Overproduction and purification of Haa1 protein envisaging functional and structural studies of this transcription factor involved in yeast adaptive response to acetic acid

Ricardo Aguilar Andrade Ribeiro

Thesis to obtain the Master of Science Degree in

Biotechnology

Supervisor: Prof. Dr. Isabel Maria de Sá-Correia Leite de Almeida

Examination Committee

Chairperson: Prof. Dr. Leonilde de Fátima Morais Moreira

Supervisor: Prof. Dr. Isabel Maria de Sá-Correia Leite de Almeida

Members of the Committee: Dr. Margarida Isabel Rosa Bento Palma

November 2016

Acknowledgments

First of all, I would like to acknowledge my supervisor, Professor Isabel Sá-Correia for provide me the opportunity to join the Group of Biological Sciences Research Group (BSRG) to develop such interesting work. I'm thankful for her patience, motivation, guidance and the help provided to improve my analytical and writing skills.

I would also like to express my gratitude to Sílvia Henriques, for the immense support during all this stage, guidance to develop the thesis's experimental part and my writing scientific skills.

I would also like to acknowledge Dr. Mário Rodrigues, with whom we have carried out Atomic Force Microscopy experiments at Faculdade de Ciências da Universidade de Lisboa. In this matter, a very special thanks to Paulo Neves, who so kindly helped me in the experiments and data analysis.

A very fond thanks to my friends and family for always being present.

This work was supported by Project contracts PTDC/BBB-BEP/0385/2014-"Structural and functional analysis of the Haa1 transcription factor required for yeast response and resistance to acetic acid" and PTDC/FIS-NAN/6101/2014-"Molecular and Mechanical Forces in Biology measured with Force Feedback Microscopy through "Fundação para a Ciência e a Tecnologia" (FCT) /MEC (PIDDAC). Funding received by iBB-Institute for Bioengineering and Biosciences from FCT (UID/BIO/04565/2013) and Programa Operacional Regional de Lisboa 2020 (Project N. 007317) is also acknowledged.

Index

Figure Index

Table Index

Notation

AFM	Atomic Force Microscopy
ATP	Adenosine triphosphate
BMGY	Buffered Glycerol-complex
BMMY	Buffered Methanol-complex
DBD	DNA-binding domain
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E.coli	*Escherichia coli*
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum associated protein degradation
ESR	Environment stress response
FPLC	Fast Protein Liquid Chromatography
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IMAC	Ion Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria Broth
MS	Mass spectrometry
P.pastoris	*Pichia pastoris*
PAD	Pichia Adenine Dropout
pK_a	$-\log(K_a)$, where K_a is the acid dissociation constant
ROS	Reactive oxygen species
S.cerevisiae	*Saccharomyces cerevisiae*
SGD	*Saccharomyces* Genome Database
TAD	Transactivation domain

TOR	Target of Rapamycin
UPR	Unfolded protein response
YNB	Yeast Nitrogen Base
YPD	Yeast Peptone Dextrose

Abstract

The transcription factor Haa1 is the main regulator of yeast genomic response to acetic acid stress, regulating, directly or indirectly, the transcription of 80% of the acetic acid-activated genes. Using error-prone PCR, a more tolerant Haa1 mutant (S135F), with a serine residue exchange to a phenylalanine residue in position 135, was selected in a previous work. This Haa1 mutation leads to increased transcriptional activation of Haa1-target genes, presumably due to the effect of altered $Haa1_{S135F}$ structure in binding interactions with the target genes promoter binding sites. In this work, it was attempted the expression of the Haa1 protein using the *Pichia pastoris* host system. Given that this strategy was not successful, the *E.coli*-pET expression system was used to overproduce the his-tagged-Haa1 and $Haa1_{S135F}$ proteins, followed by their purification using Ion Metal Affinity Chromatography and subsequent polishing by Fast Protein Liquid Chromatography.

In order to get some insight into the complexes formed between the Haa1 proteins and the promoter region of one of their target genes, *TPO3*, Atomic Force Microscopy was used. Haa1 and $Haa1_{S135F}$ proteins were found to be bind at 1/3 of the *TPO3* promoter region, where the Haa1-binding site maps. Preliminary results appear to suggest the existence of differences in the height of Haa1 proteins bound to DNA, presumable related with their binding.

