][76].

Peptidoglycan amidation - genetic determinants and enzymes

S. aureus is considered the most important human pathogen due to its acquired antibiotic resistance, causing high levels of mortality in patients with nosocomial infections. This specific group of staphylococci that are resistant to all β-lactam antibiotics, including cephalosporin and staphylococcal penicillin as well as glycopeptide antibiotics, are called methicillin-resistant *Staphylococcus aureus* (MRSA). [78]

At the present date, eight MRSA genomes have been sequenced: COL, NCTC, 8325, N315, Mu50, MW2, UK EMR-SA-16 and MSSA476 and it is possible to find up to 20% variability in their genome sequence. Further DNA manipulation studies have been performed and revealed the existence of mutations that affect methicillin resistance in *S. aureus* and are correlated with the amidation of the glutamic acid residues in the peptidoglycan. [5][79]

The mechanism responsible for the amidation of the glutamic acid residues of the peptidoglycan in *S. aureus*, and many other bacteria, is still unknown. However, in the genome of *S.aureus* COL, an operon containing two genes, designated as *murT* and *gatD*, was found to be responsible for the amidation of the glutamic acid residues (Figure III.4). [5]

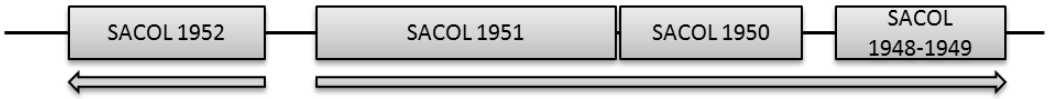


FIGURE III.4 – Schematic representation of *S. aureus* COL genome region that includes the two genes of interest: SACOL 1951 – *murT* and SACOL 1950 – *gatD* and the upstream and downstream regions. *murT* and *gatD* genes are transcribed in the same direction and no promoter was found between these two genes. Adapted from [5]

The genes *murT* and *gatD* occur as a syntenic sequence and their products show homology to murein ligases and to CobB/CobQ-like glutamine amidotransferases, respectively. [5][80]

Once the amino acid sequence of these two proteins is analysed and compared to the known homologs, several aspects become apparent. When MurT is compared with the other Mur ligases from *S. aureus*, the only common feature is the central domain (Figure III.5). When GatD is compared to the members of the glutamine-dependent amidotransferase family, the most evident aspect is that GatD only shares the glutaminase motifs (Figure III.6).

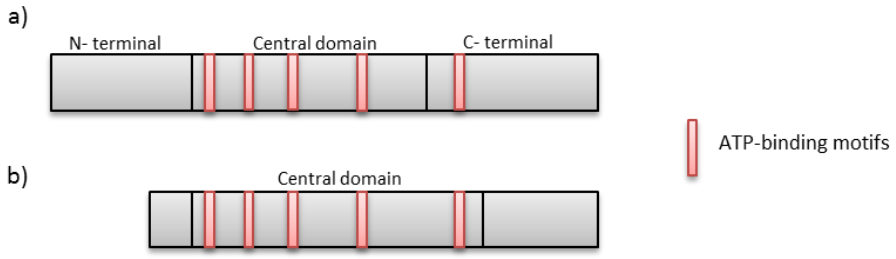


FIGURE III.5 – Comparison between the three domains characteristics of Mur ligases with MurT. a) This scheme represents the general three domains topology of Mur ligase proteins. N-terminal domain is responsible for the UDP-MurNAc-peptide binding, the C-terminal domain is responsible for the binding of the incoming amino acid and the central domain is where the ATP and co-factor Mg^{2+} bind. **b)** MurT topology lacks the flanking N- and C-terminal sequences and the C-terminal has a domain of unknown function that is conserved in more than 900 prokaryotic proteins – DUF1727. Adapted from [5]

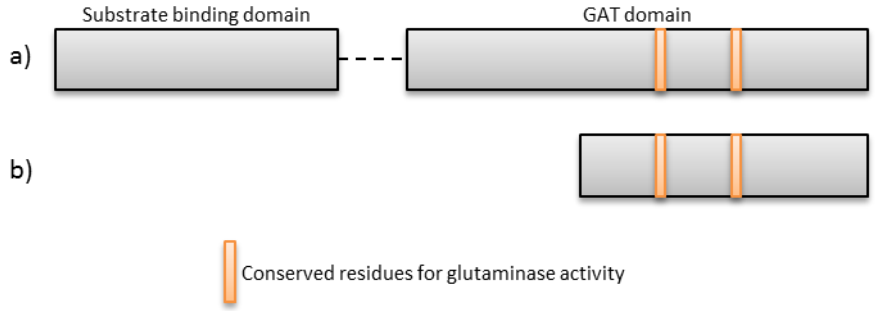


FIGURE III.6 – Comparison of glutamine-dependent amidotransferase structure with GatD. a) glutamine-dependent amidotransferases modular structure representation: a substrate binding domain (synthase domain) and a GAT domain with its conserved residues for glutaminase activity. b) GatD shares the GAT domain and the glutaminase activity motifs. Adapted from [5]

Although the mechanistic details are not yet known, it has been proven that the amidation of *S. aureus* peptidoglycan glutamic acid residues is the outcome of the action of this binary system. The proposed model refers MurT as responsible for the recognition of the reaction substrate, the lipid-linked peptidoglycan precursor and ATP, while GatD could be the catalytic subunit which transfers the amino acid group from free glutamine to the peptidoglycan precursor, since its sequence lacks the ATP binding motif (Figure III.6) suggesting that the catalytic activity depends on MurT.

Therefore, this complex seems to be, as far as one can tell, the last missing piece in order to explain the structural variation in *S. aureus* peptidoglycan.

III.2. Amidation of *S. aureus* peptidoglycan residues: experimental procedure

The objective of this project was the 3D structure determination of both proteins – MurT and GatD – and to further explore this protein-protein interaction.

The experimental procedure regarding the cloning, expression, purification and crystallization of the two proteins – MurT and GatD – was performed in a high throughput laboratory – Oxford Protein Production Facility (OPPF-UK) from Oxford University and the followed protocols are available online at http://www.oppf.rc-harwell.ac.uk/OPPF.

Data was collected at the synchrotron Diamond Light Source, beamlines I02 and I04.

The coding sequences of the *gatD* and *murT* gene were amplified from *S. aureus* COL strain genomic DNA and cloned into the pOPIN vector system using the In-Fusion™ method, a direct process where the PCR product is cloned straight into the expression vector. [81][82]

TABLE III.1 – Constructs designed at Oxford Protein Production Facility. Further details on appendix D.

Construct number	Gene	Vector	Description
12142		pOPINE-3C-HALO7	POI-3C-HALO7-KHIS$_6$
12141		pOPINE-3C-eGFP	POI-3C-eGFP-KHIS$_6$
12140	*murT*	pOPINK	HIS$_6$-GST-3C-POI
12139		pOPINS3C	HIS$_6$-SUMO-3C-POI
12137		pOPINB	HIS$_6$-3C-POI
12136		pOPINA	POI-KHIS$_6$
12144	*gatD*	pOPINB	HIS$_6$-3C-POI
12143		pOPINF	HIS$_6$-3C-POI

The murT and gatD constructs were transformed into *E. coli* expression strains – Rosetta(DE3) and Lemo21(DE3) – and the protein expression was tested using two different systems – auto-induction and IPTG induction – and the results analysed by SDS-PAGE.

The best results for the combination of fusion tag, *E.coli* expression strain and expression protocol combination were obtain for the construct 12139 (His$_6$-SUMO-MurT) and 12143 (His$_6$-GatD) transformed in *E. coli* Lemo21(DE3) and expressed through the auto-induction method. [56]

GatD

Expression and purification

GatD is a 27KDa type I glutamine amidotransferase-like protein that together with MurT, catalyses the amidation reaction of the glutamic acid residues of *S. aureus* peptidoglycan.

TABLE III.2 – Macromolecule production information. GatD construct 12143. Further details on appendix D.

Source organism	*S. aureus* COL strain
DNA source	*S. aureus* COL strain
Forward primer	AAGTTCTGTTTCAGGGCCCGCATGAATTGACTATTTATCATTTTATGTCAG
Reverse primer	ATGGTCTAGAAAGCTTTAACGAGATTTCTTCTGTCTATTTGCTC
Cloning vector	pOPINF
Expression vector	pOPINF
Expression host	*E. coli* Lemo21(DE3)

Native protein

Bacterial cells were inoculated in auto-induction media and after an incubation period of 4 hours at 37ºC followed by 16 hours at 18ºC, the cells were harvested and resuspended in 50mM Tris-HCl pH7.5, 500mM NaCl, 20mM imidazole and 0.2% Tween 20 supplemented with protease inhibitors and 400U.ml^{-1} DNAse type I.

As the protein of interest is cytosolic, the cells were lysed using a Basic-Z cell disruptor at 207MPa and clarified by centrifugation at 30000*g* for 30 minutes at 4ºC.

In order to purify the protein sample, the supernatant was loaded onto an ÄKTA Express system that is designed for computerized, multistep protein purification. The first purification step is an immobilized metal ion affinity chromatography (IMAC). The supernatant was loaded onto a 5ml HisTrap FF column previously equilibrated with 50mM Tris-HCl pH7.5, 500mM NaCl and 20mM imidazole (wash buffer). Once the supernatant was loaded onto the column, an extensive wash step was performed with wash buffer and eluted with 50mM Tris-HCl pH7.5, 500mM NaCl and 500mM imidazole (elution buffer). Once the sample is eluted, it is loaded onto a Superdex 200 HiLoad 16/60 column, equilibrated with 20mM Tris-HCl pH7.5, 200mM NaCl and 1mM TCEP, for a second purification step – size exclusion chromatography and the eluted fractions were analysed by SDS-PAGE (Figure III.7), and the aliquots containing the protein of interest were pooled together and the N-terminal tag was removed by cleavage with 3C protease. To cleave the hexahistidine tag, the protein solution was incubated with 5 units of 3C protease per mg of fusion protein, at 4ºC for 12 hours. The mixture was then purified by reverse affinity chromatography and fractions containing the protein of interest were combined and

stored in gel filtration buffer supplemented with 5% glycerol, at -80°C. In the case of GatD, the protein was stored at -80°C until necessary, without loss of its crystallographic capability.

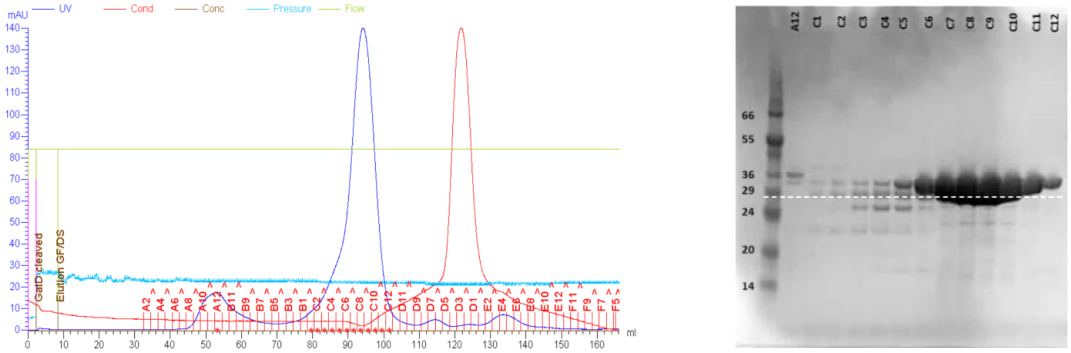


FIGURE III.7 – On the left: Typical chromatographic profile of N-His_6-GatD after the second purification step, the size exclusion chromatography. Injected sample volume of 6ml; Buffer: 20mM Tris-HCl pH7.5, 200mM NaCl, 1mM TCEP. Superdex 200 HiLoad 16/60 column (GE Healthcare). The fractions A12, C1 to C12 (*) were aliquoted and analysed by SDS-PAGE. The blue line corresponds to the A280nm and the red line corresponds to the sample conductivity. **On the right:** HT-SDS-PAGE analysis of the purified protein – Invitrogen NuPAGE Gels – run at 200V for 40 minutes in BisTris buffer and stained with a Coomassie based-solution. Lane 1 – molecular weight marker (kDa) and following lanes have the same nomenclature as the chromatogram. N-His_6-GatD expected MW: 27kDa.

Selenomethionine labelled protein

In order to produce the selenomethionine labelled protein a different expression protocol was used. Nevertheless the purification process is exactly as described above and the results are presented on figure III.8.

The protocol to produce selenomethionine labelled proteins with auto-induction media in methionine prototrophic strains uses a glucose-free selenomethionine media (Molecular Dimensions) and the *overnight* express system (Novagen) with additional amino acid solutions to inhibit the methionine biosynthetic pathway. [83]

Prior to crystallization, the SeGatD was submitted to a mass spectrometry experiment - liquid chromatography-electrospray ionisation-mass spectrometry (LC-ESI-MS), in order to confirm if the selenomethionines had been incorporated into the polypeptide. Comparing the deconvoluted spectrum of both native and labelled proteins (Figure III.9) it is possible to observe a shift in the monoisotopic peak, corresponding to an addition of two atoms of selenium.

With the labelling confirmed, the sample was then concentrated and sent to crystallization.

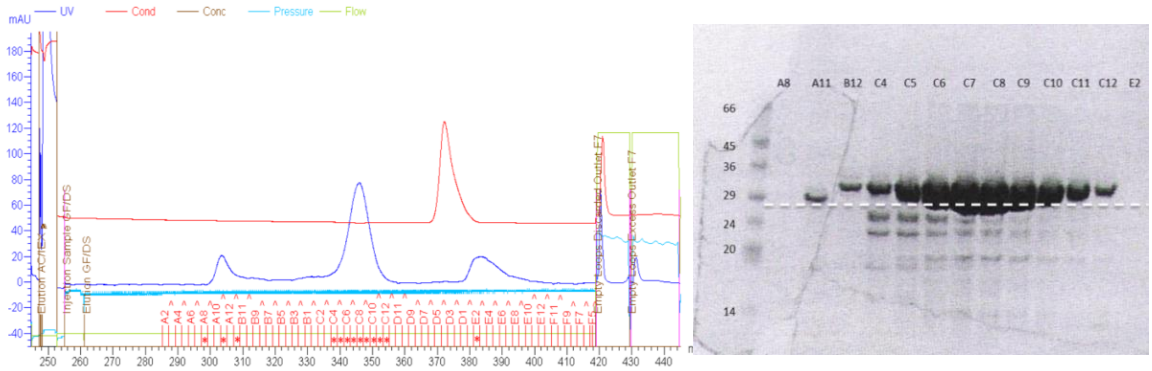


FIGURE III.8 – On the left: Typical chromatographic profile of N-His_6-SeGatD after the second purification step, the size exclusion chromatography. Injected sample volume of 5ml; Buffer: 20mM Tris-HCl pH7.5, 200mM NaCl, 1mM TCEP. Superdex 200 HiLoad 16/60 column (GE Healthcare). The fractions A8, A11, B12, C4 to C12 and E2 (*) were aliquoted and analysed by SDS-PAGE. The blue line corresponds to the A280nm and the red line corresponds to the sample conductivity. **On the right:** HT-SDS-PAGE analysis of the purified protein – Invitrogen NuPAGE Gels – run at 200V for 40 minutes in BisTris buffer and stained with a Coomassie based-solution. First and last lanes – molecular weight marker (kDa) and following lanes have the same nomenclature as the chromatogram. N-His_6-GatD expected MW: 27kDa.

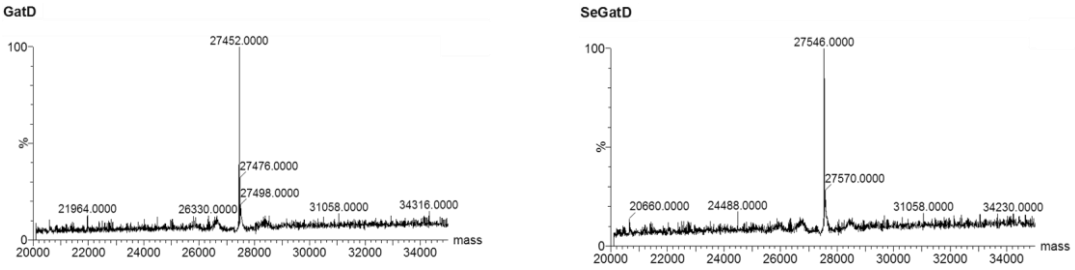


FIGURE III.9 – Mass spectra acquired for native (on the left) and labelled (on the right) GatD, by manual drop analysis and direct injection MS in a LC-ESI-MS. Sample concentration: 20µM. Service available at OPPF-UK and kindly performed by Dr. Joanne Nettleship.

Crystallization

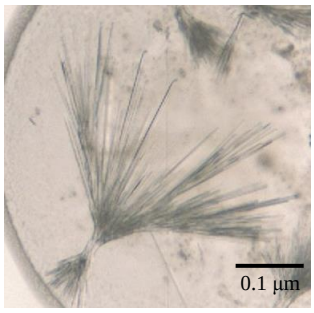
Crystallization screens were performed with a Cartesian instrument, the Digilab MicroSys liquid-handling system.

In a 96-well Greiner Bio-One plate, a 100nl protein sample was mixed with 100nl of crystallization solution and equilibrated over 90µl reservoir solution, in a sitting-drop vapour-diffusion method. The protein was concentrated to 20 and 45mg.ml^{-1} in 20mM Tris-HCl pH7.5, 200mM NaCl and 1mM TCEP for crystallization. The commercial crystallization screens used are listed in following table.

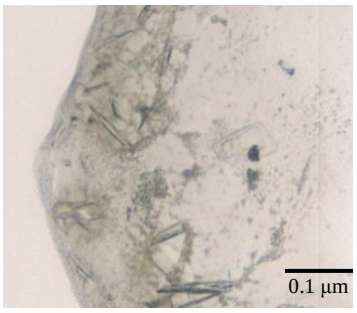
TABLE III.3 – List of the 96-well Greiner plate's barcodes, the commercial crystallization screens, and concentrations and temperatures trialled in order to crystalize the native protein sample: GatD.

Plate Barcode	Commercial Screen	Protein concentration (mg.ml^{-1})	Temperature (ºC)
4413005009723	Index	20	21
4413005009716	Emerald Wizard I+II		
4413005009594	Morpheus		
4413005009587	Emerald Wizard III+IV		
4413005009570	PACT premier		
4413005009709	JCSG+		
4413005008281	JCSG+	45	21
4413005008267	Emerald Wizard III+IV		
4413005008274	Emerald Wizard I+II		
4413005008298	Morpheus		
4413005008304	Index		
4413005008311	PACT premier		
4413005101335	Morpheus		4
4413005101311	Emerald Wizard III+IV		
4413005101328	Emerald Wizard I+II		
4413005101304	JCSG+		
4413005101298	Index		

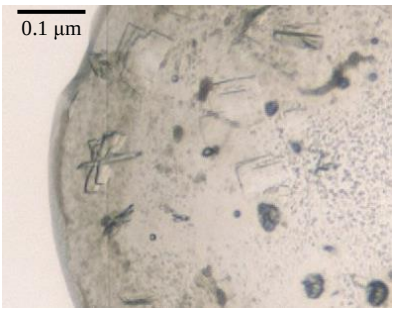
Crystals appeared in several conditions (Figure III.10) from screens Index and Emerald Wizard after 48 hours at 4ºC with a protein concentration of 45mg.ml^{-1}. The next step was to cryoprotect and flash-cool the crystals in order to be stored until data collection.



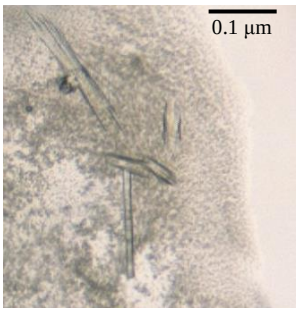
441305101311
Emerald Wizard III+IV (A3)
20% (w/v) PEG 3350
200 mM Magnesium formate



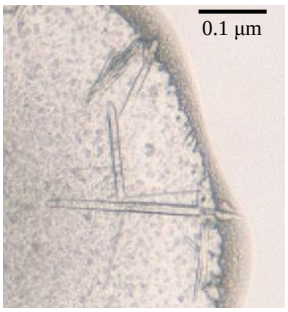
441305101311
Emerald Wizard III+IV (B5)
20% (w/v) PEG 4000
100 mM Sodium citrate/ Citric acid pH 5.5
10% (v/v) 2-Propanol



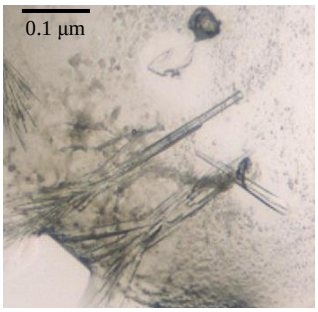
441305101311
Emerald Wizard III+IV (C8)
16% (w/v) PEG 8000
40 mM Potassium phosphate monobasic
20% (v/v) Glycerol



441305101328
Emerald Wizard I+II (D8)
30.0% v/v PEG 400
0.1 M sodium acetate pH4.5
0.2 M Calcium acetate

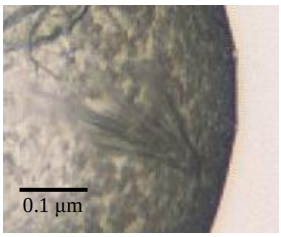


441305101304
JCSG-plus HT-96 (A5)
20.0% w/v PEG 3350
0.2 M Magnesium Formate

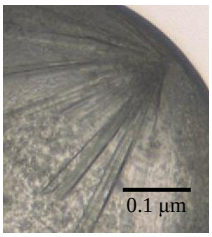


441305101298
Index (G10)
0.1 M Bis-Tris pH 5.5
0.2 M Magnesium Chloride
25.0% w/v PEG 3350

441305012518
Index (E06)
0.1 M bis-Tris pH 6.5
0.05 M Calcium Chloride
30.0% v/v PEG Monomethyl Ether 550



441305012518
Index (G06)
0.1 M Bis-Tris pH 5.5
0.2 M Ammonium Acetate
25.0% w/v PEG 3350



441305012518
Index (G10)
0.1 M Bis-Tris pH 5.5
0.2 M Magnesium Chloride
25.0% w/v PEG 3350

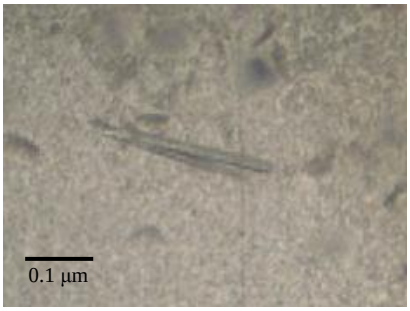
FIGURE III.10 – GatD protein crystals. Representative selection of different morphology obtained for GatD protein crystals and their conditions. Common to all crystallization solutions is the use of polyethylene glycol (PEG) as precipitant.

To determine the crystallographic phases for GatD it was necessary to produce crystals of the selenomethionine labelled protein (derivative crystals). The first approach was to reproduce the conditions that yielded the native crystals but replacing the native protein with the labelled one.

TABLE III.4 – List of the 96-well Greiner plate's barcodes, the commercial crystallization screens, concentrations and temperatures trialled in order to crystalize the labelled protein SeGatD.

Plate Barcode	**Commercial Screen**	**Protein concentration ($mg.ml^{-1}$)**	**Temperature (ºC)**
441305012518	Index	42	4
441305012501	Emerald Wizard III+IV		
441305012525	Emerald Wizard I+II		

Figure III.11 represents possible SeGatD protein crystals produced in two condition from the screen Emerald Wizard III+IV yielded. The crystals were harvested and cryoprotected with a solution of 50% (w/v) polyethylene glycol 3350 and 4000 and stored in liquid nitrogen until data collection.



441305012501
Emerald Wizard III+IV (A3)
20% (w/v) PEG 3350
200 mM Magnesium formate

441305012501
Emerald Wizard III+IV (B5)
20% (w/v) PEG 4000
100 mM Sodium citrate/ Citric acid pH 5.5
10% (v/v) 2-Propanol

FIGURE III.11 - SeGatD protein crystals. Representative selection of different morphology obtained for SeGatD protein crystals and their conditions. Common to all crystallization solutions is the use of polyethylene glycol (PEG) as precipitant.

MurT

MurT is a 49KDa ligase protein that, together with GatD, catalyses the amidation reaction of the glutamic acid residues of *S. aureus* peptidoglycan.

Expression and purification

The expression of MurT was a challenging process in the sense that it wasn't expressed with IPTG induction and with auto-induction the yield of soluble protein was very low (< 0.1mg per litre of culture).

MurT purification followed the same HTP protocol as the purification of GatD. Once the cells were lysed, the supernatant was uploaded onto an ÄKTAxpress twin system and two consecutive chromatographic steps were performed: IMAC and GF using the same buffers as the previously described as lysis buffer, wash buffer, elution buffer and gel filtration buffer.

MurT proved to be a very unstable protein in all conditions tested. Several attempts have been made to express and purify MurT with enough stable protein solution to pursue crystallization trials. The only viable option was to use the protein without cleaving its tag. When the sample was hydrolysed with 3C protease, by the end of the hydrolysis process the protein would have mostly precipitated. To make matters worse, it couldn't be stored as it would precipitate.

TABLE III.5 - Macromolecule production information. MurT construct 12139. Further information on appendix D.

Source organism	*S. aureus* COL strain
DNA source	*S. aureus* COL strain
Forward primer	AAGTTCTGTTTCAGGGCCCGAGACAGTGGACGGCAATCCATC
Reverse primer	ATGGTCTAGAAAGCTTTATGATTGACCTCCTTCAAACGAACGG
Cloning vector	pOPINS3C
Expression vector	pOPINS3C
Expression host	*E. coli* Lemo21(DE3)

Crystallization

With the small amount of protein obtained, two 96 well plates were set up with two commercial screens (table III.6). A 100nl protein sample was mixed with 100nl of crystallization solution and equilibrated over 90µl of reservoir solution in a sitting-drop vapour diffusion method. The protein was concentrated to a maximum of 7mg.ml^{-1} (higher concentrations induced protein

precipitation). From these crystallization essays, no crystals were obtained with the majority of the drops showing a dark precipitate.

TABLE III.6 – List of the 96-well Greiner plate's barcodes, the commercial crystallization screens, concentrations and temperatures trialled in order to crystalize the protein N-His_6-Sumo-MurT.

Plate Barcode	Commercial Screen	Protein concentration ($mg.ml^{-1}$)	Temperature (ºC)
441305019197	Emerald Wizard I+II	7	4
441305019203	Emerald Wizard III+IV		

The theory behind these two proteins describes a coordinated function between MurT and GatD in the amidation of the glutamic acid residues in the peptidoglycan of *S. aureus*. In this sense, and with the results obtained, where GatD is a stable protein and MurT showed to be very unstable in solution, the next approach was to co-purify and co-transform these two proteins in order to determine if the protein-protein interaction was the central key in MurT stability.

MurT-GatD complex

In order to obtain a solution with the binary complex MurT-GatD, two different approaches were followed. The first one involved the co-purification of N-His_6-Sumo-MurT with N-His_6-GatD. Both proteins were expressed independently following the protocol described above and the supernatant pooled together and loaded onto an AKTA Express system for purification.
The second approach was the co-transformation of both constructs into the expression *E.coli* strain and further co-purification in the high-throughput system described previously.

Co-purification of MurT and GatD

The supernatant resultant from independently expressed proteins was pooled together and loaded onto the AKTA Express system, for co-purification and the chromatographic profile obtained is presented on figure III.12.

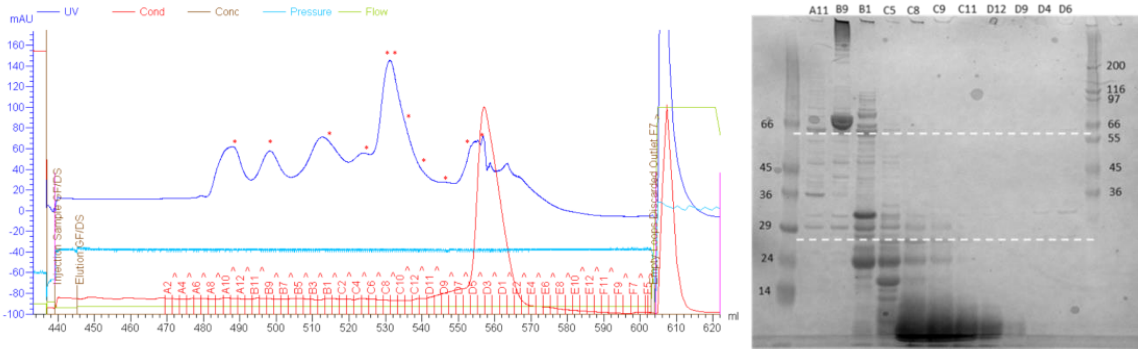


FIGURE III.12 – On the left: Chromatographic profile of N-His_6-GatD co-purified with N-His_6-Sumo-MurT after the second purification step, the size exclusion chromatography. Injected sample volume of 5ml; Buffer: 20mM Tris-HCl pH7.5, 200mM NaCl, 1mM TCEP. Superdex 200 HiLoad 16/60 column (GE Healthcare). The fractions A11, B9 B1, C5, C8, C9, C11, C12, D12, D9, D6 and D4 (*) were aliquoted and analysed by SDS-PAGE. The blue line corresponds to the A_{280nm} and the red line corresponds to the sample conductivity. **On the right:** HT-SDS-PAGE analysis of the purified protein – Invitrogen NuPAGE Gels – run at 200V for 40 minutes in BisTris buffer and stained with a Coomassie based-solution. First and last lanes – molecular weight marker (kDa) and following lanes have the same nomenclature as the chromatogram. N-His_6-GatD expected MW: 27kDa and N-His_6-Sumo-MurT MW: 61kDa.

In this case, it is possible to observe a band on aliquot B9 with the expected molecular weight of N-His_6-Sumo-MurT (expected MW: 61kDa) however no band corresponding to the fusion protein N-His_6-GatD (expected MW: 27kDa) neither in the same fraction nor in another one from this purification.
The next attempt to co-purify both proteins was to perform the IMAC purification first for both proteins separately and pool the eluted fractions together before injecting onto the Superdex 200

HiLoad 16/60 column from GE Healthcare for the second purification step. In this case, the profile obtained was exactly the obtained on figure III.12.

The last attempt of co-purification comprised the addition of pure GatD into the supernatant of N-His_6-Sumo-MurT and loading into the AKTA Express system, which presented the same chromatographic profile (data not shown) as the previous.

Since all the co-purifications failed, the next step with the aim of obtaining a MurT-GatD soluble complex was to co-express both genes.

Co-expression of MurT-GatD

In order to co-transform both genes – *murT and gatD* – a fresh PCR product using the construct 12143 (N-His_6-GatD) was obtained and cloned into the pOPINRSF HK vector, with kanamycin selection marker.

This new construct 12143HK (N-His_6-GatD) was then co-transformed into expression strains Rosetta(DE3)pLysS and Lemo21(DE3) in combination with one of each *murT* constructs (12142, 12141, 12140, 12139, 12138, 12137, 12136).

The analysis of the small scale expression tests yielded only one positive result: co-transformation of constructs N-His_6-GatD × N-His_6-SUMO-MurT.

Expression and purification

With a positive hit from the small scale analysis, further large scale expression and purification was performed. The protocol followed was the one described previously for the expression and purification of native GatD.

Bacterial cells were inoculated in auto-induction media and after an incubation period of 6 hours at 37ºC followed by 16 hours at 20ºC, the cells were harvested and resuspended in lysis buffer supplemented with protease inhibitors and 400$U.ml^{-1}$ DNAse I. The cells were lysed using a Basic-Z cell disruptor at 207MPa and clarified by centrifugation at 30000*g* for 30 minutes at 4ºC.

In order to purify the protein sample, the supernatant was loaded onto a 5ml HisTrap FF column previously equilibrated with wash buffer. Once the supernatant was loaded onto the column, an extensive wash step was performed with wash buffer and the protein eluted with 50mM Tris-HCl pH7.5, 500mM NaCl and 500mM imidazole. Once the sample was eluted, it was loaded onto a Superdex 75 HiLoad 16/60 column (GE Healthcare), equilibrated with 20mM Tris-HCl pH7.5, 200mM NaCl and 1mM TCEP, for a second purification step. The eluted fractions were analysed by SDS-PAGE, and the aliquots containing both proteins were pooled together and the N-terminal tag was removed by cleavage with 3C protease. To cleave both tags (His_6 and His_6-

Sumo) the protein solution was incubated with 5 units of 3C protease per mg of fusion protein, at 4ºC for 12 hours. The mixture was then purified by reverse affinity chromatography and fractions containing the proteins were combined and stored in gel filtration buffer.
The MurT-GatD purification profile after its co-expression is the one on figure III.113. Aliquots from the selected fractions were analysed by SDS-PAGE.

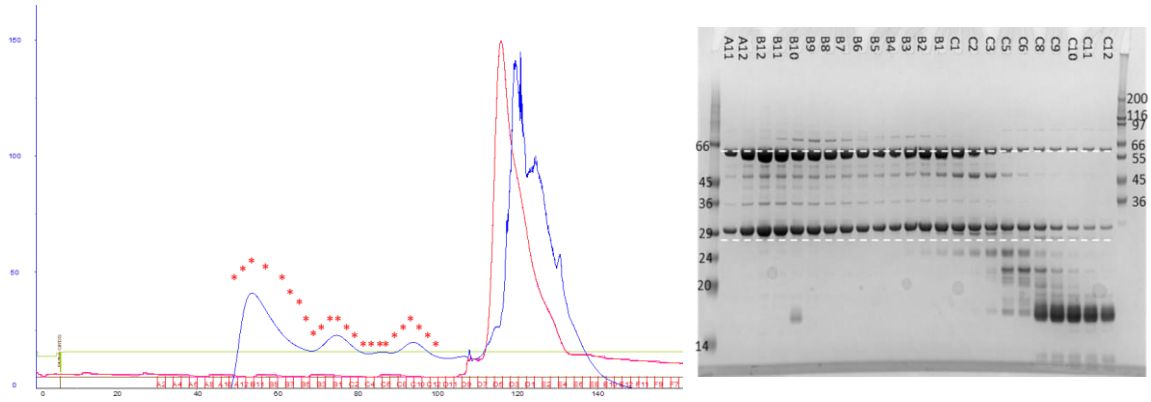


FIGURE III.13 – On the left: Chromatographic profile of co-expression of N-His_6-GatD with N-His_6-Sumo-MurT after the second purification step, the size exclusion chromatography. Injected sample volume of 5ml; Buffer: 20mM Tris-HCl pH7.5, 200mM NaCl, 1mM TCEP. Superdex 75 HiLoad 16/60 column (GE Healthcare). The fractions A11, A12, B12, B11, B10, B9, B8, B7, B6, B5, B4, B3, B2, B1, C1, C2, C3, C5, C6, C8, C9, C10, C11 and C12 (*) were aliquoted and analysed by SDS-PAGE. The blue line corresponds to the A_{280nm} and the red line corresponds to the sample conductivity. **On the right:** HT-SDS-PAGE analysis of the purified protein – Invitrogen NuPAGE Gels – run at 200V for 40 minutes in BisTris buffer and stained with a Coomassie based-solution. First and last lanes – molecular weight marker (kDa) and following lanes have the same nomenclature as the chromatogram. N-His_6-GatD expected MW: 27kDa and N-His_6-Sumo-MurT MW: 61kDa.

By observation of figure III.13 it is possible to verify the existence of bands approximately at the expected molecular weight of each protein and in this case, in the same fraction, indicating that they co-purify. However, they are present in all fractions collected instead of in a single fraction/peak. To determine the dispersity of the protein solution, the sample was submitted to dynamic light scattering (DLS) analysis – data not shown, prior to the crystallization trials. [71][72][86]

Crystallization

The crystallization trials were performed with protein straight after the reverse affinity purification. For the first trials attempted it was impossible to concentrate the protein solution over 7mg.ml^{-1}.

Further experiments were performed to increase the solubility of the MurT-GatD complex and the best results were obtained when the buffer included NDSB 256 (Hampton Research). In this case, it was possible to increase the protein concentration to a maximum of 30mg.ml^{-1}.
In order to facilitate protein-protein crystallization experiments, and as GatD crystals were easy to reproduce, some micro-seeding batches have been performed as well.
Table III.7 is a summary of all the prepared crystallization screens.

TABLE III.7 - List of the 96-well Greiner plate's barcodes, the commercial crystallization screens, concentrations and temperatures trialled in order to crystalize the protein complex MurT-GatD.

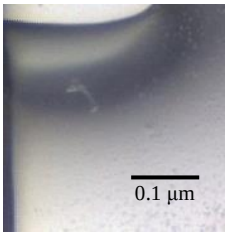
Plate Barcode	**Commercial Screen**	**Protein concentration (mg.ml^{-1})**	**Temperature (ºC)**	**Drop ratio P:R**
441305014222	ProPlex HT-96	7	20	1:1
441305014260	Morpheus			
441305016608	JCSG+			
441305016639	PACT premier			
441305016653	Emerald Wizard I+II			
441305016677	Emerald Wizard III+IV			
441305014239	ProPlex HT-96		4	
441305014253	Morpheus			
441305016615	JCSG+			
441305016646	PACT premier			
441305016660	Emerald Wizard I+II			
441305016684	Emerald Wizard III+IV			
441300488448	PACT premier	18	20	
441300488424	JCSG+			
441300488431	Morpheus			
441300488455	ProPlex HT-96			
441305023286	ProPlex HT-96	30	20	
441305023262	PACT premier			
441305023248	Morpheus			
441305023279	ProPlex HT-96		4	
441305023255	PACT premier			
441305023231	Morpheus			
441305022654	JCSG+	10 + seeds GatD	20	
441305022661	ProPlex HT-96			
441305022678	Morpheus			

A couple of conditions from PACT premier screen, with protein solution at 7mg.ml^{-1}, incubated at 4ºC yielded positive hits that have been further optimized, since the dimensions of the crystal were too small, even to be measured *in situ* (Table III.8).

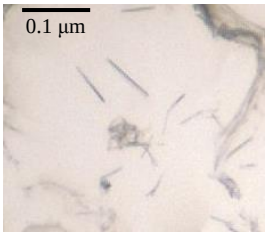
TABLE III.8 – List of the 96-well Greiner plate's barcodes, the commercial crystallization screens, concentration, temperature and protein:reservoir drop ratio (P:R) trialled in order to optimize the hits obtained for the protein complex MurT-GatD.

Plate Barcode	Commercial Screen	Protein concentration (mg.ml^{-1})	Temperature (ºC)	Drop ratio P:R
441305030192	PACT (E5)			
441305030208	PACT (D5)			1:1, 2:1, 3:1
441305030215	PACT (C8)			
441305030185	PACT	8	4	1:1
441305025662	PACT (C8)			4:1
441305025679	PACT (E5)			4:1
441305025655	PACT (D5)			4:1

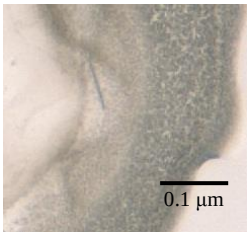
From the optimization drops, some crystals appeared after 5 days (Figure III.14). When analysed in the X-ray home source, these were proven to be salt crystals rather than protein crystals.



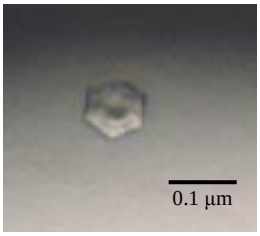
441300488455
ProPlex HT-96 (A12)
0.2 M sodium acetate
0.1 M sodium citrate pH 5.5
5 % w/v PEG 4000



441305023262
PACT premier (C08)
0.2 M Ammonium Chloride
0.1 M HEPES pH 7.0
20.0% w/v PEG 6000



441305023262
PACT premier (E05)
20.0% w/v PEG 3350
0.2 M Potassium Nitrate



441305030185
PACT premier (C12)
0.1 M HEPES pH 7.0
20.0% w/v PEG 6000
0.01 M Zinc Chloride

FIGURE III.14 – Possible MurT-GatD protein crystals. Representative selection of different morphology obtained Common to all crystallization solutions is the use of polyethylene glycol (PEG) as precipitant.

Data collection and structure determination

The tested GatD crystals were cryoprotected in a solution of 50% (w/v) PEG 400 and flash-cooled in liquid nitrogen before they were transferred to a gaseous nitrogen stream for data collection. [10]

GatD

Some of the tested native GatD crystals diffracted to ~10Å and ~3Å *in situ* but when flash-cooled and cryoprotected diffraction spots could be observed beyond 1.9Å.

The native data set was collected at I02 at Diamond Light Source (DLS, Didcot, UK), at a wavelength of 1.000Å on a Pilatus 6M detector (Figure III.15 on the left). The native GatD crystals belonged to space group $P2_12_12_1$, with unit cell dimensions a = 48.61Å, b = 93.92Å, c = 110.08Å.

The derivative crystal data set was collected at I04 at Diamond Light Source (DLS, Didcot, UK), at a wavelength of 0.9796Å on an ADSC Q315r detector (Figure III.15 on the right).

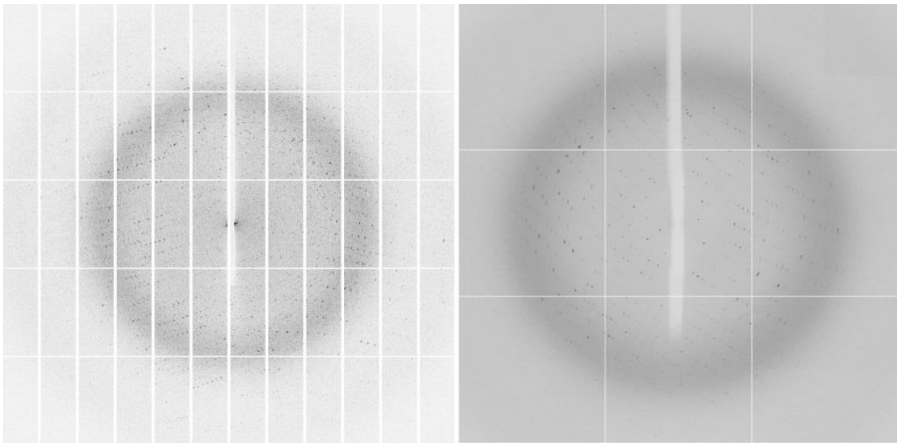


FIGURE III.15 – Representative diffraction patterns of the crystals. Left: native; Right: SeMet derivative. The resolution at the edge of the detector is 2.1 and 2.5Å, respectively.

The crystal belonged to the same space group as the native crystal, $P2_12_12_1$ and unit cell dimensions are a = 48.28Å, b = 93.00Å, c = 109.30 Å.

The experimental phases to determine the 3D structure of GatD were obtained by single-wavelength anomalous diffraction (SAD) using data collected from the SeMet derivative (Figure III.16) at the Se edge peak and data was automatically processed using *xia*2. [87]

Diffraction data were integrated, scaled, merged, and reduced with AIMLESS [88] and refined with Refmac5 within the CCP4 suite of programs [89][90][91][92].

Data-collection and processing statistics are given in table III.9.

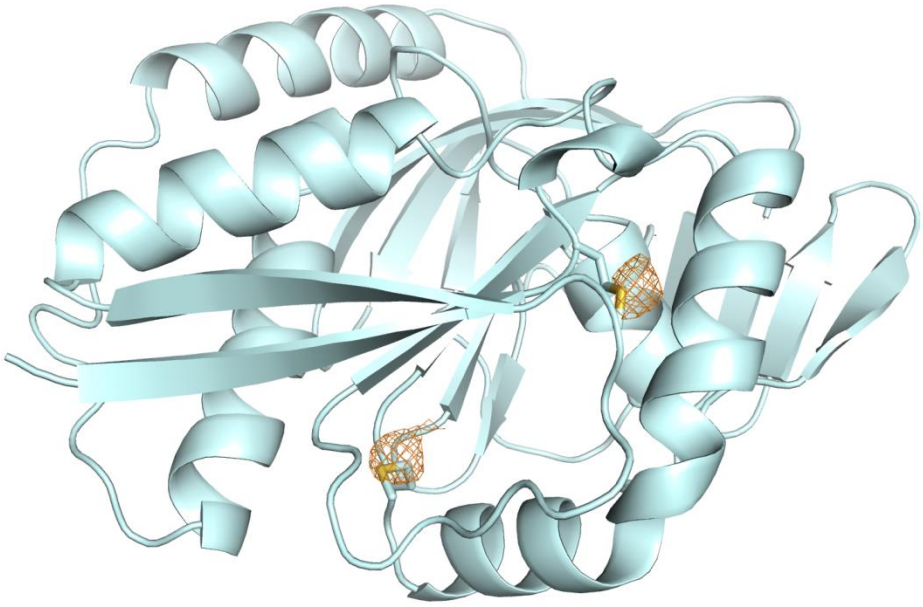


FIGURE III.16 – GatD model in a light blue cartoon representation with the anomalous electron density (in orange at a 3σ contour) for the two (out of three) atoms of selenium in the SeGatD derivative. This figure and other structure-related figures reported in this thesis were prepared using WinPymol. [60]

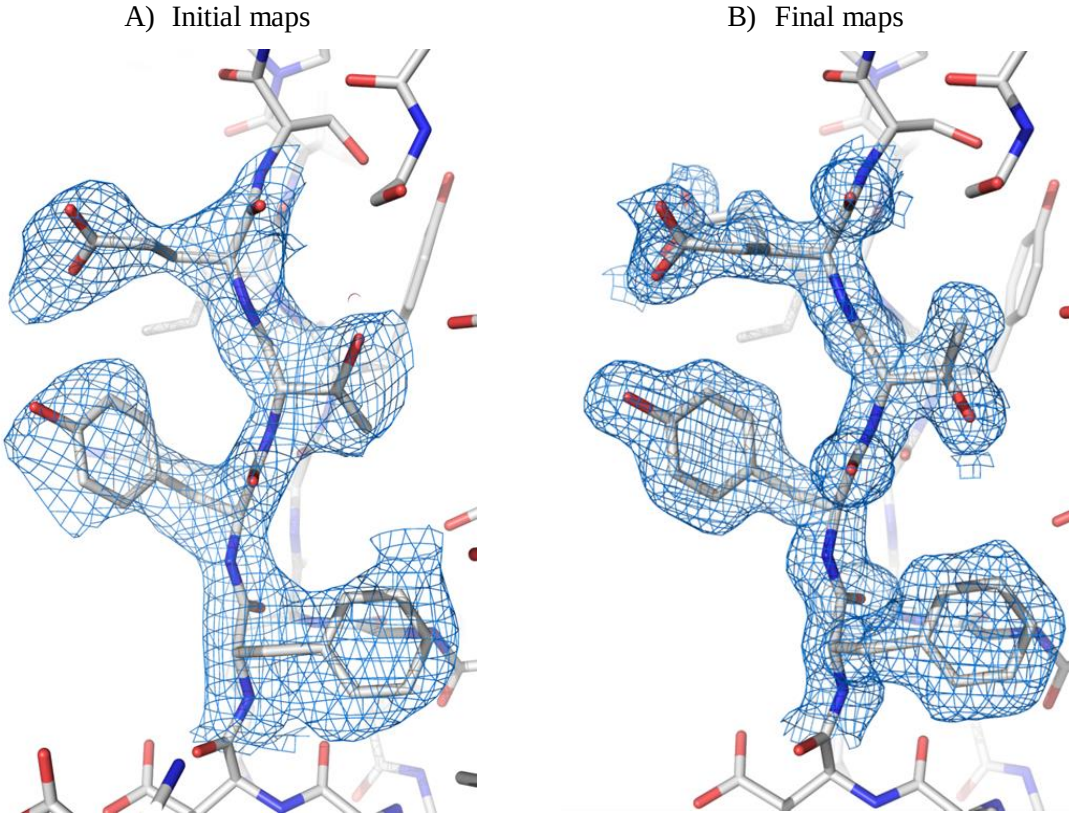


FIGURE III.17 - The experimental phases to determine the 3D structure of GatD were obtained by single-wavelength anomalous diffraction (SAD) using data collected from the SeMet derivative at the Se edge peak. The model was refined with Refmac5 from CCP4. The 2mFo-DFc electron density maps are contoured at 1σ (blue).

TABLE III.9 - X-ray crystallography data-collection, processing and refinement statistics. (Values in parenthesis correspond to the outer shell) [93]

	Native	SAD
Diffraction source	I02, DLS	I04, DLS
Wavelength (Å)	1.0000	0.9796
Temperature (K)	100	
Detector	Pilatus 6M	ADSC Q315r
Matthews coefficient ($Å^3.Da^{-1}$)	2.33	2.27
Number of molecules per AU	2	
Solvent content (%)	47.17	45.90
Space group	$P2_12_12_1$	
a, *b*, *c* (Å)	48.61, 93.92, 110.08	48.28, 93.00, 109.30
α, β, γ (Å)	90, 90, 90	
Crystal system	Orthorhombic	
Mosaicity (º)	0.141	0.387
Resolution range (Å)	36.43-1.85	47.12-2.25
Total number of reflections	839410	301512
Number of unique reflections	42987	24020
Completeness (%)	97.9 (83.7)	98.7 (89.2)
Multiplicity	19.5 (10.1)	12.6 (6.7)
I / σ(I)	20.7 (2.2)	16.5 (2.2)
$R_{p.i.m.}$	0.028 (0.313)	0.078 (0.477)
Overall B factor from Wilson plot ($Å^2$)	15.237	9.429
Anomalous completeness (%)		98.7 (89.2)
Anomalous multiplicity		6.7 (3.4)
R_{work}	0.164	
R_{free}	0.197	
RMSD bond length (Å)	0.020	
RMSD angle length (º)	1.888	
Ramachandran plot (%)		
Residues in favoured regions	97.89	
Residues in allowed regions	2.11	
Residues in disallowed regions	0	

III.3. Amidation of *S. aureus* peptidoglycan residues: results and discussion

As described previously, in structural biology, depending on the aim of the experiment, there are some aspects that need to be carefully analysed before heading to the wet lab and starting the cloning, expression and purification process.

The aim of this project is to structurally analyse two proteins that were proven to be involved in the amidation of glutamic acid residues in the bacterial cell wall of *S. aureus* peptidoglycan. [5]

To achieve this, it's important to analyse and select the sequence to clone and express.