Keywords: Transcription factor, Haa1, *Saccharomyces cerevisiae*, Acetic acid response, Recombinant protein expression, Atomic Force Microscopy

Resumo

O fator de transcrição Haa1 é o regulador principal da resposta genómica da levedura a stress por ácido acético, regulando, de forma direta ou indireta, 80% dos genes, cuja transcrição é ativada em resposta a este ácido. Num trabalho anterior, com base na introdução de erros por PCR, foi gerado e selecionado um mutante Haa1-S135F (por substituição de uma serina por uma fenilalanina na posição 135) que apresenta uma maior tolerância a ácido acético. Esta mutação leva ao aumento da ativação da transcrição dos genes-alvo do Haa1, possivelmente devido a alterações na estrutura proteica a qual resulta na alteração da força de ligação com os locais de ligação nos promotores dos genes-alvo.

Neste trabalho, tentou-se superproduzir extracelularmente a proteína Haa1, usando o sistema de expressão em *Pichia pastoris*. Dado que esta estratégia não foi bem sucedida, o sistema de expressão *E.coli*-pET foi usado para superproduzir as proteínas Haa1 e $Haa1_{S135F}$—fundidas a uma cauda de histidinas, seguido de purificação usando cromatografia de afinidade com iões metálicos imobilizados e um polimento posterior, usando FPLC.

De modo a analisar os complexos formados entre as proteínas Haa1 e a região promotora de um dos genes-alvo (*TPO3*) foi usada Microscopia de Força Atómica. Foi possível observar as ligações das proteínas Haa1 e $Haa1_{S135F}$ numa posição cerca de 1/3 da região promotora do gene *TPO3*, onde foi mapeado o local de ligação do Haa1. Resultados preliminares parecem sugerir que existam diferenças no perfil de alturas das proteínas Haa1 e $Haa1_{S135F}$ resultante da ligação ao DNA.

Palavras-chave: Factor de transcrição, Haa1, *Saccharomyces cerevisiae*, Resposta a ácido acético, Expressão de proteína recombinante, Microscopia de Força Atómica

1 Introduction

1.1 Cell response to environment stress

All different living cell types have the ability to respond to changes in environmental conditions, in particular to environmental insults. Undertaking the complexity of such cellular responses is one of the major challenges in biology. Triggering such responses requires a complex sensing network and signal transduction that will lead to in cell growth adaptation and proliferation. This process involves gene expression modulation (controlling transcription initiation and/or chromatin remodelling) in response to the different stresses and, consequently, the attenuation of metabolic activity and other features of the cell. [1 2]

Cell stress can be defined as environmental conditions that can compromise cell survival, or prevent the cell from performing optimally. [1 2]

In nature, single-celled organisms such as yeasts have to cope with different environmental challenges simultaneously and it is crucial to assure response mechanisms to guarantee cell survival and to maintain their proliferative capacity and growth at the new condition. Environmental challenges can be of chemical or physical nature such as temperature, radiation, pressure, presence of certain ions, high or low osmolality, chemical toxicants, pH, nutrient depletion or bioavailability. [3]

Cell response to a stress involves cellular changes as the direct consequence of stress exposure and/or damage and defence mechanisms are triggered. After the adaptation process cells may resume growth and proliferate when exposed to a sub-lethal stress. Cell response to stress conditions is aimed at protecting cells from detrimental effects and at repairing cellular damage, if possible. The response will lead to an acquisition of stress tolerance. Once cells are submitted to a mild stress, they become more tolerant to the same or to other more severe stresses. Acquisition of stress tolerance is considered one of the main purposes of cellular stress response. It has been clear that when the cell is exposed to one type of stress this may lead to increased tolerance to other types of stress as well (cross-tolerance), suggesting that different stress stimuli might involve common cell responses in order to overcome stress conditions.[3 4 5 6 7 8]

Environment stress response (ESR) is usually transient, and the response involves a number of changes in genomic expression after the exposure to a new environment. However, gradually over time, the expression differences usually diminish, and transcript levels return to near pre-stress levels. This transient pattern represents an adaptation phase before resuming growth and proliferation.[9]

Among the different genes, whose transcription is activated in response to environmental stress, are those related to carbohydrate metabolism, fatty acid metabolism, maintenance of the cellular redox potential, solutes transport, protein folding and degradation, vacuolar and mitochondrial functions, intracellular signalling, detoxification of reactive oxygen species (ROS), cell wall modulation, autophagy and secretion, DNA-damage repair mechanisms, and others. [3 10] Transcriptional repressed genes of the ECR are related with protein synthesis and growth-related functions. [3 10]