To start the sequence analysis, the best option is to perform a basic local alignment search tool (BLAST) to search for sequence homologs and conserved domains within the protein. With this tool, GatD was identified as a type 1 glutamine amidotransferase (GATase1)-like with a putative conserved domain. [40][41][47]

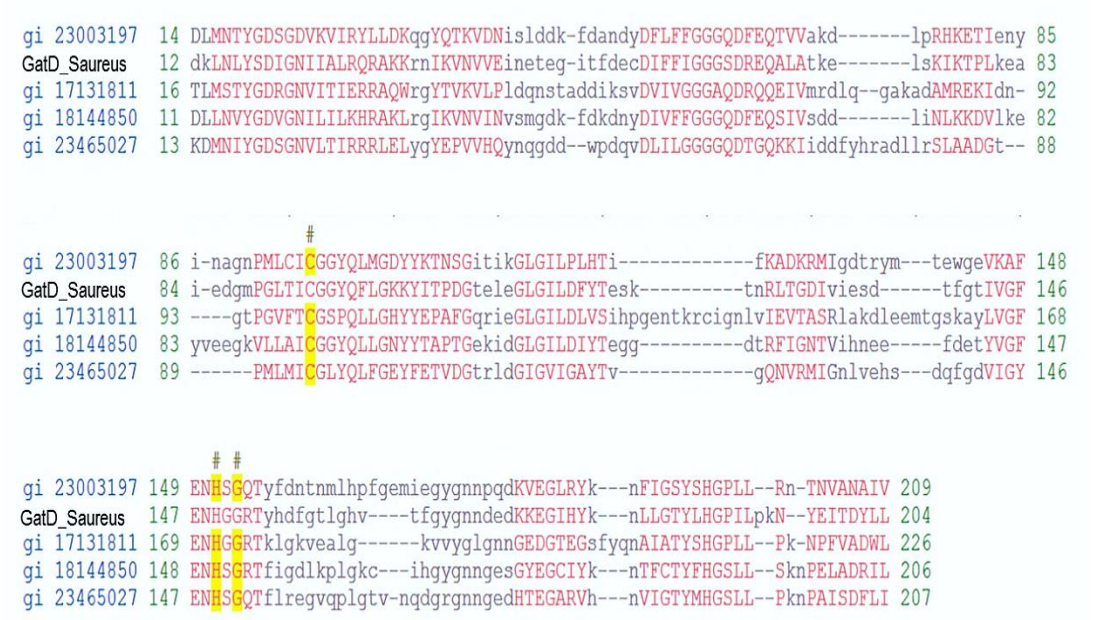


FIGURE III.18 – Sequence alignment of four proteins with conserved glutamine amide transfer domains. The highlighted residues correspond to the conserved triad of residues responsible for the catalytic function. gi_23465027: cobyric acid synthase CobQ from *Bifidobacterium longum*; gi_18144850: probable cobyric acid synthase from *Clostridium perfringens* str. 13; gi_17131811: alr2718 from *Nostoc sp.*; gi_23003197: hypothetical protein Lgas_03000924 from *Lactobacillus gasseri.* [47]

Overall Structure Description

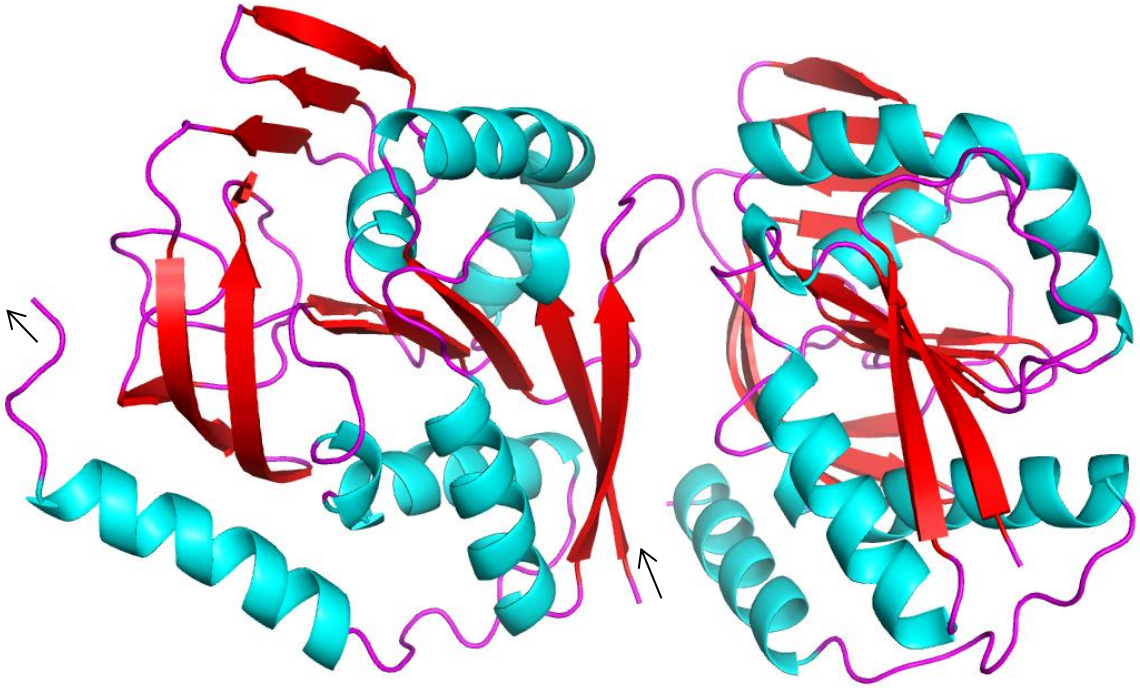


FIGURE III.19 – Crystallographic dimer of GatD in a cartoon representation and coloured according to the secondary structure: *α*-helices in blue, *β*-sheets in red and loops in purple. This figure and other structure-related figures reported in this thesis were prepared using WinPymol. [60]

The crystal structure is presented as the crystallographic dimer (two molecules in the asymmetric unit) although GatD is a monomer composed of 243 amino acids that are folded in a mixed topology, instead of the α-β-α fold characteristic of glutaminase structures.

GatD proposed model has no mutations in its sequence as the model was manually built using the GatD amino acid sequence in UniProt reference (Q5HEN2).

GatD shows electron density for the overall structure, including the N-terminal helix, enabling the complete assignment of the amino acid sequence on both molecules of the asymmetric unit, except for residues 1 and 239-243 from molecule B.

Analysing its secondary structure (Figure III.20) and correlating it with the position of the residues that compose the triad, it is possible to locate the reactive cysteine in the so called 'nucleophile elbow' - a region that contains a β-α structural motif. The tightness of the strand-turn-helix motif induces the nucleophilic amino acid residue to adopt energetically unfavourable main chain torsion angles and imposes steric restrictions on residues located in its proximity.

Although the residues of the catalytic triad are far from each other within the primary structure, they are brought together in the final fold.

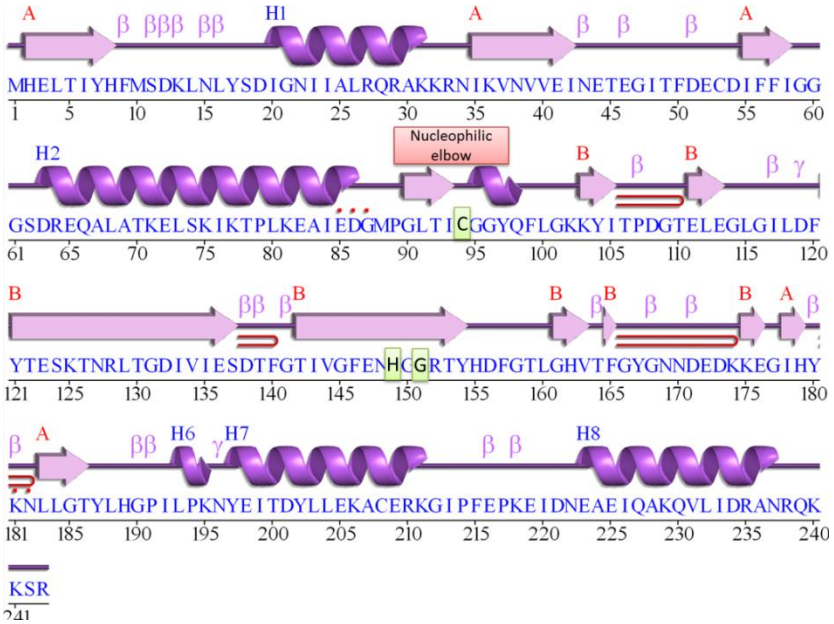


FIGURE III.20 – GatD secondary structure analysis. The catalytic triad has a reactive cysteine at position 94 in the turn between a β-strand and α-helix designated as the 'nucleophile elbow'. The residues highlighted in a green box are GatD catalytic triad. Image generated by PDBSum. [94]

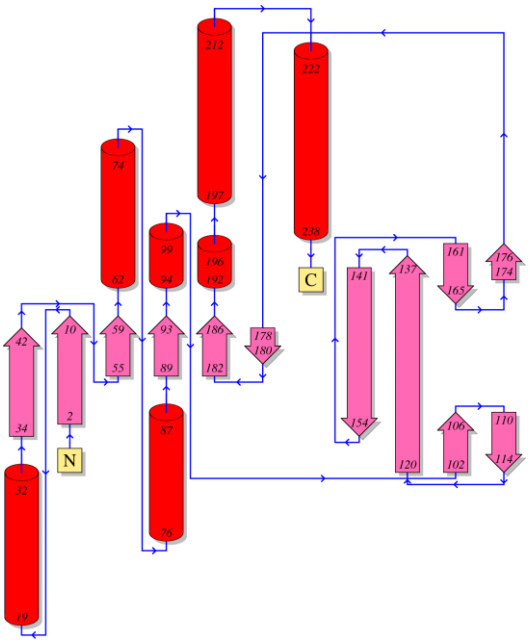


FIGURE III.21 - GatD topology**.** GatD has a mixed topology, instead of a α-β-α fold characteristic of glutaminase structures. Image generated by PDBSum. [94]

As referred previously, GatD is a type I glutamine amidotransferase and its active centre is composed of the catalytic triad found in all members of this family: a nucleophilic cysteine, a basic histidine and a glycine. In a close-up view at the active site, it is very easy to identify the catalytic triad at positions Cys_{94}-His_{149}-Gly_{151} (Figure III.22).

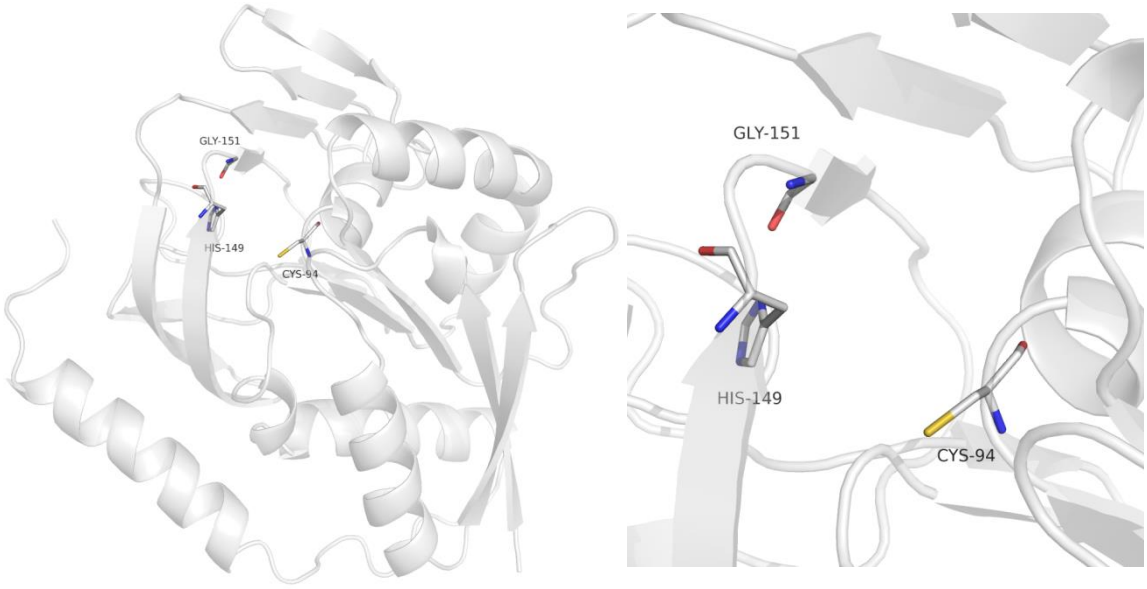


FIGURE III.22 – On the left: Overall representation of GatD and the position of the residues that compose the triad within the molecule, located at positions 94, 149 and 151. **On the right:** closer view of the catalytic triad residues. The cartoon representation is from the monomeric GatD. The residues C_{94}-His_{149}-Gly_{151} are represented as sticks and coloured by atoms: carbons in white, nitrogen in blue, oxygen in red and sulfur in yellow.

The distances between the atoms from the catalytic triad (Figure III.23) suggest that GatD has been crystallized in an inactive conformation. The current accepted mechanism of the triad's action involves the activation of the nucleophilic cysteine by the histidine. In this structure, the distance between these two residues is >3Å, which suggests no interaction between the two and, consequently, no activation of the nucleophile to initiate the conversion reaction of glutamine into glutamate and ammonia.

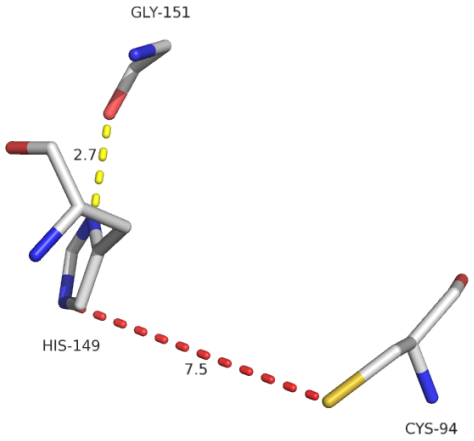


FIGURE III.23 - Distances (Å) between the catalytic triad residues. The mean donor-acceptor distances in proteins structure elements are approximately 3Å. In this structure, the distance between this two residues is >3Å, which suggest no interaction between histidine and cysteine residues and consequently, no activation of the nucleophile. The residues are coloured by atoms: carbons in white, nitrogen in blue, oxygen in red and sulfur in yellow.

Comparison with homologous structures

The closest homologue to GatD found in the PDB is HISF from *Thermotoga maritima*, a protein involved in the purine biosynthesis, that bears both glutamine and synthase activities in a bi-enzyme complex, with the glutaminase site separated from the synthase site by an ammonia tunnel of 25Å. [95]

A structural superposition of GatD with already identified structures of class I GATases shows a high similarity in the overall structure, particularly in the glutaminase active centre (Figure III.24 and III.25). [96]

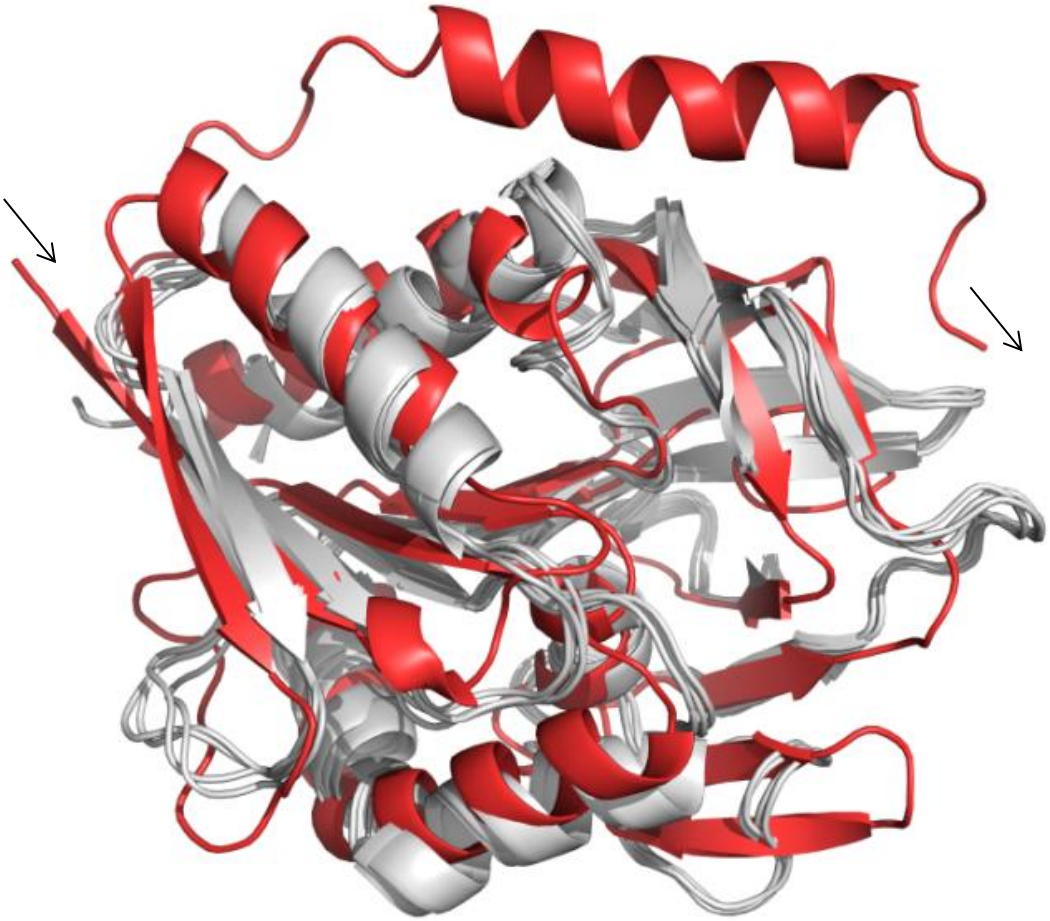


FIGURE III.24 – Superimposition of GatD structure with three other type 1 GATases [96]**.** Glutamine Amidotransferase from *Thermotoga maritima* (PDB ID: 1KXJ); Imidazole glycerol phosphate synthase subunit from *Thermotoga maritima* (PDB ID: 3ZR4); HISF Protein from *Thermotoga maritima* (PDB ID: 1GPW) and Amidotransferase HisH from *Thermotoga maritima* (PDB ID: 1K9V) in grey. The proteins are represented as cartoon and GatD is coloured in red. RMSD: 5.238 (1GPW, 1K9V); 5.250 (1KXJ); 5.220 (3ZR4).

In a close-up view at the active site superimposition it is possible to recognize the structural homology of the conserved catalytic triad.

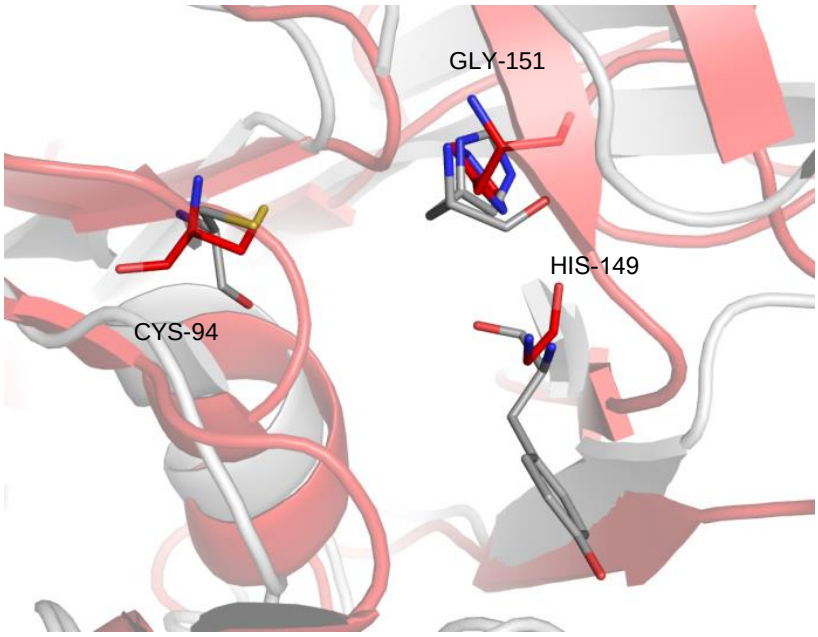


FIGURE III.25 – Superimposition of the catalytic triad Cys_{94}-His_{149}-Gly_{151} from GatD with the catalytic triad Cys_{84}-His_{178}-Glu_{180} of the homologous structures: Glutamine Amidotransferase from *Thermotoga maritima* (PDB ID: 1KXJ); HISF Protein from *Thermotoga maritima* (PDB ID: 1GPW) in grey. GatD is represented in red and the catalytic triad's residues are shown as ribbon and coloured by atom type: carbon in white, nitrogen in blue, oxygen in red and sulfur in yellow.

The catalytic triad generates a nucleophilic residue for covalent catalysis using an acid-base-nucleophile triad. Each amino acid in the triad has a specific task in the catalytic process.

The cysteine has an -SH group that is able to act as a nucleophile, attacking the carbonyl carbon of the substrate. Since the naturally occurring amino acids are not nucleophilic enough, the basic residue in the catalytic triad - histidine - deprotonates the nucleophile - cysteine - in order to increase its reactivity.

The acid role is played by glycine, although in the homologous structures the third member of the triad is glutamic acid. Besides this difference the catalytic process does not suffer any change as the carboxyl group on the glycine, hydrogen bonds with the histidine making it more electronegative. [97][98]

Ligand binding prediction

Protein–ligand modelling has become, in the past few years, a powerful tool for drug development. Due to its biological function in the synthesis of peptidoglycan, GatD is a potential target for pharmaceuticals.

Structural docking of a ligand:macromolecule complex rely on the most energetically favourable binding. The prediction was done using AutoDock Vina and analysed with WinPymol. [99][100] The cluster in the active site cleft has nine possible conformations for glutamine which are ranked according to their binding energy.

The most energetically favourable mode is the first presented on table III.10 and is represented in Figure III.26.

TABLE III.10 – Affinity binding energies from the 9 modes of glutamine predicted for the glutamine:GatD complex predicted by AutoDock Vina. [99]

Mode	Affinity (kcal/mol)	RMSD
1	-5.0	0.000
2	-5.0	1.104
3	-4.7	1.818
4	-4.7	2.007
5	-4.6	2.255
6	-4.6	1.284
7	-4.6	2.576
8	-4.5	2.384
9	-4.4	2.250

The glutamine analogue, 6-diazo-5-oxo-L-norlucine (DON), is a diazo compound and is known to interfere with both nucleotide and protein synthetic pathways where glutamine acts as a substrate. [101]

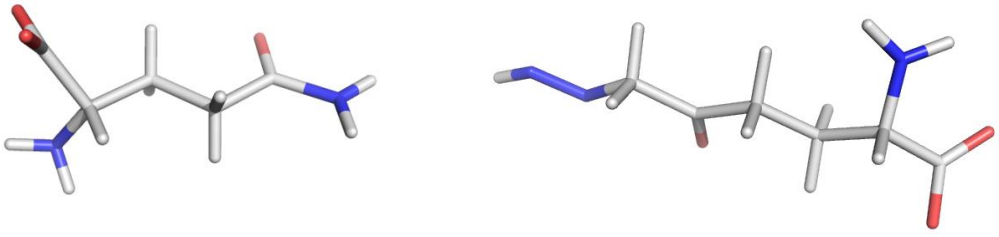


FIGURE III.26 - Comparison between glutamine (left) and its analogue 6-diazo-5-oxo-L-norlucine (DON).

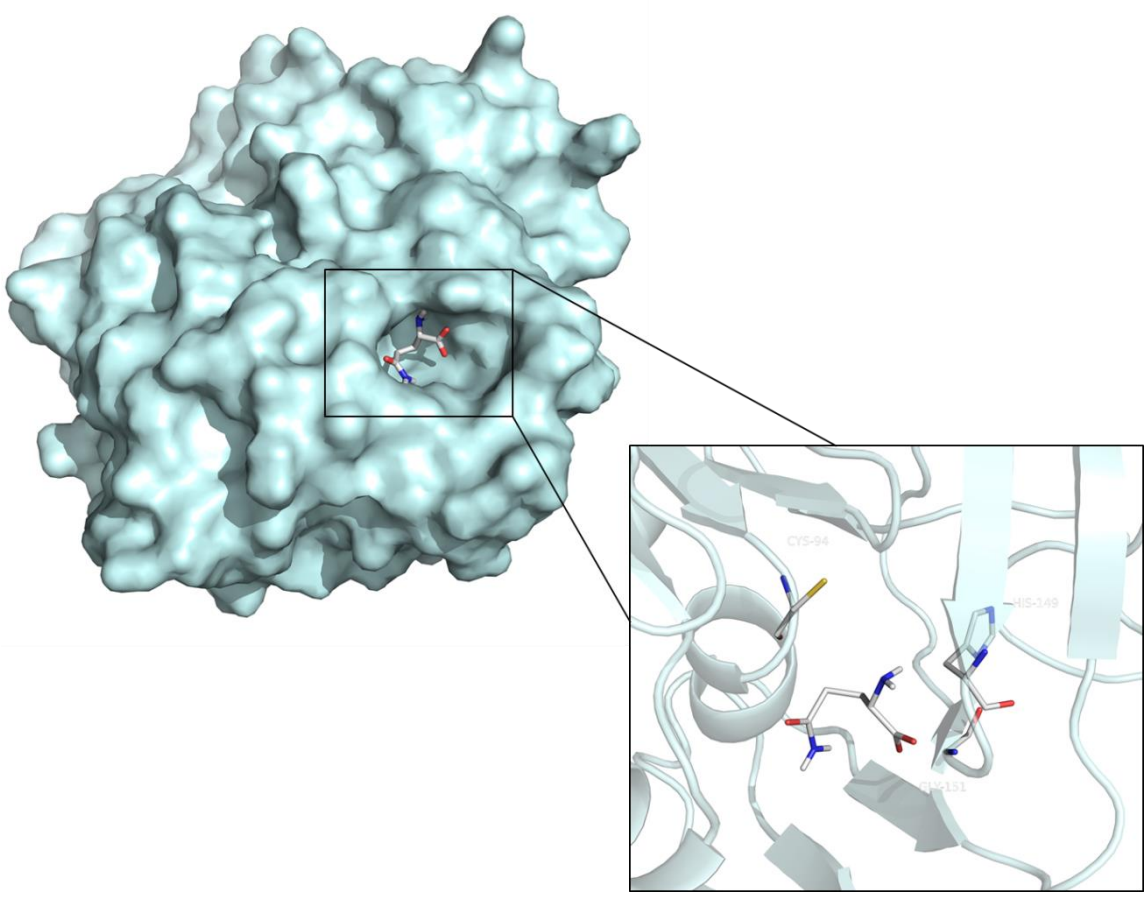


FIGURE III.27 – Cluster predicted by AutoDock Vina in a GatD:glutamine interaction at the active site and a closer look at the active site. The residues are coloured by atom type: carbon in white, nitrogen in blue, oxygen in red and sulfur in yellow. [99]

Structural docking of a DON:GatD complex rely on the most energetically favourable binding as for the previous prediction. Table III.11 presents the affinity binding energies for the seven predicted DON conformations.