The severity of a stress affects the response profile in a significant way; the more severe a stress, the longer it may take for a cell to respond, regarding signal transduction pathways and gene

expression. For this reason, it is important to study stress responses over time and to monitor changes, complementing information by doing comparative gene and protein expression analysis on mutant and wild type stains for a better elucidation of the mechanisms related to stress responses. A gene is considered a tolerance determinant to a specific stress when its deletion or increased expression leads to a phenotype of susceptibility/tolerance. However, cell cultures may have an heterologous population concerning the stress response. For instance, mRNA levels detected by Northern blot or other techniques represent the average values of whole cell population, not the values for the stress response of an cell that may be more or less affected by the deleterious effect of a particular level of stress, compared with the average. [1]

1.2 The physiological action of carboxylic acids in yeast

1.2.1 Weak acids as food preservatives

Man has been using applying natural preservatives for centuries, even without knowing how they protect food from spoiling. One source of natural occurring preservatives are organic acids, which include acetic, benzoic, propionic and sorbic acid. These acids are currently approved for use in large-scale food and soft drink preservation (Figure 1). Acetic and lactate are also widespread in nature as microbial fermentation products. [2 11]

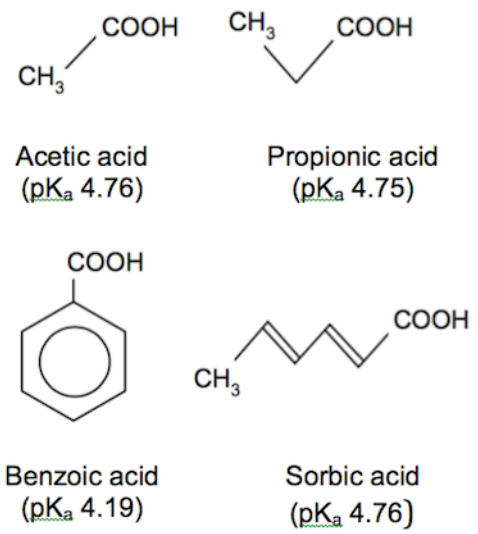

Figure 1 – Some of the carboxylic acids legally approved as food preservatives and the corresponding pKas. Adapted from Mollapour et al. (2008) [12]

Usually, the conditions imposed by many preserved food materials in the food and beverage industry do not represent a suitable environment for microbial growth. These antimicrobial conditions include low pH, low water activity (*aw*), presence of high levels of carbon dioxide preservatives and/or of ethanol, absence or low concentrations of oxygen. However, yeasts and fungi represent a major threat regarding acidic food spoilage, being able to proliferate and grow at low pH in the presence of excess of legally-permitted concentrations of weak acid preservatives used, thus preventing the decrease of its levels. [2 12 13]

Spoilage fungi, whose physiological properties mainly define their spoilage behaviour, can be distinguished in 3 different groups. Group 1 consists in the most dangerous spoilage yeasts, resistant to acid preservatives, osmotolerant, vitamin requiring and highly fermentative, leading to bottle explosions due to excessive gas formation. [14] Group 2 consists in opportunistic yeasts able to cause spoilage following manufacturing mistakes. The majority of contaminant yeasts are considered in Group 3 (do not cause spoilage and serve as an hygiene indicator). [15] Among these, *Zygosaccharomyces bailii* is the most notorious species of Group 1 of spoilage fungi [12]

The resistance of spoilage yeasts to weak acids poses a serious problem regarding weak acid usefulness as preservative agents, bringing major economic losses. [13]

Acetic acid is also a byproduct of *S.cerevisiae* alcoholic fermentation, and together with high concentrations of ethanol and other toxic metabolites, it may also contribute to fermentation arrest and thus decreased ethanol productivity. This weak acid, among others, is also present in lignocellulosic hydrolysates, which represents a promising substrate for sustainable bioethanol production and for integrated biorefineries in industrial biotechnology. [16]

Understanding the molecular basis of yeast weak acid response and resistance is important in order to elaborate more efficient preservation strategies and engineer more robust industrial strains. [13 16 17]

1.2.2 Molecular mechanisms underlying the action of weak acids

At low pH of the growth medium, depending of the concentration present, weak acids may have a strong effect on microbial growth and viability. Weak organic acids such as acetic, benzoic, lactic or sorbic acids are present in solution in a pH-dependent equilibrium, coexisting the dissociated and undissociated forms. Weak acids have a p*Ka* value, representing the pH at which half of the molecules dissociate in the counter-anion. The undissociated form (XCOOH) of the acid will increase, as the pH decreases from the p*Ka* value. At the neutral cytosolic pH, the dissociated form ($XCOO^-$) prevails. [2 12 17 18]