TABLE III.11 - Affinity binding energies from the 9 modes of DON predicted for the GatD:DON complex predicted by AutoDock Vina. [99]

Mode	Affinity (kcal/mol)	RMSD
1	-4.6	0.000
2	-4.5	2.880
3	-4.4	2.037
4	-4.2	2.178
5	-4.2	3.876
6	-4.1	2.298
7	-3.9	2.649

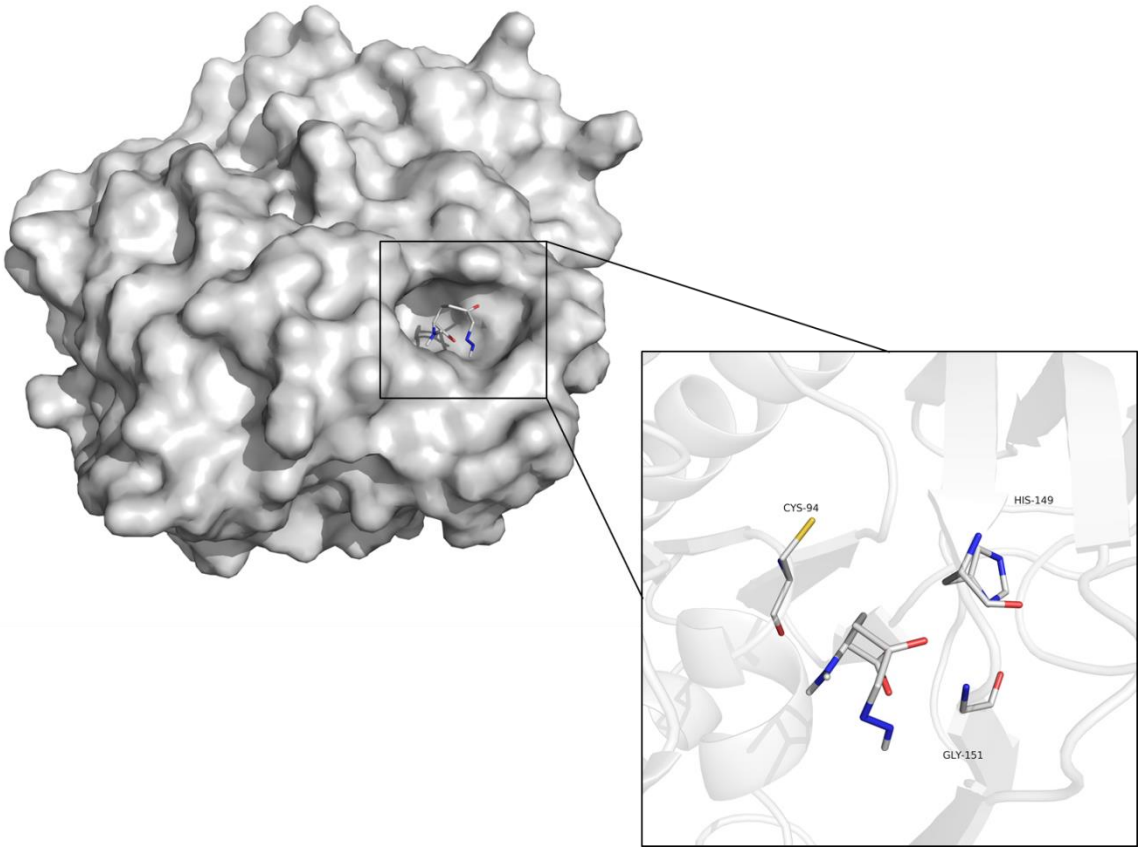


FIGURE III.28 - Cluster predicted by AutoDock Vina in a DON:GatD interaction at the active site and a closer look at the active site. The residues are coloured by atom type: carbon in white, nitrogen in blue, oxygen in red and sulfur in yellow. [99]

Comparing the predictions of GatD:glutamine and GatD:DON the binding affinity is higher in the case of the glutamine (lower affinity energy). This prediction supports the *in vitro* experiments that suggest that the GatD-MurT amidation reaction is highly specific for glutamine. [6]

Proposed mechanism for GatD-MurT

In vitro analyses have demonstrated that amidation of peptidoglycan in *S. aureu*s is catalysed by the bi-enzyme complex GatD-MurT. In the basic mechanism of most GATases, ammonia is sent through a solvent channel from the glutaminase active site to a synthase active site. [6]
Münch *et al* suggested a model that combine the action of these two enzymes in which, glutamine is the nitrogen donor and ammonia is transferred from GatD to the synthase domain, possibly present in MurT structure, across an ammonia tunnel, where the lipid-II precursor is bound, which finalizes the lipid-linked peptidoglycan precursor amidation (Figure III.29).

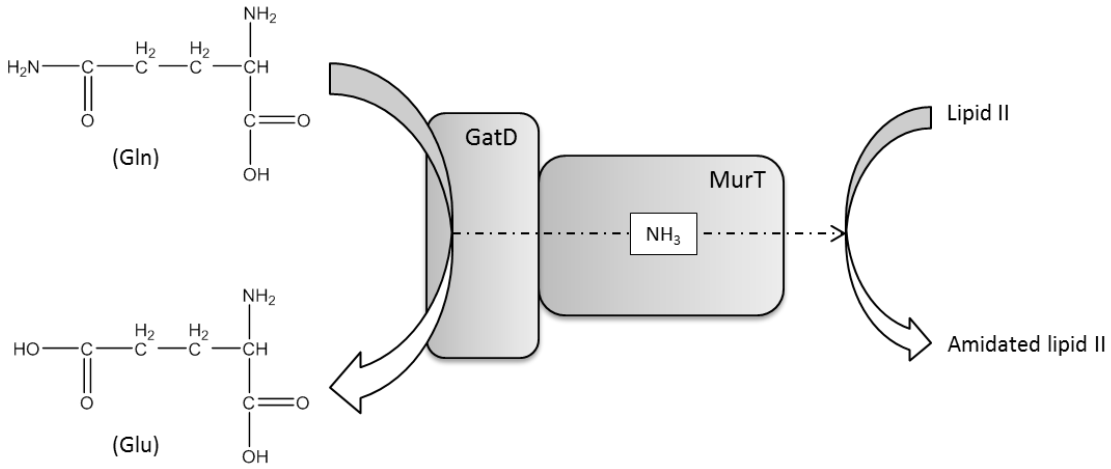


FIGURE III.29 – Proposed mechanism for GatD-MurT combined action. Glutamine (Gln) is used as a nitrogen donor and is conversed at the glutaminase site in Glutamate (Glu) and ammonia (NH_3). Ammonia is then transferred from GatD to the synthase domain, possibly present in MurT, across an ammonia channel, to finally amidate the lipid-II precursor. [6]

Considering the important role of these proteins in *S. aureus* and their high conservation within Gram-positive bacteria, its structural identification may enable the development of new drugs or combining therapeutics with the existing β-lactam antibiotics to overcome the MRSA threat.

III.4. Amidation of *S. aureus* peptidoglycan residues: conclusion and future work

Conclusions

The recently identified MurT-GatD enzymatic complex represents an unexplored stage as a potential antimicrobial target. MurT shares considerable similarity with the topology of the Mur ligases of *S. aureus*, which are cytoplasmic enzymes that are responsible for the sequential addition of amino acid residues to the growing peptidoglycan.

GatD is a 27KDa type I glutamine amidotransferase-like protein, shown to be responsible for the amidation reaction of the glutamic acid residues of the peptidoglycan of *S. aureus* in combination with MurT ligase. [5] GatD shows similarity to the glutamine amidotransferases, which show glutamine amide-transfer activity to a wide variety of substrates. Typically GATases catalyse two distinct reactions, one in the glutaminase site and the other in the synthase site. GatD correspond only to the glutaminase domain, which is responsible for the conversion of glutamine into ammonia and glutamic acid.

In order to determine the three dimensional structure of these two proteins independently and in complex, the encoding region of both genes was cloned into pOPIN vectors and expressed in *E. coli* Lemo21(DE3). The crystallization trials were performed at a high-throughput facility.

The crystallization hits obtained were from GatD and the complex. When submitted to an X-ray beam, the crystals from MurT-GatD complex showed no protein diffraction signal, while GatD crystals diffracted to several resolutions. The best crystal diffracted beyond 1.9Å and its diffraction data was collected at Diamond Light Source. The initial phases were obtained through SAD using data collected at the Se edge peak, in selenomethionine labelled protein crystals.

The crystals belong to space group $P2_12_12_1$, with unit cell dimensions of a = 48.29, b = 93.00, c = 109.31Å in an orthorhombic crystalline system, with two molecules in the asymmetric unit.

Structural analysis of GatD shows a mixed topology instead of the α-β-α fold characteristic of glutaminase structures, but as a type I glutamine amidotransferase, has a characteristic glutaminase active center composed of three residues in a nucleophile-base-acid conformation. GatD catalytic triad is Cys_{94}-His_{149}-Gly_{151} showing a positional homology with a highly conserved catalytic cysteine, the basic residue histidine and the acid member of the triad which in GatD is glycine whereas in the structural homologous is glutamate.

Considering the important role of these proteins in *S. aureus* and their high conservation within Gram-positive bacteria, its structural identification may enable the development of new drugs or combining therapeutics with the existing β-lactam antibiotics to overcome the MRSA threat.

Future work plan

In this project the main goal the structure determination of the MurT-GatD complex by X-ray crystallography. The knowledge of the 3D structure will show the important residues for substrate binding and/or catalytic mechanism and will help in efforts of structure based drug discovery.

This is still an ongoing project and there are several experiments to be done in a short-term and experiments to be done in long-term.

GatD structure was determined and structural analysis and modelling experiments showed the triad with catalytic function in this type of enzymes. Knowing the substrate, the next item in our to-do list is to obtain GatD structure in its active form, with glutamine, in order to demonstrate its mechanism. To obtain a crystal of a protein:ligand complex, there are strategies that can be followed. Prior to the crystallization step, we can co-express the protein with the ligands of interest or use the ligand during protein purification. In the case of GatD since we already know the condition to obtain the crystal [93] the most direct approach would be soaks and/or co-crystallization. Although soaking ligands into crystals may be the method of choice, validation of the soaking system with co-crystallization experiments should be done, as sometimes conformational changes are not seen when the ligand is soaked into the crystal. [102] In co-crystallization experiments, the ligand is incubated with the protein to form a complex and afterwards used in crystallization. [102]

MurT has shown to be a difficult protein to express and purify, due to its low level of expression. Recent *in vitro* results (data not shown) have demonstrated that MurT is stabilized if expressed with DUF1727, a domain of unknown function found at its C-terminus. [5] All the efforts in expressing and purifying MurT are now connected with MurT-DUF1727 expression and purification. Once this protocol is optimized, further crystallization experiments will be performed including co-crystallization with GatD. The crystal structures of GatD-MurT will insight the specific residues responsible for the interaction between these two proteins, which will become a target when designing effective inhibitors.

To cross-check the residues important in the interactions between proteins and proteins:ligands, we intend to follow site-directed mutagenesis to better understand the catalytic triad. In this process we will convert into alanine each of the triad's residues and determine the catalytic rate of the enzyme. Expectantly all the data from native crystal structures, protein:ligand crystal structures and biochemical assays will corroborate the mechanistic theory of the catalytic triad in which the glycine-histidine pair act together to generate the nucleophile attack of the carbonyl carbon atom of the peptide bond. Furthermore mutagenesis experiments can also clarify the importance of the oxyanion hole for catalysis and if the Tyr_{97} residue is involved in this tetrahedral structure.

References

[1] B. J. Cha, B. S. Koppetsch, and W. E. Theurkauf, “In vivo analysis of Drosophila bicoid mRNA localization reveals a novel microtubule-dependent axis specification pathway.,” *Cell*, vol. 106, no. 1, pp. 35–46, Jul. 2001.

[2] C.-W. Chang, D. Nashchekin, L. Wheatley, U. Irion, K. Dahlgaard, T. G. Montague, J. Hall, and D. St Johnston, “Anterior-posterior axis specification in Drosophila oocytes: identification of novel bicoid and oskar mRNA localization factors.,” *Genetics*, vol. 188, no. 4, pp. 883–96, Aug. 2011.

[3] J.-R. Huynh and D. St Johnston, “The origin of asymmetry: early polarisation of the Drosophila germline cyst and oocyte.,” *Curr. Biol.*, vol. 14, no. 11, pp. R438–49, Jun. 2004.

[4] H. De Lencastre, S. W. Wu, M. G. Pinho, A. M. Ludovice, S. Filipe, S. Gardete, R. Sobral, S. Gill, M. Chung, and A. Tomasz, “Antibiotic Resistance As a Stress Response: Complete Sequencing of a Large Number of Chromosomal Loci in Staphylococcus aureus Strain COL That Impact on the Expression of Resistance to Methicillin,” *Microb. Drug Resist.*, vol. 5, no. 3, pp. 163–175, Jan. 1999.

[5] T. a Figueiredo, R. G. Sobral, A. M. Ludovice, J. M. F. De Almeida, N. K. Bui, W. Vollmer, H. de Lencastre, and A. Tomasz, “Identification of genetic determinants and enzymes involved with the amidation of glutamic acid residues in the peptidoglycan of Staphylococcus aureus.,” *PLoS Pathog.*, vol. 8, no. 1, p. e1002508, Jan. 2012.

[6] D. Münch, T. Roemer, S. H. Lee, M. Engeser, H. G. Sahl, and T. Schneider, “Identification and in vitro analysis of the GatD/MurT enzyme-complex catalyzing lipid II amidation in Staphylococcus aureus.,” *PLoS Pathog.*, vol. 8, no. 1, p. e1002509, Jan. 2012.

[7] G. Rhodes, *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*, Third Edit. Academic Press is an imprint of Elsevier, 2006, p. 333.

[8] G. Taylor, “The phase problem,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 59, no. 11, pp. 1881–1890, Oct. 2003.

[9] W. Hol, “SOLVING THE PHASE.” Winter, pp. 1–25, 2011.

[10] B. J. Fuller, “Cryoprotectants: the essential antifreezes to protect life in the frozen state.,” *Cryo Letters*, vol. 25, no. 6, pp. 375–88, 2004.

[11] D. Jeruzalmi, “First analysis of macromolecular crystals: biochemistry and x-ray diffraction.,” *Methods Mol. Biol.*, vol. 364, pp. 43–62, Jan. 2007.

[12] M. G. Rossmann and E. Arnold, Eds., *International Tables for Crystallography*, 1st Editio., vol. F. Chester, England: International Union of Crystallography, 2001.

[13] T. Hahn, Ed., *International Tables for Crystallography Volume A: Space-group symmetry*, vol. A. Chester, England: International Union of Crystallography, 2006.

[14] C. J. Humphreys, "The significance of Bragg's law in electron diffraction and microscopy, and Bragg's second law.," *Acta Crystallogr. A.*, vol. 69, no. Pt 1, pp. 45–50, Jan. 2013.

[15] D. Blow, *Outline of Crystallography for Biologists*. Oxford University Press, 2005, p. 236.

[16] B. W. Matthews, "Solvent Content of Protein Crystals," *J. Mol. Biol.*, vol. 33, pp. 491–497, 1968.

[17] K. A. Kantardjieff and B. Rupp, "Matthews coefficient probabilities : Improved estimates for unit cell contents of proteins , DNA , and protein – nucleic acid complex crystals," *Protein Sci.*, vol. 12, pp. 1865–1871, 2003.

[18] B. Rupp, *Biomolecular Crystallography*, 1st ed. Garland Science, 2010.

[19] Z. Li, M. Nimtz, and U. Rinas, "Optimized procedure to generate heavy isotope and selenomethionine-labeled proteins for structure determination using Escherichia coli-based expression systems.," *Appl. Microbiol. Biotechnol.*, vol. 92, no. 4, pp. 823–33, Nov. 2011.

[20] D. E. Tronrud, "Introduction to macromolecular refinement.," *Methods Mol. Biol.*, vol. 364, pp. 231–54, Jan. 2007.

[21] G. J. Kleywegt and T. a Jones, "Model building and refinement practice.," *Methods Enzymol.*, vol. 277, no. 1995, pp. 208–30, Jan. 1997.

[22] M. S. Weiss, "Global indicators of X-ray data quality," *J. Appl. Crystallogr.*, vol. 34, no. 2, pp. 130–135, Apr. 2001.

[23] G. J. Kleywegt, "Validation of protein crystal structures.," *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 56, no. Pt 3, pp. 249–65, Mar. 2000.

[24] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank.," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–42, Jan. 2000.

[25] M. D. Adams, "The Genome Sequence of Drosophila melanogaster," *Science (80-.).*, vol. 287, no. 5461, pp. 2185–2195, Mar. 2000.

[26] L. Cycle, "Development of the Fruit Fly," p. 27.

[27] S. Minakhina and R. Steward, "Axes formation and RNA localization.," *Curr. Opin. Genet. Dev.*, vol. 15, no. 4, pp. 416–21, Aug. 2005.

[28] V. Riechmann and A. Ephrussi, "Par-1 regulates bicoid mRNA localisation by phosphorylating Exuperantia.," *Development*, vol. 131, no. 23, pp. 5897–907, Dec. 2004.

[29] a Ephrussi, L. K. Dickinson, and R. Lehmann, “Oskar organizes the germ plasm and directs localization of the posterior determinant nanos.,” *Cell*, vol. 66, no. 1, pp. 37–50, Jul. 1991.

[30] J. E. Wilhelm and R. D. Vale, “RNA on the move: the mRNA localization pathway.,” *J. Cell Biol.*, vol. 123, no. 2, pp. 269–74, Oct. 1993.

[31] J. E. Wilhelm, J. Mansfield, N. Hom-Booher, S. Wang, C. W. Turck, T. Hazelrigg, and R. D. Vale, “Isolation of a ribonucleoprotein complex involved in mRNA localization in Drosophila oocytes.,” *J. Cell Biol.*, vol. 148, no. 3, pp. 427–40, Feb. 2000.

[32] Scott F Gilbert, “Developmental Biology,” 6th editio., 2000.

[33] D. St Johnston, “Moving messages: the intracellular localization of mRNAs.,” *Nat. Rev. Mol. Cell Biol.*, vol. 6, no. 5, pp. 363–75, May 2005.

[34] T. Berleth, M. Burri, G. Thoma, D. Bopp, S. Richstein, G. Frigerio, M. Noll, and C. Nüsslein-Volhard, “The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo.,” *EMBO J.*, vol. 7, no. 6, pp. 1749–56, Jun. 1988.

[35] M. Wilsch-Bräuninger, H. Schwarz, and C. Nüsslein-Volhard, “A sponge-like structure involved in the association and transport of maternal products during Drosophila oogenesis.,” *J. Cell Biol.*, vol. 139, no. 3, pp. 817–29, Nov. 1997.

[36] D. Marcey, W. S. Watkins, and T. Hazelrigg, “The temporal and spatial distribution pattern of maternal exuperantia protein: evidence for a role in establishment but not maintenance of bicoid mRNA localization.,” *EMBO J.*, vol. 10, no. 13, pp. 4259–66, Dec. 1991.

[37] T. Hazelrigg and C. Tu, “Sex-specific processing of the Drosophila exuperantia transcript is regulated in male germ cells by the tra-2 gene.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, no. 22, pp. 10752–6, Oct. 1994.

[38] T. Hazelrigg, W. S. Watkins, D. Marcey, C. Tu, M. Karow, and X. R. Lin, “The exuperantia gene is required for Drosophila spermatogenesis as well as anteroposterior polarity of the developing oocyte, and encodes overlapping sex-specific transcripts.,” *Genetics*, vol. 126, no. 3, pp. 607–17, Nov. 1990.

[39] J. Glotzer and a Ephrussi, “mRNA localization and the cytoskeleton,” *Semin. Cell Dev. Biol.*, vol. 7, no. 3, pp. 357–365, Jun. 1996.

[40] S. F. Altschul, T. L. Madden, a a Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.,” *Nucleic Acids Res.*, vol. 25, no. 17, pp. 3389–402, Sep. 1997.

[41] S. F. Altschul, J. C. Wootton, E. M. Gertz, R. Agarwala, A. Morgulis, A. A. Schäffer, and Y. Yu, “Protein database searches using compositionally adjusted substitution matrices.,” *FEBS J.*, vol. 272, no. 20, pp. 5101–9, Oct. 2005.

[42] S. Rogers, R. Wells, and M. Rechsteiner, “Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis.,” *Science*, vol. 234, no. 4774, pp. 364–8, Oct. 1986.

[43] R. Linding, “GlobPlot: exploring protein sequences for globularity and disorder,” *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3701–3708, Jul. 2003.

[44] Y. Zhang, “I-TASSER server for protein 3D structure prediction.,” *BMC Bioinformatics*, vol. 9, p. 40, Jan. 2008.

[45] L. Slabinski, L. Jaroszewski, L. Rychlewski, I. a Wilson, S. a Lesley, and A. Godzik, “XtalPred: a web server for prediction of protein crystallizability.,” *Bioinformatics*, vol. 23, no. 24, pp. 3403–5, Dec. 2007.

[46] C. P. a M. Kloks, C. a E. M. Spronk, E. Lasonder, A. Hoffmann, G. W. Vuister, S. Grzesiek, and C. W. Hilbers, “The solution structure and DNA-binding properties of the cold-shock domain of the human Y-box protein YB-1.,” *J. Mol. Biol.*, vol. 316, no. 2, pp. 317–26, Feb. 2002.

[47] A. Marchler-Bauer, C. Zheng, F. Chitsaz, M. K. Derbyshire, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, S. Lu, G. H. Marchler, J. S. Song, N. Thanki, R. a Yamashita, D. Zhang, and S. H. Bryant, “CDD: conserved domains and protein three-dimensional structure.,” *Nucleic Acids Res.*, vol. 41, no. Database issue, pp. D348–52, Jan. 2013.

[48] D. N. Ermolenko and G. I. Makhatadze, “Bacterial cold-shock proteins.,” *Cell. Mol. Life Sci.*, vol. 59, no. 11, pp. 1902–13, Nov. 2002.

[49] L. Lo Conte, B. Ailey, T. J. Hubbard, S. E. Brenner, a G. Murzin, and C. Chothia, “SCOP: a structural classification of proteins database.,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 257–9, Jan. 2000.

[50] G. Horn, R. Hofweber, W. Kremer, and H. R. Kalbitzer, “Structure and function of bacterial cold shock proteins.,” *Cell. Mol. Life Sci.*, vol. 64, no. 12, pp. 1457–70, Jun. 2007.

[51] A. Cléry and A. Frédéric, “From Structure to Function of RNA Binding Domains,” in *RNA Binding Proteins*, Z. J. Lorković, Ed. Landes Bioscience, 2012.

[52] U. B. Ericsson, B. M. Hallberg, G. T. Detitta, N. Dekker, and P. Nordlund, “Thermofluor-based high-throughput stability optimization of proteins for structural studies.,” *Anal. Biochem.*, vol. 357, no. 2, pp. 289–98, Oct. 2006.

[53] L. Reinhard, H. Mayerhofer, A. Geerlof, J. Mueller-Dieckmann, and M. S. Weiss, “Optimization of protein buffer cocktails using Thermofluor.,” *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.*, vol. 69, no. Pt 2, pp. 209–14, Feb. 2013.

[54] J. E. Nettleship, J. Brown, M. R. Groves, and A. Geerlof, “Methods for protein characterization by mass spectrometry, thermal shift (ThermoFluor) assay, and multiangle or static light scattering.,” *Methods Mol. Biol.*, vol. 426, pp. 299–318, Jan. 2008.

[55] J. Jancarik and S. H. Kim, “Sparse matrix sampling: a screening method for crystallization of proteins,” *J. Appl. Crystallogr.*, vol. 24, no. 4, pp. 409–411, Aug. 1991.

[56] F. W. Studier, “Protein production by auto-induction in high-density shaking cultures,” *Protein Expr. Purif.*, vol. 41, no. 1, pp. 207–234, May 2005.

[57] S. M. Singh and A. K. Panda, “Solubilization and refolding of bacterial inclusion body proteins.,” *J. Biosci. Bioeng.*, vol. 99, no. 4, pp. 303–10, Apr. 2005.

[58] R. S. Donovan, C. W. Robinson, and B. R. Glick, “Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter.,” *J. Ind. Microbiol.*, vol. 16, no. 3, pp. 145–54, Mar. 1996.

[59] D. Mattanovich, P. Branduardi, L. Dato, B. Gasser, M. Sauer, and D. Porro, *Recombinant protein production in yeasts.*, vol. 824. 2012, pp. 329–58.

[60] J. Yang and M. Reth, “Drosophila S2 Schneider Cells: A Useful Tool for Rebuilding and Redesigning Approaches in Synthetic Biology,” in *Synthetic Gene Networks: Methods and Protocols*, vol. 813, W. Weber and M. Fussenegger, Eds. Totowa, NJ: Humana Press, 2012, pp. 331–341.

[61] C. Kollewe and A. Vilcinskas, “Production of Recombinant Proteins in Insect Cells,” *Am. J. Biochem. Biotechnol.*, vol. 9, no. 3, pp. 255–271, Mar. 2013.

[62] R. V Lehr, L. C. Elefante, K. K. Kikly, S. P. O’Brien, and R. B. Kirkpatrick, “A modified metal-ion affinity chromatography procedure for the purification of histidine-tagged recombinant proteins expressed in Drosophila S2 cells.,” *Protein Expr. Purif.*, vol. 19, no. 3, pp. 362–8, Aug. 2000.

[63] C. Janke and J. C. Bulinski, “Post-translational regulation of the microtubule cytoskeleton: mechanisms and functions.,” *Nat. Rev. Mol. Cell Biol.*, vol. 12, no. 12, pp. 773–86, Dec. 2011.

[64] K. H. Schleifer and O. Kandler, “Peptidoglycan types of bacterial cell walls and their taxonomic implications.,” *Bacteriol. Rev.*, vol. 36, no. 4, pp. 407–77, Dec. 1972.

[65] J. T. Park and J. L. Strominger, “Mode of Action of Penicillin: Biochemical Basis for the Mechanism of Action of Penicillin and for Its Selective Toxicity,” *Science (80-.).*, vol. 125, no. 3238, pp. 99–101, Jan. 1957.

[66] P. Courvalin, “Vancomycin resistance in gram-positive cocci.,” *Clin. Infect. Dis.*, vol. 42 Suppl 1, pp. S25–34, Jan. 2006.

[67] T. J. Silhavy, D. Kahne, and S. Walker, “The bacterial cell envelope.,” *Cold Spring Harb. Perspect. Biol.*, vol. 2, no. 5, p. a000414, May 2010.

[68] D. Scheffers and M. G. Pinho, “Bacterial cell wall synthesis: new insights from localization studies.,” *Microbiol. Mol. Biol. Rev.*, vol. 69, no. 4, pp. 585–607, Dec. 2005.

[69] J. van Heijenoort, “Formation of the glycan chains in the synthesis of bacterial peptidoglycan.,” *Glycobiology*, vol. 11, no. 3, p. 25R–36R, Mar. 2001.