In low pH cultures, the undissociated form of the weak acid will enter in yeast cell by simple diffusion across the lipid bilayer. Therefore, the uncharged acid will diffuse across the plasma membrane to the cytoplasm more easily than the charged form. The charged form is retained if a specific permease is not present to facilitate diffusion or active transport. Once the acid passes through the plasma membrane to the cytoplasm in the undissociated form, it will dissociates in the cell

as a consequence of the near neutral intracellular pH of the cytosol, generating a proton and the acid anion (H^+, $XCOO^-$, respectively). The charged anion accumulates in the cytosol because it will not readily diffuse through the cellular lipid bilayer of the cell. [12 17 19] In addition, the accumulation of weak acids counter-anions in the cytosol causing an increase in internal turgor pressure and the consequent inhibitory effects. [20]

The protons present in the cytosol will lead to a decrease in the intracellular pH (pHi), resulting in decreased metabolic activity, affecting signal transduction, protein interactions and cell division rate. Such intracellular acidification of the cell due to weak acid dissociation can be counteracted by the combined effects of intracellular buffering and the action of ATP-driven proton efflux pump (Pma1), promoting proton extrusion from the cell (figure 2 B). The proton efflux generates an electrochemical potential across the cell membrane (ZΔpH) that is disrupted by weak acids, affecting the active transport of nutrients and active transport toxic metabolites out of the cell. [21 22 23] Pumping protons out of the cell and anions through the activity of plasma membrane H^+ATPase together with the inhibition of the glycolytic pathways by the weak acids will lead to ATP depletion (figure 2 A). [24 25 20 26 27 28 21]