[70] A. L. Lovering, S. S. Safadi, and N. C. J. Strynadka, “Structural perspective of peptidoglycan biosynthesis and assembly.,” *Annu. Rev. Biochem.*, vol. 81, pp. 451–78, Jan. 2012.

[71] R. Wheeler, “Peptidoglycan architecture and dynamics in Gram-positive bacteria,” University of Sheffield, 2012.

[72] M. S. VanNieuwenhze, S. C. Mauldin, M. Zia-Ebrahimi, J. a Aikins, and L. C. Blaszczak, “The total synthesis of lipid I.,” *J. Am. Chem. Soc.*, vol. 123, no. 29, pp. 6983–8, Jul. 2001.

[73] N. Mills, “ChemDraw Ultra 10.0,” *J. Am. Chem. Soc.*, vol. 128, no. 41, pp. 13649–13650, Oct. 2006.

[74] W. Vollmer, “Bacterial outer membrane evolution via sporulation?,” *Nat. Chem. Biol.*, vol. 8, no. 1, pp. 14–8, Jan. 2012.

[75] S. Sharif, M. Singh, S. J. Kim, and J. Schaefer, “Staphylococcus aureus peptidoglycan tertiary structure from carbon-13 spin diffusion.,” *J. Am. Chem. Soc.*, vol. 131, no. 20, pp. 7023–30, May 2009.

[76] A. Typas, M. Banzhaf, C. a Gross, and W. Vollmer, “From the regulation of peptidoglycan synthesis to bacterial growth and morphology.,” *Nat. Rev. Microbiol.*, vol. 10, no. 2, pp. 123–36, Feb. 2012.

[77] P. A. Rei, “On the track of β-lactam resistance : Studies on the regulation of methicillin-resistance in Staphylococcus aureus,” 2012.

[78] D. M. Livermore, “Antibiotic resistance in staphylococci.,” *Int. J. Antimicrob. Agents*, vol. 16 Suppl 1, pp. S3–10, Nov. 2000.

[79] J. Gustafson, a Strässle, H. Hächler, F. H. Kayser, and B. Berger-Bächi, “The femC locus of Staphylococcus aureus required for methicillin resistance includes the glutamine synthetase operon.,” *J. Bacteriol.*, vol. 176, no. 5, pp. 1460–7, Mar. 1994.

[80] M. Y. Galperin and N. V Grishin, “The synthetase domains of cobalamin biosynthesis amidotransferases cobB and cobQ belong to a new family of ATP-dependent amidoligases, related to dethiobiotin synthetase.,” *Proteins*, vol. 41, no. 2, pp. 238–47, Nov. 2000.

[81] L. E. Bird, “High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli.,” *Methods*, vol. 55, no. 1, pp. 29–37, Sep. 2011.

[82] L. E. Bird, H. Rada, J. Flanagan, J. M. Diprose, R. J. C. Gilbert, and R. J. Owens, “Application of In-Fusion™ cloning for the parallel construction of E. coli expression vectors.,” *Methods Mol. Biol.*, vol. 1116, pp. 209–34, Jan. 2014.

[83] H. Walden, “Selenium incorporation using recombinant techniques.,” *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 66, no. Pt 4, pp. 352–7, Apr. 2010.

[84] G. E. O. Borgstahl, “How to use dynamic light scattering to improve the likelihood of growing macromolecular crystals.,” *Methods Mol. Biol.*, vol. 363, pp. 109–29, Jan. 2007.

[85] B. Lorber, F. Fischer, M. Bailly, H. Roy, and D. Kern, “Protein analysis by dynamic light scattering: methods and techniques for students.,” *Biochem. Mol. Biol. Educ.*, vol. 40, no. 6, pp. 372–82, 2012.

[86] D. Arzensek, R. Podgornik, and D. Kuzman, “Dynamic light scattering and application to proteins in solutions.” Ljubljana, p. 19, 2010.

[87] G. Winter, "Xia2 : an Expert System for Macromolecular Crystallography Data Reduction," *J. Appl. Crystallogr.*, vol. 43, no. 1, pp. 186–190, Dec. 2009.

[88] P. Evans, "Scaling and assessment of data quality.," *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 62, no. Pt 1, pp. 72–82, Jan. 2006.

[89] E. Potterton, P. Briggs, M. Turkenburg, and E. Dodson, "A graphical user interface to the CCP4 program suite research papers A graphical user interface to the CCP 4 program suite," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, pp. 1131–1137, 2003.

[90] G. N. Murshudov, P. Skubák, A. a Lebedev, N. S. Pannu, R. a Steiner, R. a Nicholls, M. D. Winn, F. Long, and A. a Vagin, "REFMAC5 for the refinement of macromolecular crystal structures.," *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 67, no. Pt 4, pp. 355–67, Apr. 2011.

[91] G. N. Murshudov, a a Vagin, and E. J. Dodson, "Refinement of macromolecular structures by the maximum-likelihood method.," *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 53, no. Pt 3, pp. 240–55, May 1997.

[92] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson, "Overview of the CCP4 suite and current developments.," *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 67, no. Pt 4, pp. 235–42, Apr. 2011.

[93] D. Vieira, T. a. Figueiredo, A. Verma, R. G. Sobral, A. M. Ludovice, H. de Lencastre, and J. Trincao, "Purification, crystallization and preliminary X-ray diffraction analysis of GatD, a glutamine amidotransferase-like protein from Staphylococcus aureus peptidoglycan," *Acta Crystallogr. Sect. F Struct. Biol. Commun.*, vol. 70, no. 5, pp. 1–4, Apr. 2014.

[94] R. a Laskowski, "PDBsum new things.," *Nucleic Acids Res.*, vol. 37, no. Database issue, pp. D355–9, Jan. 2009.

[95] A. Douangamath, M. Walker, S. Beismann-Driemeyer, M. C. Vega-Fernandez, R. Sterner, and M. Wilmanns, "Structural evidence for ammonia tunneling across the (beta alpha)(8) barrel of the imidazole glycerol phosphate synthase bienzyme complex.," *Structure*, vol. 10, no. 02, pp. 185–93, Feb. 2002.

[96] L. Holm and P. Rosenström, "Dali server: conservation mapping in 3D.," *Nucleic Acids Res.*, vol. 38, no. Web Server issue, pp. W545–9, Jul. 2010.

[97] Y. Zhang, J. Kua, and J. A. McCammon, "Role of the catalytic triad and oxyanion hole in acetylcholinesterase catalysis: an ab initio QM/MM study.," *J. Am. Chem. Soc.*, vol. 124, no. 35, pp. 10572–7, Sep. 2002.

[98] L.-C. Lee, Y.-L. Lee, R.-J. Leu, and J.-F. Shaw, "Functional role of catalytic triad and oxyanion hole-forming residues on enzyme activity of Escherichia coli thioesterase I/protease I/phospholipase L1.," *Biochem. J.*, vol. 397, no. 1, pp. 69–76, Jul. 2006.

[99] O. Trott and A. J. Olson, "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.," *J. Comput. Chem.*, vol. 31, no. 2, pp. 455–61, Jan. 2010.

[100] Schrodinger LLC, "The PyMOL Molecular Graphics System, Version 1.3r1." 2010.

[101] K. Thangavelu, Q. Y. Chong, B. C. Low, and J. Sivaraman, "Structural basis for the active site inhibition mechanism of human kidney-type glutaminase (KGA).," *Sci. Rep.*, vol. 4, p. 3827, Jan. 2014.

[102] A. M. Hassell, G. An, R. K. Bledsoe, J. M. Bynum, H. L. Carter, S.-J. J. Deng, R. T. Gampe, T. E. Grisard, K. P. Madauss, R. T. Nolte, W. J. Rocque, L. Wang, K. L. Weaver, S. P. Williams, G. B. Wisely, R. Xu, and L. M. Shewchuk, "Crystallization of protein-ligand complexes.," *Acta Crystallogr. D. Biol. Crystallogr.*, vol. 63, no. Pt 1, pp. 72–9, Jan. 2007.

[103] C. a O'Callaghan, J. Tormo, B. E. Willcox, C. D. Blundell, B. K. Jakobsen, D. I. Stuart, a J. McMichael, J. I. Bell, and E. Y. Jones, "Production, crystallization, and preliminary X-ray analysis of the human MHC class Ib molecule HLA-E.," *Protein Sci.*, vol. 7, no. 5, pp. 1264–6, May 1998.

[104] M. E. Goldberg, N. Expert-Bezançon, L. Vuillard, and T. Rabilloud, "Non-detergent sulphobetaines: a new class of molecules that facilitate in vitro protein renaturation.," *Fold. Des.*, vol. 1, no. 1, pp. 21–7, Jan. 1996.

[105] D. Evangelopous and J. D. Guzman, "The Art of Protein Crystallization," *Methods*, vol. 6, no. 27, pp. 74–76, 2010.

Appendix

A. DNA and protein sequences

A.1. Translation of exuperantia (*exu*)

Species: *Drosophila melanogaster* (fruit fly)

DNA Sequence:

ATGGTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCCG
TGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCAC
TGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGTGTTA
TTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAATCCAAGTC
TGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACCAAGGCGGGTCCC
AGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATTCCCTACATGATCCTGG
AGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCGGTGAAGTCGTTTGCAAACAG
CATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATAAAGAACTACAGTCTGCGAAAGCTT
TCCAAGATCTTGTCCACGACCAAGGAGGAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTG
GATCTGGCTTGGGATCTGGATCTAGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAG
TACCGTGACCAATGGTGACGATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTG
TTTGATGGCAATGCCAGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACT
CCGACGGCAAGCCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCC
GTTTGCCAAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGT
CAAAACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTGCGTG
CCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGTGCCATCTG
GGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGGCTAAAGAGCAAG
GACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCCAAGAAGACCACGGTGAAGC
CGGTTGTCAAGGGTAACAGCAACAATAATAACAATTATCGTCGTCGCAATCGACGCGGCGGCCG
CCAATCGGTGAAGGACGCTAGGCCATCCAGCTCCCCATCGGCTAGCACCGAGTTTGGAGCAGGA
GGTGACAAGTCGCGCAGCGTTTCCTCGTTGCCCGACTCTACTACCAAGACTCCCTCGCCAAACA
AGCCGCGTATGCACCGTAAGCGCAACTCGCGTCAAAGCTTGGGTGCAACCCCCAATGGACTCAA
GGTTGCTGCTGAAATATCCTCCTCAGGCGTGTCTGAGCTCAACAATAGCGCTCCACCTGCCGTA
ACCATAAGCCCAGTCGTGGCACAACCCTCACCGACACCGGTGGCCATCACGGCCTCCAAC

Protein Sequence:

MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFE
QYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGP
SSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKL
SKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNGDDKQSSKNAVQGKREL
FDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPFAKLVVSDVLELDIQIENLER
QNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLSAIWAEKNIEGLDIALQSIGRLKSK
DKAELLELLDSYFDPKKTTVKPVVKGNSNNNNNYRRRNRRGGRQSVKDARPSSSPSASTEFGAG
GDKSRSVSSLPDSTTKTPSPNKPRMHRKRNSRQSLGATPNGLKVAAEISSSGVSELNNSAPPAV
TISPVVAQPSPTPVAITASN

A.2. Translation of ypsilon-schachtel (*yps*)

Species: *Drosophila melanogaster* (fruit fly)

DNA Sequence:

ATGGCTGATGCCGCGGAGAGTAAGCCACTGGCCGCCGAACAGCAGCAGGCGCAGCAGCAG
CCGGAGCAGCAGCAGAATCCGCCGAATCCGCAGGAGCAGGATCACGAGCAGGAGCCGCTG
GACGAGCTGCAGGGACAGCAGGGCCAGCCCGCTCCGCCCACCAAGGAAGTCATCGCCACC
AAAGTCACCGGCACCGTCAAGTGGTTCAACGTGAAGAGCGGCTACGGCTTCATCAACCGC
AACGACACCAGAGAGGATGTCTTTGTGCACCAGAGCGCCATTGCGCGGAACAACCCCAAA
AAGGCGGTCCGCTCGGTGGGCGACGGTGAGGTCGTTGAGTTCGACGTGGTCATTGGTGAG
AAGGGCAACGAGGCGGCCAACGTGACCGGTCCCTCCGGTGAGCCGGTGCGGGGCAGTCAG
TTTGCAGCGGACAAGCGCCGTAACTTCCGTCCCTGGATGAAGAAGAATCGCCGCAAGGAT
GGCGAAGTGGAAGGCGAAGACGCCGAATCGTCGGCCCAGCAGCAACAGCAGCAGGCGGCA
CCAATCGTTGACGGGCAGCCGCAGCAGCAGGTGCAGTCCGGCCCGCGTCAACCACGACAG
AACTTCCGCCGAGGACCACCCGGTGGACCGCCCGGAGGACCTCGAGGTGGCCCCCGAGGA
CCGCCTGGTGGAGCTCCCGGTGGCCCCCGGCGCTACAACAACTATTATCTGCGTCAGCCG
AGACGCGGCCTGGGCGGCGGTGACGGCAGCGCCGAGCCCGGTGTCCACGATCAGAATCCC
GAGGGTCTGCAGCGCGGCGAGGGTCAGGGACCACGTCGCGGTGGAGGCCCACCGGGCGGA
CCCCAGAGGCGCTTCTTCCGACGCAACTTCAACAATGGTCCACCGCCACCACGCCGCGAC
GGCGGAGAATACATTCAAGGCCAGGGACCGCCACGCCCACAACAGCCACGTCCACGTCGC
CAGAGGAAACCAAATGGCCCTGGCGGTGGTTCGGAGCAGCAGCCAGAGAAGAACGGCGCT
CAAGAGCTGCAAAATACAACCACAGAGAGCACTGCA

Protein Sequence:

MADAAESKPLAAEQQQAQQQPEQQQNPPNPQEQDHEQEPLDELQGQQGQPAPPTKEVIAT
KVTGTVKWFNVKSGYGFINRNDTREDVFVHQSAIARNNPKKAVRSVGDGEVVEFDVVIGE
KGNEAANVTGPSGEPVRGSQFAADKRRNFRPWMKKNRRKDGEVEGEDAESSAQQQQQQAA
PIVDGQPQQQVQSGPRQPRQNFRRGPPGGPPGGPRGGPRGPPGGAPGGPRRYNNYYLRQP
RRGLGGGDGSAEPGVHDQNPEGLQRGEGQGPRRGGGPPGGPQRRFFRRNFNNGPPPPRRD
GGEYIQGQGPPRPQQPRPRRQRKPNGPGGGSEQQPEKNGAQELQNTTTESTA

A.3. Translation of gatD

Species: *Staphylococcus aureus*

DNA Sequence:

ATGCATGAATTGACTATTTATCATTTTATGTCAGATAAATTGAATTTATACAGTGATATA
GGAAATATTATTGCTTTAAGACAACGTGCTAAAAAACGAAATATTAAAGTTAATGTCGTA
GAAATCAATGAAACAGAAGGTATTACCTTTGATGAATGTGATATTTTCTTTATCGGTGGT
GGAAGTGATAGAGAACAAGCATTAGCAACAAAAGAATTAAGTAAAATTAAGACACCACTT
AAAGAAGCGATTGAAGATGGTATGCCGGGATTAACGATTTGTGGAGGCTATCAATTTTTA
GGGAAAAAATATATCACGCCTGATGGTACAGAATTAGAAGGGTTAGGTATTTTAGATTTT
TATACTGAATCAAAGACAAACCGATTAACAGGAGATATTGTTATCGAAAGTGATACTTTT
GGAACTATTGTAGGTTTTGAAAATCACGGTGGTAGAACATATCATGATTTCGGTACACTT
GGTCATGTTACTTTTGGTTATGGTAATAATGATGAAGATAAAAAAGAAGGCATTCATTAT
AAAAATTTATTAGGTACTTATTTACATGGACCAATTTTACCTAAAAATTACGAAATCACT
GATTATCTGTTAGAAAAAGCTTGTGAACGTAAGGGTATTCCGTTTGAGCCTAAAGAAATA
GATAATGAAGCGGAAATACAAGCGAAACAAGTATTAATAGACAGAGCAAATAGACAGAAG
AAATCTCGT

Protein sequence:

MHELTIYHFMSDKLNLYSDIGNIIALRQRAKKRNIKVNVVEINETEGITFDECDIFFIGG
GSDREQALATKELSKIKTPLKEAIEDGMPGLTICGGYQFLGKKYITPDGTELEGLGILDF
YTESKTNRLTGDIVIESDTFGTIVGFENHGGRTYHDFGTLGHVTFGYGNNDEDKKEGIHY
KNLLGTYLHGPILPKNYEITDYLLEKACERKGIPFEPKEIDNEAEIQAKQVLIDRANRQK
KSR

A.4. Translation of murT

Species: *Staphylococcus aureus*

DNA Sequence:

ATGAGACAGTGGACGGCAATCCATCTAGCGAAATTGGCGCGTAAAGCAAGTAGAGCAGTA
GGTAAAAGAGGAACAGATTTACCTGGACAAATCGCTAGAAAAGTGGATACAGATATATTA
AGAAAATTAGCAGAGCAAGTTGATGATATTGTATTTATCAGTGGAACAAATGGTAAAACA
ACGACTTCAAACTTAATTGGACATACTTTAAAAGCAAATAATATTCAAATTATACACAAT
AATGAAGGTGCTAATATGGCTGCAGGTATAACTTCTGCATTCATCATGCAATCAACACCT
AAGACTAAAATTGCGGTAATCGAAATTGATGAAGGTTCGATTCCACGTGTGTTAAAAGAA
GTTACACCTTCAATGATGGTATTTACTAATTTCTTTAGAGATCAAATGGATCGCTTCGGT
GAAATTGATATTATGGTTAATAACATTGCAGAGACAATTAGTAATAAAGGCATCAAATTA
TTGCTAAATGCTGATGATCCATTTGTGAGTCGTTTGAAAATCGCAAGTGATACGATTGTG
TACTATGGTATGAAAGCACATGCCCATGAATTTGAACAAAGTACGATGAATGAAAGTAGA
TATTGTCCAAACTGTGGTCGCTTATTGCAATACGATTATATTCATTATAATCAAATTGGT
CATTATCACTGTCAGTGTGGTTTCAAACGAGAGCAAGCAAAATATGAAATATCAAGTTTT
GATGTGGCACCGTTTTTATATTTAAATATCAATGATGAAAAATATGATATGAAAATTGCA
GGTGACTTTAACGCTTATAACGCGTTAGCAGCATATACTGTTTTAAGAGAGCTAGGGTTA
AATGAACAAACAATTAAAAATGGCTTTGAAACGTATACATCAGACAATGGTCGTATGCAG
TACTTTAAAAAAGAACGAAAAGAAGCGATGATCAATTTAGCTAAAAATCCTGCAGGAATG
AATGCAAGTTTATCAGTTGGTGAACAATTAGAAGGCGAAAAAGTGTATGTTATTTCGCTA
AATGATAACGCTGCAGATGGTCGAGATACTTCATGGATTTATGATGCAGATTTTGAAAAA
TTATCTAAGCAACAAATTGAAGCTATCATCGTGACAGGTACACGAGCAGAAGAACTTCAA
TTGCGATTGAAGTTAGCAGAGGTTGAAGTACCAATTATAGTTGAGCGTGATATTTATAAA
GCAACGGCAAAGACTATGGATTATAAAGGTTTCACAGTTGCAATACCAAACTATACATCA
TTAGCGCCTATGCTTGAACAATTAAACCGTTCGTTTGAAGGAGGTCAATCA

Protein sequence:

MRQWTAIHLAKLARKASRAVGKRGTDLPGQIARKVDTDILRKLAEQVDDIVFISGTNGKT
TTSNLIGHTLKANNIQIIHNNEGANMAAGITSAFIMQSTPKTKIAVIEIDEGSIPRVLKE
VTPSMMVFTNFFRDQMDRFGEIDIMVNNIAETISNKGIKLLLNADDPFVSRLKIASDTIV
YYGMKAHAHEFEQSTMNESRYCPNCGRLLQYDYIHYNQIGHYHCQCGFKREQAKYEISSF
DVAPFLYLNINDEKYDMKIAGDFNAYNALAAYTVLRELGLNEQTIKNGFETYTSDNGRMQ
YFKKERKEAMINLAKNPAGMNASLSVGEQLEGEKVYVISLNDNAADGRDTSWIYDADFEK
LSKQQIEAIIVTGTRAEELQLRLKLAEVEVPIIVERDIYKATAKTMDYKGFTVAIPNYTS
LAPMLEQLNRSFEGGQS

B. Exuperantia and Ypsilon-Schachtel cloning and expression assays

The following tables combine all the genes, vectors, *E. coli* expression strains and protein expression induction systems tried in the experiments carried out in this thesis.
The results range from no protein expression (x), low level of protein expression (†) and protein expression suitable for purification and crystallization assays (‡).

Vectors	**Genes**	***E. coli* expression strains**	**Induction methods**	
			IPTG	**Auto-induction**
pGEX-6P	Exu205C	BL21	†	
	Exu399C		x	
	Exu531C		x	
pET-15b	Exu205C		x	x
	Exu399C		x	x
	Exu531C		x	x
pET-14b	Exu205C		x	x
	Exu399C		x	x
	Exu531C		x	x
pET-28a	Exu205C		x	x
	Exu210C		x	†
	Exu399C		x	x
	Exu531C		x	x
pET-Sumo-28a	Exu210C	BL21 (DE3)	x	x
		BL21 Star(DE3)pLysS	x	x
		BL21 Gold(DE3)pLysS	x	x
		Rosetta2(DE3)pLysS	x	x
pET-MAL-28b-Prescission		BL21 (DE3)	x	x
		BL21 Star(DE3)pLysS	x	x
		BL21 Gold(DE3)pLysS	x	x
		Rosetta2(DE3)pLysS	x	x
pET-MAL-29b-Prescission		BL21 (DE3)	x	x
		BL21 Star(DE3)pLysS	x	x
		BL21 Gold(DE3)pLysS	x	x
		Rosetta2(DE3)pLysS	x	x
pET-GFP-19b-TEV		BL21 (DE3)	x	x
		BL21 Star(DE3)pLysS	x	x
		BL21 Gold(DE3)pLysS	x	x
		Rosetta2(DE3)pLysS	x	x

Vectors	Genes	*E. coli* expression strains	Induction methods	
			IPTG	Auto-induction
pET-28a	sExu210C	BL21 (DE3)	x	x
		BL21 Star(DE3)pLysS	x	†
		BL21 Gold(DE3)pLysS	x	x
		Rosetta2(DE3)pLysS	x	x
pET-Sumo-28a		BL21 (DE3)	x	x
		BL21 Star(DE3)pLysS	x	x
		BL21 Gold(DE3)pLysS	x	x
		Rosetta2(DE3)pLysS	x	x
pET-MAL-28b-Prescission		BL21 (DE3)	x	x
		BL21 Star(DE3)pLysS	x	x
		BL21 Gold(DE3)pLysS	x	x
		Rosetta2(DE3)pLysS	x	x
pET-MAL-29b-Prescission		BL21 (DE3)	x	x
		BL21 Star(DE3)pLysS	x	x
		BL21 Gold(DE3)pLysS	x	x
		Rosetta2(DE3)pLysS	x	x
pET-GFP-19b-TEV		BL21 (DE3)	x	x
		BL21 Star(DE3)pLysS	x	x
		BL21 Gold(DE3)pLysS	x	x
		Rosetta2(DE3)pLysS	x	x
pET-15b	Yps_CSD	BL21	x	x
pET-28a			†	†
pET-Sumo-28a		BL21 (DE3)	†	†
		BL21 Star(DE3)pLysS	†	‡
		BL21 Gold(DE3)pLysS	†	†
		Rosetta2(DE3)pLysS	†	†
pET-MAL-28b-Prescission		BL21 (DE3)	x	x
		BL21 Star(DE3)pLysS	x	x
		BL21 Gold(DE3)pLysS	x	x
		Rosetta2(DE3)pLysS	x	x
pET-MAL-29b-Prescission		BL21 (DE3)	x	x
		BL21 Star(DE3)pLysS	x	x
		BL21 Gold(DE3)pLysS	x	x
		Rosetta2(DE3)pLysS	x	x
pET-GFP-19b-TEV		BL21 (DE3)	x	x
		BL21 Star(DE3)pLysS	†	†
		BL21 Gold(DE3)pLysS	†	†
		Rosetta2(DE3)pLysS	x	x

Name: Exu205C
Amplified region: 1 – 615
Vector: pGEX-6P, pET-15b, pET-14b, pET-28a

DNA Sequence:

ATGGTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCG
CCCGTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATC
GACACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACC
GACCACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGT
CATCAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTAT
AAGATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAG
CAACTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAG
CGCAAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGC
TTCACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGC
GATGCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAG
GAGGAAGACGCGGCT

Protein Sequence (untagged):

MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPT
DHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLE
QLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIG
DANIKNYSLRKLSKILSTTKEEDAA

Oligonucleotide sequence:

Forward primer	GACGGATCCCATATGGTTGCCGATAACATCG
Reverse Primer	GTCGAATTCGGATCCTTAAGCCGCGTCTTCCTC

Name: Exu210C
Amplified region: 1 – 630
Vector: pET-28a

DNA Sequence:

ATGGTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCG
CCCGTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATC
GACACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACC
GACCACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGT
CATCAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTAT
AAGATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAG
CAACTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAG
CGCAAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGC
TTCACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGC
GATGCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAG
GAGGAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPT
DHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLE
QLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIG
DANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer	GACGGATCCCATATGGTTGCCGATAACATCG
Reverse Primer	CACCTCGAGTTAATGTGGACGCAGAGCAAGCC

Name: Exu210C
Amplified region: 1 – 630
Vector: pET-Sumo-28a

DNA Sequence:

ATGGTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCG
CCCGTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATC
GACACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACC
GACCACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGT
CATCAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTAT
AAGATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAG
CAACTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAG
CGCAAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGC
TTCACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGC
GATGCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAG
GAGGAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPT
DHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLE
QLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIG
DANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer	CTCGGATCCATGGTTGCCGATAACATCGATG
Reverse Primer	CACCTCGAGTTAATGTGGACGCAGAGCAAGCC

Name: Exu210C
Amplified region: 1 – 630
Vector: pET-MAL-28b-Prescission

DNA Sequence:

ATGGTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCG
CCCGTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATC
GACACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACC
GACCACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGT
CATCAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTAT
AAGATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAG
CAACTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAG
CGCAAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGC
TTCACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGC
GATGCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAG
GAGGAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPT
DHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLE
QLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIG
DANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer	CTCGGATCCATGGTTGCCGATAACATCGATG
Reverse Primer	CACCTCGAGTTAATGTGGACGCAGAGCAAGCC

Name: Exu210C
Amplified region: 1 – 630
Vector: pET-MAL-29b-Prescission

DNA Sequence:

ATGGTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCG
CCCGTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATC
GACACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACC
GACCACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGT
CATCAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTAT
AAGATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAG
CAACTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAG
CGCAAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGC
TTCACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGC
GATGCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAG
GAGGAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPT
DHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLE
QLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIG
DANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer	CTCGGATCCATGGTTGCCGATAACATCGATG
Reverse Primer	CACCTCGAGATGTGGACGCAGAGCAAGCC

Name: Exu210C
Amplified region: 1 – 630
Vector: pET-GFP-19b-TEV

DNA Sequence:

ATGGTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCG
CCCGTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATC
GACACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACC
GACCACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGT
CATCAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTAT
AAGATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAG
CAACTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAG
CGCAAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGC
TTCACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGC
GATGCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAG
GAGGAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPT
DHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLE
QLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIG
DANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer	CTCCATATGATGGTTGCCGATAACATCGATG
Reverse Primer	CACCTCGAGTTAATGTGGACGCAGAGCAAGCC

Name: Exu399C
Amplified region: 1 – 1197
Vector: pGEX-6P, pET-15b, pET-14b, pET-28a

DNA Sequence:

ATGGTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCG
CCCGTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATC
GACACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACC
GACCACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGT
CATCAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTAT
AAGATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAG
CAACTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAG
CGCAAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGC
TTCACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGC
GATGCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAG
GAGGAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGA
TCTAGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAGTACCGTGACCAATGGT
GACGATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGATGGCAAT
GCCAGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGC
AAGCCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTT
GCCAAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGT
CAAAACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTG
CGTGCCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGT
GCCATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGG
CTAAAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCC

Protein Sequence (untagged):

MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPT
DHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLE
QLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIG
DANIKNYSLRKLSKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNG
DDKQSSKNAVQGKRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPF
AKLVVSDVLELDIQIENLERQNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLS
AIWAEKNIEGLDIALQSIGRLKSKDKAELLELLDSYFDP

Oligonucleotide sequence:

Forward primer	GACGGATCCCATATGGTTGCCGATAACATCG
Reverse Primer	GTCGAATTCGGATCCTTAGGGATCGAAATAGCT

Name: Exu531C
Amplified region: 1 – 1593
Vector: pGEX-6P, pET-15b, pET-14b, pET-28a

DNA Sequence:

ATGGTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCG
CCCGTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATC
GACACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACC
GACCACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGT
CATCAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTAT
AAGATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAG
CAACTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAG
CGCAAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGC
TTCACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGC
GATGCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAG
GAGGAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGA
TCTAGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAGTACCGTGACCAATGGT
GACGATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGATGGCAAT
GCCAGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGC
AAGCCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTT
GCCAAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGT
CAAAACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTG
CGTGCCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGT
GCCATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGG
CTAAAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCCAAG
AAGACCACGGTGAAGCCGGTTGTCAAGGGTAACAGCAACAATAATAACAATTATCGTCGT
CGCAATCGACGCGGCGGCCGCCAATCGGTGAAGGACGCTAGGCCATCCAGCTCCCCATCG
GCTAGCACCGAGTTTGGAGCAGGAGGTGACAAGTCGCGCAGCGTTTCCTCGTTGCCCGAC
TCTACTACCAAGACTCCCTCGCCAAACAAGCCGCGTATGCACCGTAAGCGCAACTCGCGT
CAAAGCTTGGGTGCAACCCCCAATGGACTCAAGGTTGCTGCTGAAATATCCTCCTCAGGC
GTGTCTGAGCTCAACAATAGCGCTCCACCTGCCGTAACCATAAGCCCAGTCGTGGCACAA
CCCTCACCGACACCGGTGGCCATCACGGCCTCC

Protein Sequence (untagged):

MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPT
DHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLE
QLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIG
DANIKNYSLRKLSKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNG
DDKQSSKNAVQGKRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPF
AKLVVSDVLELDIQIENLERQNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLS
AIWAEKNIEGLDIALQSIGRLKSKDKAELLELLDSYFDPKKTTVKPVVKGNSNNNNNYRR
RNRRGGRQSVKDARPSSSPSASTEFGAGGDKSRSVSSLPDSTTKTPSPNKPRMHRKRNSR
QSLGATPNGLKVAAEISSSGVSELNNSAPPAVTISPVVAQPSPTPVAITAS

Oligonucleotide sequence:

Forward primer	GACGGATCCCATATGGTTGCCGATAACATCG
Reverse Primer	GTCGAATTCGGATCCTTAGTTGGAGGCCGTG

Name: sExu210C (synthetic gene)
Vector: pET-Sumo-28a, pET-MAL-28b-Prescission

DNA Sequence:

ATGGTTGCGGATAACATTGATGCTGGCGTTGCGATTGCGGTGGCTGACCAAAGCTCGTCC
CCGGTGGGCGATAAAGTGGAACTGCCGGCGGGCAACTACATTCTGGTGGGTGTTGATATC
GACACCACGGGCCGTCGCCTGATGGATGAAATTGTGCAGCTGGCGGCCTATACCCCGACG
GACCATTTTGAACAATATATCATGCCGTACATGAACCTGAATCCGGCAGCTCGTCAGCGC
CACCAAGTCCGTGTGATTAGTATCGGTTTTTATCGCATGCTGAAATCAATGCAGACCTAC
AAAATCATCAAAAGTAAATCCGAAATCGCGGCCCTGAAAGATTTCCTGAATTGGCTGGAA
CAACTGAAAACGAAAGCAGGCCCGAGCTCTGACGGTATTGTTCTGATCTATCATGAAGAA
CGTAAATTCATTCCGTACATGATCCTGGAAAGCCTGAAAAAATATGGCCTGCTGGAACGC
TTTACCGCGTCAGTCAAATCGTTCGCCAACAGCATCAATCTGGCACAGGCTTCTATTGGT
GATGCGAACATCAAAAATTACTCCCTGCGTAAACTGTCAAAAATCCTGTCCACCACCAAA
GAAGAAGACGCCGCCTGTTCCGCCTCAACC

Protein Sequence (untagged):

MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPT
DHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLE
QLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAQASIG
DANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer	CTCGGATCCATGGTTGCCGATAACATCGATG
Reverse Primer	CACCTCGAGTTAATGTGGACGCAGAGCAAGCC

Name: sExu210C (synthetic gene)
Vector: pET-MAL-29b-Prescission

DNA Sequence:

ATGGTTGCGGATAACATTGATGCTGGCGTTGCGATTGCGGTGGCTGACCAAAGCTCGTCC
CCGGTGGGCGATAAAGTGGAACTGCCGGCGGGCAACTACATTCTGGTGGGTGTTGATATC
GACACCACGGGCCGTCGCCTGATGGATGAAATTGTGCAGCTGGCGGCCTATACCCCGACG
GACCATTTTGAACAATATATCATGCCGTACATGAACCTGAATCCGGCAGCTCGTCAGCGC
CACCAAGTCCGTGTGATTAGTATCGGTTTTTATCGCATGCTGAAATCAATGCAGACCTAC
AAAATCATCAAAAGTAAATCCGAAATCGCGGCCCTGAAAGATTTCCTGAATTGGCTGGAA
CAACTGAAAACGAAAGCAGGCCCGAGCTCTGACGGTATTGTTCTGATCTATCATGAAGAA
CGTAAATTCATTCCGTACATGATCCTGGAAAGCCTGAAAAAATATGGCCTGCTGGAACGC
TTTACCGCGTCAGTCAAATCGTTCGCCAACAGCATCAATCTGGCACAGGCTTCTATTGGT
GATGCGAACATCAAAAATTACTCCCTGCGTAAACTGTCAAAAATCCTGTCCACCACCAAA
GAAGAAGACGCCGCCTGTTCCGCCTCAACC

Protein Sequence (untagged):

MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPT
DHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLE
QLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAQASIG
DANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer	CTCGGATCCATGGTTGCCGATAACATCGATG
Reverse Primer	CACCTCGAGATGTGGACGCAGAGCAAGCC

Name: sExu210C (synthetic gene)
Vector: pET-GFP-19b-TEV

DNA Sequence:

ATGGTTGCGGATAACATTGATGCTGGCGTTGCGATTGCGGTGGCTGACCAAAGCTCGTCC
CCGGTGGGCGATAAAGTGGAACTGCCGGCGGGCAACTACATTCTGGTGGGTGTTGATATC
GACACCACGGGCCGTCGCCTGATGGATGAAATTGTGCAGCTGGCGGCCTATACCCCGACG
GACCATTTTGAACAATATATCATGCCGTACATGAACCTGAATCCGGCAGCTCGTCAGCGC
CACCAAGTCCGTGTGATTAGTATCGGTTTTTATCGCATGCTGAAATCAATGCAGACCTAC
AAAATCATCAAAAGTAAATCCGAAATCGCGGCCCTGAAAGATTTCCTGAATTGGCTGGAA
CAACTGAAAACGAAAGCAGGCCCGAGCTCTGACGGTATTGTTCTGATCTATCATGAAGAA
CGTAAATTCATTCCGTACATGATCCTGGAAAGCCTGAAAAAATATGGCCTGCTGGAACGC
TTTACCGCGTCAGTCAAATCGTTCGCCAACAGCATCAATCTGGCACAGGCTTCTATTGGT
GATGCGAACATCAAAAATTACTCCCTGCGTAAACTGTCAAAAATCCTGTCCACCACCAAA
GAAGAAGACGCCGCCTGTTCCGCCTCAACC

Protein Sequence (untagged):

MVADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPT
DHFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLE
QLKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAQASIG
DANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer	CTCCATATGATGGTTGCCGATAACATCGATG
Reverse Primer	CACCTCGAGTTAATGTGGACGCAGAGCAAGCC

C. OPPF construct design – Exuperantia and Ypsilon-Schachtel

	OPPF Number	Gene name	aa_N	aa_C	Vector
1	12511	Exuperantia	2	402	pOPINE-3C-HALO7
2	12510				pOPINE-3C-eGFP
3	12509				pOPINRSJ
4	12508				pOPINTF
5	12507				pOPINS3C
6	12506				pOPINM
7	12505				pOPINJ
8	12504				pOPINHALO
9	12503				pOPINF
10	12502				pOPINE
11	12501			210	pOPINE-3C-HALO7
12	12500				pOPINE-3C-eGFP
13	12499				pOPINRSJ
14	12498				pOPINTF
15	12497				pOPINS3C
16	12496				pOPINM
17	12495				pOPINJ
18	12494				pOPINHALO
19	12493				pOPINF
20	12492				pOPINE
21	12491			193	pOPINE-3C-HALO7
22	12490				pOPINE-3C-eGFP
23	12489				pOPINRSJ
24	12488				pOPINTF
25	12487				pOPINS3C
26	12486				pOPINM
27	12485				pOPINJ
28	12484				pOPINHALO
29	12483				pOPINF
30	12482				pOPINE
31	12481			531	pOPINF
32	12538		25	210	pOPINE-3C-HALO7
33	12537				pOPINE-3C-eGFP
34	12536				pOPINRSJ
35	12535				pOPINTF
36	12534				pOPINS3C
37	12533				pOPINM
38	12532				pOPINJ
39	12531				pOPINHALO
40	12530				pOPINF
41	12529				pOPINE
42	12528			193	pOPINE-3C-HALO7
43	12527				pOPINE-3C-eGFP
44	12526				pOPINRSJ
45	12525				pOPINTF
46	12524				pOPINS3C
47	12523				pOPINM
48	12522				pOPINJ
49	12521				pOPINHALO
50	12520				pOPINF
51	12519				pOPINE
52	12481			531	pOPINF

1	12515	Ypsilon-Schachtel	55	171	pOPINRSJ
2	12514				pOPINS3C
3	12513				pOPINF
4	12512				pOPINE

Construct #1

Name: OPPF12511
Amplified region: 4 - 1206
Vector: pOPINE-3C-HALO7
Description: C-HALO 402C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGATCT
AGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAGTACCGTGACCAATGGTGAC
GATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGATGGCAATGCC
AGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGCAAG
CCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTTGCC
AAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGTCAA
AACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTGCGT
GCCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGTGCC
ATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGGCTA
AAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCCAAGAAG
ACC

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLSKILSTTKEEDA
ACSAST

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGGTTGCCGATAACATCGATGCCG
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTGGTCTTCTTGGGATCGAAATAGCTGTCC

Construct #2

Name: OPPF12510
Amplified region: 4 - 1206
Vector: pOPINE-3C-eGFP
Description: C-GFP 402C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGATCT
AGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAGTACCGTGACCAATGGTGAC
GATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGATGGCAATGCC
AGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGCAAG
CCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTTGCC
AAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGTCAA
AACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTGCGT
GCCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGTGCC
ATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGGCTA
AAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCCAAGAAG
ACC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNGD
DKQSSKNAVQGKRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPFA
KLVVSDVLELDIQIENLERQNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLSA
IWAEKNIEGLDIALQSIGRLKSKDKAELLELLDSYFDPKKT

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGGTTGCCGATAACATCGATGCCG
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTGGTCTTCTTGGGATCGAAATAGCTGTCC

Construct #3

Name: OPPF12509
Amplified region: 4 - 1206
Vector: pOPINRSJ
Description: N-HIS (Kan^R) 402C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGATCT
AGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAGTACCGTGACCAATGGTGAC
GATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGATGGCAATGCC
AGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGCAAG
CCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTTGCC
AAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGTCAA
AACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTGCGT
GCCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGTGCC
ATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGGCTA
AAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCCAAGAAG
ACC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNGD
DKQSSKNAVQGKRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPFA
KLVVSDVLELDIQIENLERQNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLSA
IWAEKNIEGLDIALQSIGRLKSKDKAELLELLDSYFDPKKT

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGTCTTCTTGGGATCGAAATAGCTGTCC

Construct #4

Name: OPPF12508
Amplified region: 4 - 1206
Vector: pOPINTF
Description: N-TF 402C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGATCT
AGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAGTACCGTGACCAATGGTGAC
GATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGATGGCAATGCC
AGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGCAAG
CCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTTGCC
AAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGTCAA
AACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTGCGT
GCCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGTGCC
ATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGGCTA
AAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCCAAGAAG
ACC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNGD
DKQSSKNAVQGKRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPFA
KLVVSDVLELDIQIENLERQNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLSA
IWAEKNIEGLDIALQSIGRLKSKDKAELLELLDSYFDPKKT

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGTCTTCTTGGGATCGAAATAGCTGTCC

Construct #5

Name: OPPF12507
Amplified region: 4 - 1206
Vector: pOPINS3C
Description: N-SUMO 402C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGATCT
AGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAGTACCGTGACCAATGGTGAC
GATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGATGGCAATGCC
AGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGCAAG
CCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTTGCC
AAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGTCAA
AACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTGCGT
GCCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGTGCC
ATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGGCTA
AAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCCAAGAAG
ACC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNGD
DKQSSKNAVQGKRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPFA
KLVVSDVLELDIQIENLERQNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLSA
IWAEKNIEGLDIALQSIGRLKSKDKAELLELLDSYFDPKKT

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGTCTTCTTGGGATCGAAATAGCTGTCC

Construct #6

Name: OPPF12506
Amplified region: 4 - 1206
Vector: pOPINM
Description: N-MBP 402C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGATCT
AGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAGTACCGTGACCAATGGTGAC
GATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGATGGCAATGCC
AGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGCAAG
CCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTTGCC
AAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGTCAA
AACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTGCGT
GCCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGTGCC
ATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGGCTA
AAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCCAAGAAG
ACC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNGD
DKQSSKNAVQGKRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPFA
KLVVSDVLELDIQIENLERQNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLSA
IWAEKNIEGLDIALQSIGRLKSKDKAELLELLDSYFDPKKT

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGTCTTCTTGGGATCGAAATAGCTGTCC

Construct #7

Name: OPPF12505
Amplified region: 4 - 1206
Vector: pOPINJ
Description: N-GST 402C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGATCT
AGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAGTACCGTGACCAATGGTGAC
GATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGATGGCAATGCC
AGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGCAAG
CCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTTGCC
AAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGTCAA
AACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTGCGT
GCCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGTGCC
ATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGGCTA
AAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCCAAGAAG
ACC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNGD
DKQSSKNAVQGKRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPFA
KLVVSDVLELDIQIENLERQNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLSA
IWAEKNIEGLDIALQSIGRLKSKDKAELLELLDSYFDPKKT

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGTCTTCTTGGGATCGAAATAGCTGTCC

Construct #8

Name: OPPF12504
Amplified region: 4 - 1206
Vector: pOPINHALO
Description: N-HALO 402C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGATCT
AGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAGTACCGTGACCAATGGTGAC
GATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGATGGCAATGCC
AGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGCAAG
CCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTTGCC
AAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGTCAA
AACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTGCGT
GCCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGTGCC
ATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGGCTA
AAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCCAAGAAG
ACC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNGD
DKQSSKNAVQGKRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPFA
KLVVSDVLELDIQIENLERQNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLSA
IWAEKNIEGLDIALQSIGRLKSKDKAELLELLDSYFDPKKT

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGTCTTCTTGGGATCGAAATAGCTGTCC

Construct #9

Name: OPPF12503
Amplified region: 4 - 1206
Vector: pOPINF
Description: N-HIS 402C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGATCT
AGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAGTACCGTGACCAATGGTGAC
GATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGATGGCAATGCC
AGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGCAAG
CCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTTGCC
AAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGTCAA
AACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTGCGT
GCCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGTGCC
ATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGGCTA
AAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCCAAGAAG
ACC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNGD
DKQSSKNAVQGKRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPFA
KLVVSDVLELDIQIENLERQNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLSA
IWAEKNIEGLDIALQSIGRLKSKDKAELLELLDSYFDPKKT

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGTCTTCTTGGGATCGAAATAGCTGTCC

Construct #10

Name: OPPF12502
Amplified region: 4 - 1206
Vector: pOPINE
Description: C-HIS 402C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACATCTGGATCTGGATCTGGCTTGGGATCTGGATCT
AGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACAGTACCGTGACCAATGGTGAC
GATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGATGGCAATGCC
AGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGCAAG
CCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTTGCC
AAGTTGGTGGTGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGTCAA
AACTCCTTCCGTCCGGTGTTCCTGAACTACTTCAAGACCACTCTGTACCATCGTGTGCGT
GCCGTAAAGTTCCGCATTGTGCTGGCCGAGAATGGGTTCGATTTGAACACGCTAAGTGCC
ATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCCTGCAGTCAATTGGCCGGCTA
AAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGATCCCAAGAAG
ACC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNGD
DKQSSKNAVQGKRELFDGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPFA
KLVVSDVLELDIQIENLERQNSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLSA
IWAEKNIEGLDIALQSIGRLKSKDKAELLELLDSYFDPKKT

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGGTTGCCGATAACATCGATGCCG
Reverse Primer (KH6-pTriExInfrev)	GTGATGGTGATGTTTGGTCTTCTTGGGATCGAAATAGCTGTCC

Construct #11

Name: OPPF12501
Amplified region: 4 - 630
Vector: pOPINE-3C-HALO7
Description: C-HALO 210C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGGTTGCCGATAACATCGATGCCG
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTTGTGGACGCAGAGCAAGCCG

Construct #12

Name: OPPF12500
Amplified region: 4 - 630
Vector: pOPINE-3C-eGFP
Description: C-GFP 210C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGGTTGCCGATAACATCGATGCCG
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTTGTGGACGCAGAGCAAGCCG

Construct #13

Name: OPPF12499
Amplified region: 4 - 630
Vector: pOPINRSJ
Description: N-HIS (Kan^R) 210C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #14

Name: OPPF12498
Amplified region: 4 - 630
Vector: pOPINTF
Description: N-TF 210C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #15

Name: OPPF12497
Amplified region: 4 - 630
Vector: pOPINS3C
Description: N-SUMO 210C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #16

Name: OPPF12496
Amplified region: 4 - 630
Vector: pOPINM
Description: N-MBP 210C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #17

Name: OPPF12495
Amplified region: 4 - 630
Vector: pOPINJ
Description: N-GST 210C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #18

Name: OPPF12494
Amplified region: 4 - 630
Vector: pOPINHALO
Description: N-HALO 210C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #19

Name: OPPF12493
Amplified region: 4 - 630
Vector: pOPINF
Description: N-HIS 210C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #20

Name: OPPF12492
Amplified region: 4 - 630
Vector: pOPINE
Description: C-HIS 210C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAG
GAAGACGCGGCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLSKILSTTKEEDAACSAST

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGGTTGCCGATAACATCGATGCCG
Reverse Primer (KH6-pTriExInfrev)	GTGATGGTGATGTTTTGTGGACGCAGAGCAAGCCG

Construct #21

Name: OPPF12491
Amplified region: 4 - 579
Vector: pOPINE-3C-HALO7
Description: C-HALO 193C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGGTTGCCGATAACATCGATGCCG
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #22

Name: OPPF12490
Amplified region: 4 - 579
Vector: pOPINE-3C-eGFP
Description: C-GFP 193C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGGTTGCCGATAACATCGATGCCG
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #23

Name: OPPF12489
Amplified region: 4 - 579
Vector: pOPINRSJ
Description: N-HIS (Kan^R) 193C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #24

Name: OPPF12488
Amplified region: 4 - 579
Vector: pOPINTF
Description: N-TF 193C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #25

Name: OPPF12487
Amplified region: 4 - 579
Vector: pOPINS3C
Description: N-SUMO 193C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #26

Name: OPPF12486
Amplified region: 4 - 579
Vector: pOPINM
Description: N-MBP 193C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #27

Name: OPPF12485
Amplified region: 4 - 579
Vector: pOPINJ
Description: N-GST 193C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #28

Name: OPPF12484
Amplified region: 4 – 579
Vector: pOPINHALO
Description: N-HALO 193C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #29

Name: OPPF12483
Amplified region: 4 - 579
Vector: pOPINF
Description: N-HIS 193C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGGTTGCCGATAACATCGATGCCG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #30

Name: OPPF12482
Amplified region: 4 – 579
Vector: pOPINE
Description: C-HIS 193C Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCC
GTGGGCGATAAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGAC
ACCACTGGACGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGAC
CACTTCGAGCAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCAT
CAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAG
ATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAA
CTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTC
ACCGCGTCGGTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGAT
GCCAACATAAAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTD
HFEQYIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQ
LKTKAGPSSDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGD
ANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGGTTGCCGATAACATCGATGCCG
Reverse Primer (KH6-pTriExInfrev)	GTGATGGTGATGTTTGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #31
Name: OPPF12481
Amplified region: 4 - 1596
Vector: pOPINF
Description: N-HIS Full Length Exu

DNA Sequence:

GTTGCCGATAACATCGATGCCGGAGTGGCCATTGCCGTCGCCGATCAATCCTCGTCGCCCGTGGGCGATA
AAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGACGTCGTCTAAT
GGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAGCAGTACATCATGCCATATATG
AATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGTGTTATTTCGATTGGCTTTTATCGTATGCTGA
AGTCGATGCAGACTTATAAGATCATTAAATCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTG
GCTTGAGCAACTGAAGACCAAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGC
AAGTTCATTCCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCGG
TGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATAAAGAACTACAG
TCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAGGAAGACGCGGCTTGCTCTGCGTCCACATCT
GGATCTGGATCTGGCTTGGGATCTGGATCTAGTATGGTGTCCGATTCGGTATCGATATCCCCCAGGGACA
GTACCGTGACCAATGGTGACGATAAGCAATCTAGCAAGAACGCCGTTCAGGGTAAGCGGGAACTGTTTGA
TGGCAATGCCAGTGTTCGTGCCAAATTGGCCTTTGATGTGGCCCTTCAGCTGAGCAACTCCGACGGCAAG
CCGGAACCAAAAAGCTCGGAGGCATTGGAGAATATGTTTAATGCCATACGTCCGTTTGCCAAGTTGGTGG
TGTCCGATGTCCTGGAGCTGGATATCCAGATCGAAAATTTGGAGCGTCAAAACTCCTTCCGTCCGGTGTT
CCTGAACTACTTCAAGACCACTCTGTACCATCGTGTGCGTGCCGTAAAGTTCCGCATTGTGCTGGCCGAG
AATGGGTTCGATTTGAACACGCTAAGTGCCATCTGGGCAGAGAAAAACATCGAGGGCCTGGACATAGCCC
TGCAGTCAATTGGCCGGCTAAAGAGCAAGGACAAGGCGGAGCTGCTCGAGCTTCTGGACAGCTATTTCGA
TCCCAAGAAGACCACGGTGAAGCCGGTTGTCAAGGGTAACAGCAACAATAATAACAATTATCGTCGTCGC
AATCGACGCGGCGGCCGCCAATCGGTGAAGGACGCTAGGCCATCCAGCTCCCCATCGGCTAGCACCGAGT
TTGGAGCAGGAGGTGACAAGTCGCGCAGCGTTTCCTCGTTGCCCGACTCTACTACCAAGACTCCCTCGCC
AAACAAGCCGCGTATGCACCGTAAGCGCAACTCGCGTCAAAGCTTGGGTGCAACCCCCAATGGACTCAAG
GTTGCTGCTGAAATATCCTCCTCAGGCGTGTCTGAGCTCAACAATAGCGCTCCACCTGCCGTAACCATAA
GCCCAGTCGTGGCACAACCCTCACCGACACCGGTGGCCATCACGGCCTCCAAC

Protein Sequence (untagged):

VADNIDAGVAIAVADQSSSPVGDKVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQ
YIMPYMNLNPAARQRHQVRVISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPS
SDGIVLIYHEERKFIPYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLS
KILSTTKEEDAACSASTSGSGSGLGSGSSMVSDSVSISPRDSTVTNGDDKQSSKNAVQGKRELF
DGNASVRAKLAFDVALQLSNSDGKPEPKSSEALENMFNAIRPFAKLVVSDVLELDIQIENLERQ
NSFRPVFLNYFKTTLYHRVRAVKFRIVLAENGFDLNTLSAIWAEKNIEGLDIALQSIGRLKSKD
KAELLELLDSYFDPKKTTVKPVVKGNSNNNNNYRRRNRRGGRQSVKDARPSSSPSASTEFGAGG
DKSRSVSSLPDSTTKTPSPNKPRMHRKRNSRQSLGATPNGLKVAAEISSSGVSELNNSAPPAVT
ISPVVAQPSPTPVAITASN

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGGTTGCCGATAACATCGATGCCG
Reverse Primer (KH6-pTriExInfrev)	GTGATGGTGATGTTTGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #32

Name: OPPF12538
Amplified region: 73 - 630
Vector: pOPINE-3C-HALO7
Description: C-HALO 25-210 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAGGAAGACGCG
GCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLSKILSTTKEEDA
ACSAST

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTTGTGGACGCAGAGCAAGCCG

Construct #33

Name: OPPF12537
Amplified region: 73 - 630
Vector: pOPINE-3C-eGFP
Description: C-GFP 25-210 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAGGAAGACGCG
GCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLSKILSTTKEEDA
ACSAST

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTTGTGGACGCAGAGCAAGCCG

Construct #34

Name: OPPF12536
Amplified region: 73 - 630
Vector: pOPINRSJ
Description: N-GST (Kan^R) 25-210 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAGGAAGACGCG
GCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLSKILSTTKEEDA
ACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #35

Name: OPPF12535
Amplified region: 73 - 630
Vector: pOPINTF
Description: N-TF 25-210 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAGGAAGACGCG
GCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLSKILSTTKEEDA
ACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #36

Name: OPPF12534
Amplified region: 73 - 630
Vector: pOPINS3C
Description: N-SUMO 25-210 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAGGAAGACGCG
GCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLSKILSTTKEEDA
ACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #37

Name: OPPF12533
Amplified region: 73 - 630
Vector: pOPINM
Description: N-MBP 25-210 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAGGAAGACGCG
GCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLSKILSTTKEEDA
ACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #38

Name: OPPF12532
Amplified region: 73 - 630
Vector: pOPINJ
Description: N-GST 25-210 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAGGAAGACGCG
GCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLSKILSTTKEEDA
ACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #39

Name: OPPF12531
Amplified region: 73 - 630
Vector: pOPINHALO
Description: N-HALO 25-210 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAGGAAGACGCG
GCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLSKILSTTKEEDA
ACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #40

Name: OPPF12530
Amplified region: 73 - 630
Vector: pOPINF
Description: N-HIS 25-210 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAGGAAGACGCG
GCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLSKILSTTKEEDA
ACSAST

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGTGGACGCAGAGCAAGCCG

Construct #41

Name: OPPF12529
Amplified region: 73 - 630
Vector: pOPINE
Description: C-HIS 25-210 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCCAAGATCTTGTCCACGACCAAGGAGGAAGACGCG
GCTTGCTCTGCGTCCACA

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLSKILSTTKEEDA
ACSAST

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (KH6-pTriExInfrev)	GTGATGGTGATGTTTTGTGGACGCAGAGCAAGCCG

Construct #42

Name: OPPF12528
Amplified region: 73 - 579
Vector: pOPINE-3C-HALO7
Description: C-HALO 25-193 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #43

Name: OPPF12527
Amplified region: 73 - 579
Vector: pOPINE-3C-eGFP
Description: C-GFP 25-193 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #44

Name: OPPF12526
Amplified region: 73 - 579
Vector: pOPINRSJ
Description: N-GST (Kan^R) 25-193 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #45

Name: OPPF12525
Amplified region: 73 - 579
Vector: pOPINTF
Description: N-TF 25-193 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #46

Name: OPPF12524
Amplified region: 73 - 579
Vector: pOPINS3C
Description: N-SUMO 25-193 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #47

Name: OPPF12523
Amplified region: 73 - 579
Vector: pOPINM
Description: N-SUMO 25-193 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #48

Name: OPPF12522
Amplified region: 73 - 579
Vector: pOPINJ
Description: N-GST 25-193 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #49

Name: OPPF12521
Amplified region: 73 - 579
Vector: pOPINHALO
Description: N-HALO 25-193 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #50

Name: OPPF12520
Amplified region: 73 - 579
Vector: pOPINF
Description: N-HIS 25-193 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Construct #51

Name: OPPF12519
Amplified region: 73 - 579
Vector: pOPINE
Description: C-HIS 25-193 Exu

DNA Sequence:

AAAGTTGAGCTGCCGGCCGGTAATTATATCCTGGTGGGCGTGGATATCGACACCACTGGA
CGTCGTCTAATGGATGAGATTGTCCAGCTGGCCGCTTATACACCCACCGACCACTTCGAG
CAGTACATCATGCCATATATGAATCTGAATCCAGCTGCTCGCCAGCGTCATCAAGTTCGT
GTTATTTCGATTGGCTTTTATCGTATGCTGAAGTCGATGCAGACTTATAAGATCATTAAA
TCCAAGTCTGAGATCGCTGCCCTCAAGGACTTCCTCAACTGGCTTGAGCAACTGAAGACC
AAGGCGGGTCCCAGCTCGGACGGGATTGTGCTTATCTATCATGAAGAGCGCAAGTTCATT
CCCTACATGATCCTGGAGTCGTTGAAGAAGTACGGCCTGTTGGAGCGCTTCACCGCGTCG
GTGAAGTCGTTTGCAAACAGCATCAATCTGGCCAAGGCTTCCATTGGCGATGCCAACATA
AAGAACTACAGTCTGCGAAAGCTTTCC

Protein Sequence (untagged):

KVELPAGNYILVGVDIDTTGRRLMDEIVQLAAYTPTDHFEQYIMPYMNLNPAARQRHQVR
VISIGFYRMLKSMQTYKIIKSKSEIAALKDFLNWLEQLKTKAGPSSDGIVLIYHEERKFI
PYMILESLKKYGLLERFTASVKSFANSINLAKASIGDANIKNYSLRKLS

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGAAAGTTGAGCTGCCGGCCGG
Reverse Primer (KH6-pTriExInfrev)	GTGATGGTGATGTTTGGAAAGCTTTCGCAGACTGTAGTTCTTTATG

Translation of ypsilon-schachtel (*yps*)

Construct #1

Name: OPPF12515
Amplified region: 163 - 513
Vector: pOPINRSJ
Description: N-HIS (Kan^R) 171C Yps

DNA Sequence:

AAGGAAGTCATCGCCACCAAAGTCACCGGCACCGTCAAGTGGTTCAACGTGAAGAGCGGC
TACGGCTTCATCAACCGCAACGACACCAGAGAGGATGTCTTTGTGCACCAGAGCGCCATT
GCGCGGAACAACCCCAAAAAGGCGGTCCGCTCGGTGGGCGACGGTGAGGTCGTTGAGTTC
GACGTGGTCATTGGTGAGAAGGGCAACGAGGCGGCCAACGTGACCGGTCCCTCCGGTGAG
CCGGTGCGGGGCAGTCAGTTTGCAGCGGACAAGCGCCGTAACTTCCGTCCCTGGATGAAG
AAGAATCGCCGCAAGGATGGCGAAGTGGAAGGCGAAGACGCCGAATCGTCG

Protein Sequence (untagged):

KEVIATKVTGTVKWFNVKSGYGFINRNDTREDVFVHQSAIARNNPKKAVRSVGDGEVVEF
DVVIGEKGNEAANVTGPSGEPVRGSQFAADKRRNFRPWMKKNRRKDGEVEGEDAESS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAGGAAGTCATCGCCACCAAAGTCAC
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTACGACGATTCGGCGTCTTCGC

Construct #2

Name: OPPF12514
Amplified region: 163 - 513
Vector: pOPINS3C
Description: N-SUMO 171C Yps

DNA Sequence:

AAGGAAGTCATCGCCACCAAAGTCACCGGCACCGTCAAGTGGTTCAACGTGAAGAGCGGC
TACGGCTTCATCAACCGCAACGACACCAGAGAGGATGTCTTTGTGCACCAGAGCGCCATT
GCGCGGAACAACCCCAAAAAGGCGGTCCGCTCGGTGGGCGACGGTGAGGTCGTTGAGTTC
GACGTGGTCATTGGTGAGAAGGGCAACGAGGCGGCCAACGTGACCGGTCCCTCCGGTGAG
CCGGTGCGGGGCAGTCAGTTTGCAGCGGACAAGCGCCGTAACTTCCGTCCCTGGATGAAG
AAGAATCGCCGCAAGGATGGCGAAGTGGAAGGCGAAGACGCCGAATCGTCG

Protein Sequence (untagged):

KEVIATKVTGTVKWFNVKSGYGFINRNDTREDVFVHQSAIARNNPKKAVRSVGDGEVVEF
DVVIGEKGNEAANVTGPSGEPVRGSQFAADKRRNFRPWMKKNRRKDGEVEGEDAESS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAGGAAGTCATCGCCACCAAAGTCAC
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTACGACGATTCGGCGTCTTCGC

Construct #3

Name: OPPF12513
Amplified region: 163 - 513
Vector: pOPINF
Description: N-HIS 171C Yps

DNA Sequence:

AAGGAAGTCATCGCCACCAAAGTCACCGGCACCGTCAAGTGGTTCAACGTGAAGAGCGGC
TACGGCTTCATCAACCGCAACGACACCAGAGAGGATGTCTTTGTGCACCAGAGCGCCATT
GCGCGGAACAACCCCAAAAAGGCGGTCCGCTCGGTGGGCGACGGTGAGGTCGTTGAGTTC
GACGTGGTCATTGGTGAGAAGGGCAACGAGGCGGCCAACGTGACCGGTCCCTCCGGTGAG
CCGGTGCGGGGCAGTCAGTTTGCAGCGGACAAGCGCCGTAACTTCCGTCCCTGGATGAAG
AAGAATCGCCGCAAGGATGGCGAAGTGGAAGGCGAAGACGCCGAATCGTCG

Protein Sequence (untagged):

KEVIATKVTGTVKWFNVKSGYGFINRNDTREDVFVHQSAIARNNPKKAVRSVGDGEVVEF
DVVIGEKGNEAANVTGPSGEPVRGSQFAADKRRNFRPWMKKNRRKDGEVEGEDAESS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAAGGAAGTCATCGCCACCAAAGTCAC
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTACGACGATTCGGCGTCTTCGC

Construct #4

Name: OPPF12512
Amplified region: 163 - 513
Vector: pOPINE
Description: C-HIS 171C Yps

DNA Sequence:

AAGGAAGTCATCGCCACCAAAGTCACCGGCACCGTCAAGTGGTTCAACGTGAAGAGCGGC
TACGGCTTCATCAACCGCAACGACACCAGAGAGGATGTCTTTGTGCACCAGAGCGCCATT
GCGCGGAACAACCCCAAAAAGGCGGTCCGCTCGGTGGGCGACGGTGAGGTCGTTGAGTTC
GACGTGGTCATTGGTGAGAAGGGCAACGAGGCGGCCAACGTGACCGGTCCCTCCGGTGAG
CCGGTGCGGGGCAGTCAGTTTGCAGCGGACAAGCGCCGTAACTTCCGTCCCTGGATGAAG
AAGAATCGCCGCAAGGATGGCGAAGTGGAAGGCGAAGACGCCGAATCGTCG

Protein Sequence (untagged):

KEVIATKVTGTVKWFNVKSGYGFINRNDTREDVFVHQSAIARNNPKKAVRSVGDGEVVEF
DVVIGEKGNEAANVTGPSGEPVRGSQFAADKRRNFRPWMKKNRRKDGEVEGEDAESS

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGAAGGAAGTCATCGCCACCAAAGTCAC
Reverse Primer (KH6-pTriExInfrev)	GTGATGGTGATGTTTCGACGATTCGGCGTCTTCGC

D. OPPF construct design – GatD and MurT

Translation of gatD

Construct #1

Name: OPPF12144
Amplified region: 4 - 729
Vector: pOPINB
Description: N-His tag Full Length (low copy)

DNA Sequence:

CATGAATTGACTATTTATCATTTTATGTCAGATAAATTGAATTTATACAGTGATATAGGA
AATATTATTGCTTTAAGACAACGTGCTAAAAAACGAAATATTAAAGTTAATGTCGTAGAA
ATCAATGAAACAGAAGGTATTACCTTTGATGAATGTGATATTTTCTTTATCGGTGGTGGA
AGTGATAGAGAACAAGCATTAGCAACAAAAGAATTAAGTAAAATTAAGACACCACTTAAA
GAAGCGATTGAAGATGGTATGCCGGGATTAACGATTTGTGGAGGCTATCAATTTTTAGGG
AAAAAATATATCACGCCTGATGGTACAGAATTAGAAGGGTTAGGTATTTTAGATTTTTAT
ACTGAATCAAAGACAAACCGATTAACAGGAGATATTGTTATCGAAAGTGATACTTTTGGA
ACTATTGTAGGTTTTGAAAATCACGGTGGTAGAACATATCATGATTTCGGTACACTTGGT
CATGTTACTTTTGGTTATGGTAATAATGATGAAGATAAAAAAGAAGGCATTCATTATAAA
AATTTATTAGGTACTTATTTACATGGACCAATTTTACCTAAAAATTACGAAATCACTGAT
TATCTGTTAGAAAAAGCTTGTGAACGTAAGGGTATTCCGTTTGAGCCTAAAGAAATAGAT
AATGAAGCGGAAATACAAGCGAAACAAGTATTAATAGACAGAGCAAATAGACAGAAGAAA
TCTCGT

Protein Sequence (untagged):

HELTIYHFMSDKLNLYSDIGNIIALRQRAKKRNIKVNVVEINETEGITFDECDIFFIGGG
SDREQALATKELSKIKTPLKEAIEDGMPGLTICGGYQFLGKKYITPDGTELEGLGILDFY
TESKTNRLTGDIVIESDTFGTIVGFENHGGRTYHDFGTLGHVTFGYGNNDEDKKEGIHYK
NLLGTYLHGPILPKNYEITDYLLEKACERKGIPFEPKEIDNEAEIQAKQVLIDRANRQKK
SR

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGCATGAATTGACTATTTATCATTTTATGTCAG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAACGAGATTTCTTCTGTCTATTTGCTC

Construct #2

Name: OPPF12143
Amplified region: 4 - 729
Vector: pOPINF
Description: N-HIS-GatD

DNA Sequence:

CATGAATTGACTATTTATCATTTTATGTCAGATAAATTGAATTTATACAGTGATATAGGA
AATATTATTGCTTTAAGACAACGTGCTAAAAAACGAAATATTAAAGTTAATGTCGTAGAA
ATCAATGAAACAGAAGGTATTACCTTTGATGAATGTGATATTTTCTTTATCGGTGGTGGA
AGTGATAGAGAACAAGCATTAGCAACAAAAGAATTAAGTAAAATTAAGACACCACTTAAA
GAAGCGATTGAAGATGGTATGCCGGGATTAACGATTTGTGGAGGCTATCAATTTTTAGGG
AAAAAATATATCACGCCTGATGGTACAGAATTAGAAGGGTTAGGTATTTTAGATTTTTAT
ACTGAATCAAAGACAAACCGATTAACAGGAGATATTGTTATCGAAAGTGATACTTTTGGA
ACTATTGTAGGTTTTGAAAATCACGGTGGTAGAACATATCATGATTTCGGTACACTTGGT
CATGTTACTTTTGGTTATGGTAATAATGATGAAGATAAAAAAGAAGGCATTCATTATAAA
AATTTATTAGGTACTTATTTACATGGACCAATTTTACCTAAAAATTACGAAATCACTGAT
TATCTGTTAGAAAAAGCTTGTGAACGTAAGGGTATTCCGTTTGAGCCTAAAGAAATAGAT
AATGAAGCGGAAATACAAGCGAAACAAGTATTAATAGACAGAGCAAATAGACAGAAGAAA
TCTCGT

Protein Sequence (untagged):

HELTIYHFMSDKLNLYSDIGNIIALRQRAKKRNIKVNVVEINETEGITFDECDIFFIGGG
SDREQALATKELSKIKTPLKEAIEDGMPGLTICGGYQFLGKKYITPDGTELEGLGILDFY
TESKTNRLTGDIVIESDTFGTIVGFENHGGRTYHDFGTLGHVTFGYGNNDEDKKEGIHYK
NLLGTYLHGPILPKNYEITDYLLEKACERKGIPFEPKEIDNEAEIQAKQVLIDRANRQKK
SR

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGCATGAATTGACTATTTATCATTTTATGTCAG
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTAACGAGATTTCTTCTGTCTATTTGCTC

Translation of murT

Construct #1

Name: OPPF12142
Amplified region: 4 - 1311
Vector: pOPINE-3C-HALO7
Description: C-HALO7 tag full length murT

DNA Sequence:

AGACAGTGGACGGCAATCCATCTAGCGAAATTGGCGCGTAAAGCAAGTAGAGCAGTAGGT
AAAAGAGGAACAGATTTACCTGGACAAATCGCTAGAAAAGTGGATACAGATATATTAAGA
AAATTAGCAGAGCAAGTTGATGATATTGTATTTATCAGTGGAACAAATGGTAAAACAACG
ACTTCAAACTTAATTGGACATACTTTAAAAGCAAATAATATTCAAATTATACACAATAAT
GAAGGTGCTAATATGGCTGCAGGTATAACTTCTGCATTCATCATGCAATCAACACCTAAG
ACTAAAATTGCGGTAATCGAAATTGATGAAGGTTCGATTCCACGTGTGTTAAAAGAAGTT
ACACCTTCAATGATGGTATTTACTAATTTCTTTAGAGATCAAATGGATCGCTTCGGTGAA
ATTGATATTATGGTTAATAACATTGCAGAGACAATTAGTAATAAAGGCATCAAATTATTG
CTAAATGCTGATGATCCATTTGTGAGTCGTTTGAAAATCGCAAGTGATACGATTGTGTAC
TATGGTATGAAAGCACATGCCCATGAATTTGAACAAAGTACGATGAATGAAAGTAGATAT
TGTCCAAACTGTGGTCGCTTATTGCAATACGATTATATTCATTATAATCAAATTGGTCAT
TATCACTGTCAGTGTGGTTTCAAACGAGAGCAAGCAAAATATGAAATATCAAGTTTTGAT
GTGGCACCGTTTTTATATTTAAATATCAATGATGAAAAATATGATATGAAAATTGCAGGT
GACTTTAACGCTTATAACGCGTTAGCAGCATATACTGTTTTAAGAGAGCTAGGGTTAAAT
GAACAAACAATTAAAAATGGCTTTGAAACGTATACATCAGACAATGGTCGTATGCAGTAC
TTTAAAAAAGAACGAAAAGAAGCGATGATCAATTTAGCTAAAAATCCTGCAGGAATGAAT
GCAAGTTTATCAGTTGGTGAACAATTAGAAGGCGAAAAAGTGTATGTTATTTCGCTAAAT
GATAACGCTGCAGATGGTCGAGATACTTCATGGATTTATGATGCAGATTTTGAAAAATTA
TCTAAGCAACAAATTGAAGCTATCATCGTGACAGGTACACGAGCAGAAGAACTTCAATTG
CGATTGAAGTTAGCAGAGGTTGAAGTACCAATTATAGTTGAGCGTGATATTTATAAAGCA
ACGGCAAAGACTATGGATTATAAAGGTTTCACAGTTGCAATACCAAACTATACATCATTA
GCGCCTATGCTTGAACAATTAAACCGTTCGTTTGAAGGAGGTCAATCA

Protein Sequence (untagged):

RQWTAIHLAKLARKASRAVGKRGTDLPGQIARKVDTDILRKLAEQVDDIVFISGTNGKTT
TSNLIGHTLKANNIQIIHNNEGANMAAGITSAFIMQSTPKTKIAVIEIDEGSIPRVLKEV
TPSMMVFTNFFRDQMDRFGEIDIMVNNIAETISNKGIKLLLNADDPFVSRLKIASDTIVY
YGMKAHAHEFEQSTMNESRYCPNCGRLLQYDYIHYNQIGHYHCQCGFKREQAKYEISSFD
VAPFLYLNINDEKYDMKIAGDFNAYNALAAYTVLRELGLNEQTIKNGFETYTSDNGRMQY
FKKERKEAMINLAKNPAGMNASLSVGEQLEGEKVYVISLNDNAADGRDTSWIYDADFEKL
SKQQIEAIIVTGTRAEELQLRLKLAEVEVPIIVERDIYKATAKTMDYKGFTVAIPNYTSL
APMLEQLNRSFEGGQS

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGAGACAGTGGACGGCAATCCATC
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTTGATTGACCTCCTTCAAACGAACGG

Construct #2

Name: OPPF12141
Amplified region: 4 - 1311
Vector: pOPINE-3C-HALO7
Description: C-HALO7 tag full length murT

DNA Sequence:

AGACAGTGGACGGCAATCCATCTAGCGAAATTGGCGCGTAAAGCAAGTAGAGCAGTAGGT
AAAAGAGGAACAGATTTACCTGGACAAATCGCTAGAAAAGTGGATACAGATATATTAAGA
AAATTAGCAGAGCAAGTTGATGATATTGTATTTATCAGTGGAACAAATGGTAAAACAACG
ACTTCAAACTTAATTGGACATACTTTAAAAGCAAATAATATTCAAATTATACACAATAAT
GAAGGTGCTAATATGGCTGCAGGTATAACTTCTGCATTCATCATGCAATCAACACCTAAG
ACTAAAATTGCGGTAATCGAAATTGATGAAGGTTCGATTCCACGTGTGTTAAAAGAAGTT
ACACCTTCAATGATGGTATTTACTAATTTCTTTAGAGATCAAATGGATCGCTTCGGTGAA
ATTGATATTATGGTTAATAACATTGCAGAGACAATTAGTAATAAAGGCATCAAATTATTG
CTAAATGCTGATGATCCATTTGTGAGTCGTTTGAAAATCGCAAGTGATACGATTGTGTAC
TATGGTATGAAAGCACATGCCCATGAATTTGAACAAAGTACGATGAATGAAAGTAGATAT
TGTCCAAACTGTGGTCGCTTATTGCAATACGATTATATTCATTATAATCAAATTGGTCAT
TATCACTGTCAGTGTGGTTTCAAACGAGAGCAAGCAAAATATGAAATATCAAGTTTTGAT
GTGGCACCGTTTTTATATTTAAATATCAATGATGAAAAATATGATATGAAAATTGCAGGT
GACTTTAACGCTTATAACGCGTTAGCAGCATATACTGTTTTAAGAGAGCTAGGGTTAAAT
GAACAAACAATTAAAAATGGCTTTGAAACGTATACATCAGACAATGGTCGTATGCAGTAC
TTTAAAAAAGAACGAAAAGAAGCGATGATCAATTTAGCTAAAAATCCTGCAGGAATGAAT
GCAAGTTTATCAGTTGGTGAACAATTAGAAGGCGAAAAAGTGTATGTTATTTCGCTAAAT
GATAACGCTGCAGATGGTCGAGATACTTCATGGATTTATGATGCAGATTTTGAAAAATTA
TCTAAGCAACAAATTGAAGCTATCATCGTGACAGGTACACGAGCAGAAGAACTTCAATTG
CGATTGAAGTTAGCAGAGGTTGAAGTACCAATTATAGTTGAGCGTGATATTTATAAAGCA
ACGGCAAAGACTATGGATTATAAAGGTTTCACAGTTGCAATACCAAACTATACATCATTA
GCGCCTATGCTTGAACAATTAAACCGTTCGTTTGAAGGAGGTCAATCA

Protein Sequence (untagged):

RQWTAIHLAKLARKASRAVGKRGTDLPGQIARKVDTDILRKLAEQVDDIVFISGTNGKTT
TSNLIGHTLKANNIQIIHNNEGANMAAGITSAFIMQSTPKTKIAVIEIDEGSIPRVLKEV
TPSMMVFTNFFRDQMDRFGEIDIMVNNIAETISNKGIKLLLNADDPFVSRLKIASDTIVY
YGMKAHAHEFEQSTMNESRYCPNCGRLLQYDYIHYNQIGHYHCQCGFKREQAKYEISSFD
VAPFLYLNINDEKYDMKIAGDFNAYNALAAYTVLRELGLNEQTIKNGFETYTSDNGRMQY
FKKERKEAMINLAKNPAGMNASLSVGEQLEGEKVYVISLNDNAADGRDTSWIYDADFEKL
SKQQIEAIIVTGTRAEELQLRLKLAEVEVPIIVERDIYKATAKTMDYKGFTVAIPNYTSL
APMLEQLNRSFEGGQS

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGAGACAGTGGACGGCAATCCATC
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTTGATTGACCTCCTTCAAACGAACGG

Construct #3

Name: OPPF12140
Amplified region: 4 - 1311
Vector: pOPINJ
Description: N-GST tag full length murT

DNA Sequence:

AGACAGTGGACGGCAATCCATCTAGCGAAATTGGCGCGTAAAGCAAGTAGAGCAGTAGGT
AAAAGAGGAACAGATTTACCTGGACAAATCGCTAGAAAAGTGGATACAGATATATTAAGA
AAATTAGCAGAGCAAGTTGATGATATTGTATTTATCAGTGGAACAAATGGTAAAACAACG
ACTTCAAACTTAATTGGACATACTTTAAAAGCAAATAATATTCAAATTATACACAATAAT
GAAGGTGCTAATATGGCTGCAGGTATAACTTCTGCATTCATCATGCAATCAACACCTAAG
ACTAAAATTGCGGTAATCGAAATTGATGAAGGTTCGATTCCACGTGTGTTAAAAGAAGTT
ACACCTTCAATGATGGTATTTACTAATTTCTTTAGAGATCAAATGGATCGCTTCGGTGAA
ATTGATATTATGGTTAATAACATTGCAGAGACAATTAGTAATAAAGGCATCAAATTATTG
CTAAATGCTGATGATCCATTTGTGAGTCGTTTGAAAATCGCAAGTGATACGATTGTGTAC
TATGGTATGAAAGCACATGCCCATGAATTTGAACAAAGTACGATGAATGAAAGTAGATAT
TGTCCAAACTGTGGTCGCTTATTGCAATACGATTATATTCATTATAATCAAATTGGTCAT
TATCACTGTCAGTGTGGTTTCAAACGAGAGCAAGCAAAATATGAAATATCAAGTTTTGAT
GTGGCACCGTTTTTATATTTAAATATCAATGATGAAAAATATGATATGAAAATTGCAGGT
GACTTTAACGCTTATAACGCGTTAGCAGCATATACTGTTTTAAGAGAGCTAGGGTTAAAT
GAACAAACAATTAAAAATGGCTTTGAAACGTATACATCAGACAATGGTCGTATGCAGTAC
TTTAAAAAAGAACGAAAAGAAGCGATGATCAATTTAGCTAAAAATCCTGCAGGAATGAAT
GCAAGTTTATCAGTTGGTGAACAATTAGAAGGCGAAAAAGTGTATGTTATTTCGCTAAAT
GATAACGCTGCAGATGGTCGAGATACTTCATGGATTTATGATGCAGATTTTGAAAAATTA
TCTAAGCAACAAATTGAAGCTATCATCGTGACAGGTACACGAGCAGAAGAACTTCAATTG
CGATTGAAGTTAGCAGAGGTTGAAGTACCAATTATAGTTGAGCGTGATATTTATAAAGCA
ACGGCAAAGACTATGGATTATAAAGGTTTCACAGTTGCAATACCAAACTATACATCATTA
GCGCCTATGCTTGAACAATTAAACCGTTCGTTTGAAGGAGGTCAATCA

Protein Sequence (untagged):

RQWTAIHLAKLARKASRAVGKRGTDLPGQIARKVDTDILRKLAEQVDDIVFISGTNGKTT
TSNLIGHTLKANNIQIIHNNEGANMAAGITSAFIMQSTPKTKIAVIEIDEGSIPRVLKEV
TPSMMVFTNFFRDQMDRFGEIDIMVNNIAETISNKGIKLLLNADDPFVSRLKIASDTIVY
YGMKAHAHEFEQSTMNESRYCPNCGRLLQYDYIHYNQIGHYHCQCGFKREQAKYEISSFD
VAPFLYLNINDEKYDMKIAGDFNAYNALAAYTVLRELGLNEQTIKNGFETYTSDNGRMQY
FKKERKEAMINLAKNPAGMNASLSVGEQLEGEKVYVISLNDNAADGRDTSWIYDADFEKL
SKQQIEAIIVTGTRAEELQLRLKLAEVEVPIIVERDIYKATAKTMDYKGFTVAIPNYTSL
APMLEQLNRSFEGGQS

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGAGACAGTGGACGGCAATCCATC
Reverse Primer (3CFusionRev)	CAGAACTTCCAGTTTTGATTGACCTCCTTCAAACGAACGG

Construct #4

Name: OPPF12139
Amplified region: 4 - 1311
Vector: pOPINJ
Description: N-GST tag full length murT

DNA Sequence:

AGACAGTGGACGGCAATCCATCTAGCGAAATTGGCGCGTAAAGCAAGTAGAGCAGTAGGT
AAAAGAGGAACAGATTTACCTGGACAAATCGCTAGAAAAGTGGATACAGATATATTAAGA
AAATTAGCAGAGCAAGTTGATGATATTGTATTTATCAGTGGAACAAATGGTAAAACAACG
ACTTCAAACTTAATTGGACATACTTTAAAAGCAAATAATATTCAAATTATACACAATAAT
GAAGGTGCTAATATGGCTGCAGGTATAACTTCTGCATTCATCATGCAATCAACACCTAAG
ACTAAAATTGCGGTAATCGAAATTGATGAAGGTTCGATTCCACGTGTGTTAAAAGAAGTT
ACACCTTCAATGATGGTATTTACTAATTTCTTTAGAGATCAAATGGATCGCTTCGGTGAA
ATTGATATTATGGTTAATAACATTGCAGAGACAATTAGTAATAAAGGCATCAAATTATTG
CTAAATGCTGATGATCCATTTGTGAGTCGTTTGAAAATCGCAAGTGATACGATTGTGTAC
TATGGTATGAAAGCACATGCCCATGAATTTGAACAAAGTACGATGAATGAAAGTAGATAT
TGTCCAAACTGTGGTCGCTTATTGCAATACGATTATATTCATTATAATCAAATTGGTCAT
TATCACTGTCAGTGTGGTTTCAAACGAGAGCAAGCAAAATATGAAATATCAAGTTTTGAT
GTGGCACCGTTTTTATATTTAAATATCAATGATGAAAAATATGATATGAAAATTGCAGGT
GACTTTAACGCTTATAACGCGTTAGCAGCATATACTGTTTTAAGAGAGCTAGGGTTAAAT
GAACAAACAATTAAAAATGGCTTTGAAACGTATACATCAGACAATGGTCGTATGCAGTAC
TTTAAAAAAGAACGAAAAGAAGCGATGATCAATTTAGCTAAAAATCCTGCAGGAATGAAT
GCAAGTTTATCAGTTGGTGAACAATTAGAAGGCGAAAAAGTGTATGTTATTTCGCTAAAT
GATAACGCTGCAGATGGTCGAGATACTTCATGGATTTATGATGCAGATTTTGAAAAATTA
TCTAAGCAACAAATTGAAGCTATCATCGTGACAGGTACACGAGCAGAAGAACTTCAATTG
CGATTGAAGTTAGCAGAGGTTGAAGTACCAATTATAGTTGAGCGTGATATTTATAAAGCA
ACGGCAAAGACTATGGATTATAAAGGTTTCACAGTTGCAATACCAAACTATACATCATTA
GCGCCTATGCTTGAACAATTAAACCGTTCGTTTGAAGGAGGTCAATCA

Protein Sequence (untagged):

RQWTAIHLAKLARKASRAVGKRGTDLPGQIARKVDTDILRKLAEQVDDIVFISGTNGKTT
TSNLIGHTLKANNIQIIHNNEGANMAAGITSAFIMQSTPKTKIAVIEIDEGSIPRVLKEV
TPSMMVFTNFFRDQMDRFGEIDIMVNNIAETISNKGIKLLLNADDPFVSRLKIASDTIVY
YGMKAHAHEFEQSTMNESRYCPNCGRLLQYDYIHYNQIGHYHCQCGFKREQAKYEISSFD
VAPFLYLNINDEKYDMKIAGDFNAYNALAAYTVLRELGLNEQTIKNGFETYTSDNGRMQY
FKKERKEAMINLAKNPAGMNASLSVGEQLEGEKVYVISLNDNAADGRDTSWIYDADFEKL
SKQQIEAIIVTGTRAEELQLRLKLAEVEVPIIVERDIYKATAKTMDYKGFTVAIPNYTSL
APMLEQLNRSFEGGQS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAGACAGTGGACGGCAATCCATC
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGATTGACCTCCTTCAAACGAACGG

Construct #5

Name: OPPF12138
Amplified region: 4 - 1311
Vector: pOPINS3C
Description: N-SUMO tag full length murT

DNA Sequence:

AGACAGTGGACGGCAATCCATCTAGCGAAATTGGCGCGTAAAGCAAGTAGAGCAGTAGGT
AAAAGAGGAACAGATTTACCTGGACAAATCGCTAGAAAAGTGGATACAGATATATTAAGA
AAATTAGCAGAGCAAGTTGATGATATTGTATTTATCAGTGGAACAAATGGTAAAACAACG
ACTTCAAACTTAATTGGACATACTTTAAAAGCAAATAATATTCAAATTATACACAATAAT
GAAGGTGCTAATATGGCTGCAGGTATAACTTCTGCATTCATCATGCAATCAACACCTAAG
ACTAAAATTGCGGTAATCGAAATTGATGAAGGTTCGATTCCACGTGTGTTAAAAGAAGTT
ACACCTTCAATGATGGTATTTACTAATTTCTTTAGAGATCAAATGGATCGCTTCGGTGAA
ATTGATATTATGGTTAATAACATTGCAGAGACAATTAGTAATAAAGGCATCAAATTATTG
CTAAATGCTGATGATCCATTTGTGAGTCGTTTGAAAATCGCAAGTGATACGATTGTGTAC
TATGGTATGAAAGCACATGCCCATGAATTTGAACAAAGTACGATGAATGAAAGTAGATAT
TGTCCAAACTGTGGTCGCTTATTGCAATACGATTATATTCATTATAATCAAATTGGTCAT
TATCACTGTCAGTGTGGTTTCAAACGAGAGCAAGCAAAATATGAAATATCAAGTTTTGAT
GTGGCACCGTTTTTATATTTAAATATCAATGATGAAAAATATGATATGAAAATTGCAGGT
GACTTTAACGCTTATAACGCGTTAGCAGCATATACTGTTTTAAGAGAGCTAGGGTTAAAT
GAACAAACAATTAAAAATGGCTTTGAAACGTATACATCAGACAATGGTCGTATGCAGTAC
TTTAAAAAAGAACGAAAAGAAGCGATGATCAATTTAGCTAAAAATCCTGCAGGAATGAAT
GCAAGTTTATCAGTTGGTGAACAATTAGAAGGCGAAAAAGTGTATGTTATTTCGCTAAAT
GATAACGCTGCAGATGGTCGAGATACTTCATGGATTTATGATGCAGATTTTGAAAAATTA
TCTAAGCAACAAATTGAAGCTATCATCGTGACAGGTACACGAGCAGAAGAACTTCAATTG
CGATTGAAGTTAGCAGAGGTTGAAGTACCAATTATAGTTGAGCGTGATATTTATAAAGCA
ACGGCAAAGACTATGGATTATAAAGGTTTCACAGTTGCAATACCAAACTATACATCATTA
GCGCCTATGCTTGAACAATTAAACCGTTCGTTTGAAGGAGGTCAATCA

Protein Sequence (untagged):

RQWTAIHLAKLARKASRAVGKRGTDLPGQIARKVDTDILRKLAEQVDDIVFISGTNGKTT
TSNLIGHTLKANNIQIIHNNEGANMAAGITSAFIMQSTPKTKIAVIEIDEGSIPRVLKEV
TPSMMVFTNFFRDQMDRFGEIDIMVNNIAETISNKGIKLLLNADDPFVSRLKIASDTIVY
YGMKAHAHEFEQSTMNESRYCPNCGRLLQYDYIHYNQIGHYHCQCGFKREQAKYEISSFD
VAPFLYLNINDEKYDMKIAGDFNAYNALAAYTVLRELGLNEQTIKNGFETYTSDNGRMQY
FKKERKEAMINLAKNPAGMNASLSVGEQLEGEKVYVISLNDNAADGRDTSWIYDADFEKL
SKQQIEAIIVTGTRAEELQLRLKLAEVEVPIIVERDIYKATAKTMDYKGFTVAIPNYTSL
APMLEQLNRSFEGGQS

Oligonucleotide sequence:

Forward primer (SUMOInffwd)	GCGAACAGATCGGTGGTAGACAGTGGACGGCAATCCATC
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGATTGACCTCCTTCAAACGAACGG

Construct #6

Name: OPPF12137
Amplified region: 4 - 1311
Vector: pOPINF
Description: N-His tag full length murT

DNA Sequence:

AGACAGTGGACGGCAATCCATCTAGCGAAATTGGCGCGTAAAGCAAGTAGAGCAGTAGGT
AAAAGAGGAACAGATTTACCTGGACAAATCGCTAGAAAAGTGGATACAGATATATTAAGA
AAATTAGCAGAGCAAGTTGATGATATTGTATTTATCAGTGGAACAAATGGTAAAACAACG
ACTTCAAACTTAATTGGACATACTTTAAAAGCAAATAATATTCAAATTATACACAATAAT
GAAGGTGCTAATATGGCTGCAGGTATAACTTCTGCATTCATCATGCAATCAACACCTAAG
ACTAAAATTGCGGTAATCGAAATTGATGAAGGTTCGATTCCACGTGTGTTAAAAGAAGTT
ACACCTTCAATGATGGTATTTACTAATTTCTTTAGAGATCAAATGGATCGCTTCGGTGAA
ATTGATATTATGGTTAATAACATTGCAGAGACAATTAGTAATAAAGGCATCAAATTATTG
CTAAATGCTGATGATCCATTTGTGAGTCGTTTGAAAATCGCAAGTGATACGATTGTGTAC
TATGGTATGAAAGCACATGCCCATGAATTTGAACAAAGTACGATGAATGAAAGTAGATAT
TGTCCAAACTGTGGTCGCTTATTGCAATACGATTATATTCATTATAATCAAATTGGTCAT
TATCACTGTCAGTGTGGTTTCAAACGAGAGCAAGCAAAATATGAAATATCAAGTTTTGAT
GTGGCACCGTTTTTATATTTAAATATCAATGATGAAAAATATGATATGAAAATTGCAGGT
GACTTTAACGCTTATAACGCGTTAGCAGCATATACTGTTTTAAGAGAGCTAGGGTTAAAT
GAACAAACAATTAAAAATGGCTTTGAAACGTATACATCAGACAATGGTCGTATGCAGTAC
TTTAAAAAAGAACGAAAAGAAGCGATGATCAATTTAGCTAAAAATCCTGCAGGAATGAAT
GCAAGTTTATCAGTTGGTGAACAATTAGAAGGCGAAAAAGTGTATGTTATTTCGCTAAAT
GATAACGCTGCAGATGGTCGAGATACTTCATGGATTTATGATGCAGATTTTGAAAAATTA
TCTAAGCAACAAATTGAAGCTATCATCGTGACAGGTACACGAGCAGAAGAACTTCAATTG
CGATTGAAGTTAGCAGAGGTTGAAGTACCAATTATAGTTGAGCGTGATATTTATAAAGCA
ACGGCAAAGACTATGGATTATAAAGGTTTCACAGTTGCAATACCAAACTATACATCATTA
GCGCCTATGCTTGAACAATTAAACCGTTCGTTTGAAGGAGGTCAATCA

Protein Sequence (untagged):

RQWTAIHLAKLARKASRAVGKRGTDLPGQIARKVDTDILRKLAEQVDDIVFISGTNGKTT
TSNLIGHTLKANNIQIIHNNEGANMAAGITSAFIMQSTPKTKIAVIEIDEGSIPRVLKEV
TPSMMVFTNFFRDQMDRFGEIDIMVNNIAETISNKGIKLLLNADDPFVSRLKIASDTIVY
YGMKAHAHEFEQSTMNESRYCPNCGRLLQYDYIHYNQIGHYHCQCGFKREQAKYEISSFD
VAPFLYLNINDEKYDMKIAGDFNAYNALAAYTVLRELGLNEQTIKNGFETYTSDNGRMQY
FKKERKEAMINLAKNPAGMNASLSVGEQLEGEKVYVISLNDNAADGRDTSWIYDADFEKL
SKQQIEAIIVTGTRAEELQLRLKLAEVEVPIIVERDIYKATAKTMDYKGFTVAIPNYTSL
APMLEQLNRSFEGGQS

Oligonucleotide sequence:

Forward primer (Opti3CInffwd)	AAGTTCTGTTTCAGGGCCCGAGACAGTGGACGGCAATCCATC
Reverse Primer (Infusion 3' site)	ATGGTCTAGAAAGCTTTATGATTGACCTCCTTCAAACGAACGG

Construct #7

Name: OPPF12136
Amplified region: 4 - 1311
Vector: pOPINE
Description: C-His tag full length murT

DNA Sequence:

AGACAGTGGACGGCAATCCATCTAGCGAAATTGGCGCGTAAAGCAAGTAGAGCAGTAGGT
AAAAGAGGAACAGATTTACCTGGACAAATCGCTAGAAAAGTGGATACAGATATATTAAGA
AAATTAGCAGAGCAAGTTGATGATATTGTATTTATCAGTGGAACAAATGGTAAAACAACG
ACTTCAAACTTAATTGGACATACTTTAAAAGCAAATAATATTCAAATTATACACAATAAT
GAAGGTGCTAATATGGCTGCAGGTATAACTTCTGCATTCATCATGCAATCAACACCTAAG
ACTAAAATTGCGGTAATCGAAATTGATGAAGGTTCGATTCCACGTGTGTTAAAAGAAGTT
ACACCTTCAATGATGGTATTTACTAATTTCTTTAGAGATCAAATGGATCGCTTCGGTGAA
ATTGATATTATGGTTAATAACATTGCAGAGACAATTAGTAATAAAGGCATCAAATTATTG
CTAAATGCTGATGATCCATTTGTGAGTCGTTTGAAAATCGCAAGTGATACGATTGTGTAC
TATGGTATGAAAGCACATGCCCATGAATTTGAACAAAGTACGATGAATGAAAGTAGATAT
TGTCCAAACTGTGGTCGCTTATTGCAATACGATTATATTCATTATAATCAAATTGGTCAT
TATCACTGTCAGTGTGGTTTCAAACGAGAGCAAGCAAAATATGAAATATCAAGTTTTGAT
GTGGCACCGTTTTTATATTTAAATATCAATGATGAAAAATATGATATGAAAATTGCAGGT
GACTTTAACGCTTATAACGCGTTAGCAGCATATACTGTTTTAAGAGAGCTAGGGTTAAAT
GAACAAACAATTAAAAATGGCTTTGAAACGTATACATCAGACAATGGTCGTATGCAGTAC
TTTAAAAAAGAACGAAAAGAAGCGATGATCAATTTAGCTAAAAATCCTGCAGGAATGAAT
GCAAGTTTATCAGTTGGTGAACAATTAGAAGGCGAAAAAGTGTATGTTATTTCGCTAAAT
GATAACGCTGCAGATGGTCGAGATACTTCATGGATTTATGATGCAGATTTTGAAAAATTA
TCTAAGCAACAAATTGAAGCTATCATCGTGACAGGTACACGAGCAGAAGAACTTCAATTG
CGATTGAAGTTAGCAGAGGTTGAAGTACCAATTATAGTTGAGCGTGATATTTATAAAGCA
ACGGCAAAGACTATGGATTATAAAGGTTTCACAGTTGCAATACCAAACTATACATCATTA
GCGCCTATGCTTGAACAATTAAACCGTTCGTTTGAAGGAGGTCAATCA

Protein Sequence (untagged):

RQWTAIHLAKLARKASRAVGKRGTDLPGQIARKVDTDILRKLAEQVDDIVFISGTNGKTT
TSNLIGHTLKANNIQIIHNNEGANMAAGITSAFIMQSTPKTKIAVIEIDEGSIPRVLKEV
TPSMMVFTNFFRDQMDRFGEIDIMVNNIAETISNKGIKLLLNADDPFVSRLKIASDTIVY
YGMKAHAHEFEQSTMNESRYCPNCGRLLQYDYIHYNQIGHYHCQCGFKREQAKYEISSFD
VAPFLYLNINDEKYDMKIAGDFNAYNALAAYTVLRELGLNEQTIKNGFETYTSDNGRMQY
FKKERKEAMINLAKNPAGMNASLSVGEQLEGEKVYVISLNDNAADGRDTSWIYDADFEKL
SKQQIEAIIVTGTRAEELQLRLKLAEVEVPIIVERDIYKATAKTMDYKGFTVAIPNYTSL
APMLEQLNRSFEGGQS

Oligonucleotide sequence:

Forward primer (Kozak(pET/TriEx)Inffwd)	AGGAGATATACCATGAGACAGTGGACGGCAATCCATC
Reverse Primer (KH6-pTriExInfrev)	GTGATGGTGATGTTTTGATTGACCTCCTTCAAACGAACGG

E. Folding of proteins expressed as inclusion bodies

Protocol 1 (Adapted from [103])

i) Solubilization of proteins from inclusion bodies

The pellet from 4L of cell culture is resuspended in 250ml of 50mM Tris-HCl pH8.0, 100mM NaCl, 5mM EDTA, 0,1%NaN_3, 0,5% Triton X-100, 0,1mM PMSF and 1mM DTT (immediately added before use). This is followed by ultrasound sonication by cycles of 30sec.

After sonication add 10mM MgSO4 to chelate EDTA, and then add DNAseI (~0.01mgml^{-1}) and lysozyme to about 0.1mg.ml^{-1} to the lysate and incubate at room temperature for 20min.

Centrifuge at 30 000*g* for 30min at 4ºC to collect inclusion bodies. Crush the pellet with a spatula, then resuspend it completely by sonication in the lysing buffer. (Another portion of DNAse and lysozyme can be added at this point to improve the purity of the pellet)

Repeat the centrifugation twice without adding DNAse and lysozyme.

Wash the inclusion bodies again with 50mM Tris-HCl pH8.0, 100mM NaCl, 5mM EDTA, 0,1%NaN_3. Collect the final inclusion bodies pellet by centrifugation.

ii) Dissolve the washed inclusion bodies

Add 30-40ml of 100mM Tris-HCl pH8.0 with 50mM glycine to the inclusion bodies pellet.

Disperse the pellets completely by sonication and then dissolve the suspension dropwise, stirring vigorously in 100mM Tris-HCl pH8.0 with 50mM glycine and 8.5M urea (200mL for every 4L of culture)

Add 5mM GSSH (reduced glutathione) and 0.5mM GSSG (oxidized glutathione) and stir overnight at 4ºC.

This is now ready to refold by dialysis.

iii) Refolding by dialysis

Refolding buffer (RB): 0.1M Tris-HCl pH8.0, 0.4M L-arginine, 1mM EDTA, 1µg.ml^{-1} protease inhibitors, 0.2mM DTT

Dialyze against RB + 4M urea for 24h.

Dialyze against RB + 2M urea for 24h.

Dialyze against RB + 1M urea for 24h.

Dialyze against RB for 24h.

Dilute the buffer in a 1:4 ratio and dialyse against it for 24h.

Dialyse against 10mM Tris-HCl pH8.0, 150mM NaCl, 1mM EDTA, 5mM DTT.

To harvest the solubilized protein, centrifuge for 20min at 10 000g at 4ºC to remove the precipitant.

Protocol 2 [104]

http://dwb4.unl.edu/Chem/CHEM869N/CHEM869NLinks/www.nwfsc.noaa.gov/protocols/inclusion.html

i) Solubilization of proteins from inclusion bodies

The pellet from 1L of bacterial suspension is resuspended in 20ml of 50mM HEPES pH7.5, 0.5M NaCl, 1mM PMSF, 5mM DTT containing 0.35mg.ml^{-1} lysozyme and then incubated for 30min at 20ºC. Triton X-100 is added to a concentration of 1% (v/v). This is followed by ultrasound sonication by bursts of 30sec followed by cooling until the solution clears.

The extract is treated with DNAseI for 1h at 37ºC in a concentration of 20mg.ml^{-1}.

The inclusion bodies are sedimented by centrifugation at 30 000*g* for 30min at 4ºC.

The pellet (inclusion bodies) is washed twice with PBS containing 1% Triton X-100 followed by centrifuging at 30 000*g* for 30min at 4ºC.

The pellet is solubilized in 2ml 50mM HEPES pH7.5, 6M guanidine-HCl, 25mM DTT and left for 1h at 4ºC.

Insoluble material is removed by centrifugation at 100 000g for 10min.

Determine protein concentration and adjust to 1mg.ml-1 using 50mM HEPES pH7.5, 6M guanidine-HCl, 25mM DTT and proceed directly to folding.

ii) Folding protocol

The solubilized proteins are diluted as quickly as possible 1:10 into cold (4ºC) folding buffer: 50mM HEPES pH7.5, 0.2M NaCl, 1mM DTT, 1M NDSB256

The final protein concentration should not exceed 0.05 to 0.1mg.ml^{-1}. A fast and efficient mix is essential. For small volumes dispensing the protein solution with a pipette directly into the folding buffer while vortexing is adequate – keep vortexing for 30sec after addition. For larger volumes one can dispense the protein solution into the folding buffer using the syringe under vigorous (magnetic stirrer) agitation. Keep stirring for 2min after addition. Leave for 1h at 4ºC. The remaining guanidine and NDSB can then be removed by dialysis into the appropriate buffer.

F. Thermofluor Screening Solutions

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
100mM Sodium phosphate pH 6.2	100mM Potassium phosphate pH 8.0	100mM MOPS pH 7.0	100mM Bis-Tris pH 7.0	100mM HEPES pH 6.5	100mM HEPES pH 8.5	100mM MES pH 5.5	100mM Tris pH 8.5	100mM Tris pH 7.5	100mM Bis-Tris pH 5.5	100mM MES pH 6.5	100mM HEPES pH 7.5
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
100mM Sodium phosphate pH 6.2 150mM NaCl	100mM Potassium phosphate pH 8.0 150mM NaCl	100mM MOPS pH 7.0 150mM NaCl	100mM Bis-Tris pH 7.0 150mM NaCl	100mM HEPES pH 6.5 150mM NaCl	100mM HEPES pH 8.5 150mM NaCl	100mM MES pH 5.5 150mM NaCl	100mM Tris pH 8.5 150mM NaCl	100mM Tris pH 7.5 150mM NaCl	100mM Bis-Tris pH 5.5 150mM NaCl	100mM MES pH 6.5 150mM NaCl	100mM HEPES pH 7.5 150mM NaCl
C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
100mM Sodium phosphate pH 6.2 500mM NaCl	100mM Potassium phosphate pH 8.0 500mM NaCl	100mM MOPS pH 7.0 500mM NaCl	100mM Bis-Tris pH 7.0 500mM NaCl	100mM HEPES pH 6.5 500mM NaCl	100mM HEPES pH 8.5 500mM NaCl	100mM MES pH 5.5 500mM NaCl	100mM Tris pH 8.5 500mM NaCl	100mM Tris pH 7.5 500mM NaCl	100mM Bis-Tris pH 5.5 500mM NaCl	100mM MES pH 6.5 500mM NaCl	100mM HEPES pH 7.5 500mM NaCl

"There is a saying around structural biologists, that protein crystallization is half art and half science. Like science, however, every art is not only based on talent but also on hard work and practice." [105]