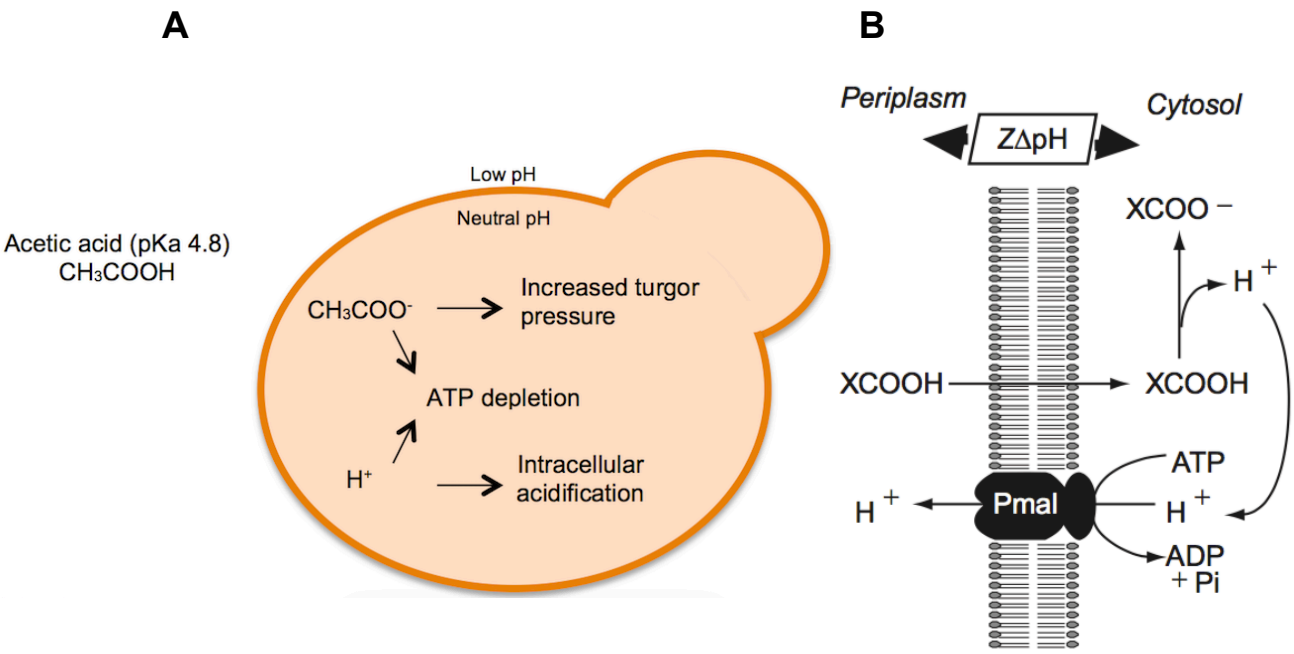

Figure 2 - A) Effects of weak acids stress in yeast. As an example of a weak acid, acetic acid (pKa 4.8) exists in a predominant undissociated form below the acid pKa. A diffusional flux of the acid through the plasma membrane to the cytosol, which is at a near intracellular neutral-pH, will result in the weak acid dissociation-mediated generation of protons and anions. B) Schematic representation of how weak acid acts on yeast plasma membrane. A near-neutral pH in the intracellular medium will lead to the dissociation of the acid. The anion ($XCOO^-$) will accumulate inside the cell. An electrochemical potential is maintained across plasma membrane (ZΔpH) by the activation of plasma membrane H^+ATPase mediating proton extrusion. This electrochemical potential is dissipated in the presence of weak acids. Adapted from Takagi, H *et al.* (2015) [10] and Mollapour *et al.* (2008) [12]

However, the inhibitory effects of weak acids in yeast are not exclusively related to intracellular acidification and anion accumulation. Stress responses and physiological adaptation mechanisms might differ according to the weak acid lipophilicity. [10 12 17 20] It was been reported that acetic acid and more lipophilic acids, such as benzoic or sorbic acids, affect intracellular pH differently. [20] Studies suggest that monocarboxylic acids become more potent inhibitors as they become more lipophilic. [29 30 31] For instance, high levels of acetic acid are required (80-150mM) to inhibit S.*cerevisiae* growth at pH 4.5, but only 1-3 mM of sorbic acid are required. Sorbic acid is strongly inhibitory to membrane trafficking. [12 32 33]

Concerning the physiological action of acetic acid in yeast, a Crabtree positive yeast *S.cerevisiae* strain do not possess the capacity to metabolize acetic acid in the presence of glucose due to catabolic repression, thus, it cannot use acetate as a carbon source as it happens in the highly tolerant *Zygosaccharomyces* yeasts. Yeast plasma membrane monocarboxylate proton antiporters for acetate uptake, as well as tricarboxylic acid cycle enzymes for the acetate catabolism in the mitochondria and glyoxylate cycle enzymes for acetate assimilation in the peroxisome are all subject to glucose repression. [32] When low pH cultures of *S.cerevisiae* are exposed to acetic acid, the undissociated form is of sufficiently low molecular weight to enter in the cell by facilitated diffusion through the open plasma membrane glycerol channel Fps1p, and/or by passive diffusion across the plasma membrane. Once inside the cell, acetic acid will dissociate leading to an increase of the intracellular pool of acetate anions, triggering a multifaceted adaptation stress response (figure 4). [13 15 20 35 34] In order to counteract acidification of cytosol and dissipation of H^+ gradient, plasma membrane H^+-ATPase Pma1 and the vacuolar H^+-ATPase (V-ATPase) activity increases (figure 4). [11 35 36]

1.3 Adaptative response to acetic acid: the Haa1 transcription factor and the Haa1-regulon

Haa1 transcription is considered the main regulator of yeast adaptive genomic response to acetic acid, since approximately 80% of the acetic acid activated genes are regulated, indirectly or directly, by this transcription factor. [37] Being a transcription factor, Haa1 is composed by a DNA binding domain (DBD) and a Transactivation domain (TAD). The DBD is able to recognize specific DNA sequences in the promoter region of the transcription factor target genes. The TAD is involved in transcription factor activation and often includes binding sites for other proteins such as transcription coregulators.[38] Haa1 has a DNA binding domain (DBD) in the first 124 N-terminal residues, and a transactivation domain (TAD) in the remaining C-terminal region (a.a. 125-694). Haa1p was firstly included in a family of copper-regulated transcription factors based on its high degree of homology at the level of the DNA binding domain (DBD) with the copper-responsive transcription factor Ace1p, and thought to be involved in copper regulation and tolerance. [39] The DBD of Haa1 and Ace1 includes a conserved zinc module and a Copper Regulatory Domain. [39] The protective role exerted by *HAA1* is observed for acetic propionic and lactic acids [40] Moreover, several genes from the Haa1-regulon are themselves important determinants of tolerance to acetic acid. [40 41]

The Haa1p-depedent genes encode proteins clustered in several functional classes such as those related with multidrug resistance transporters (42%), cell wall (4.6%), lipid metabolism (6.2%), regulation of carbohydrate metabolism (12.3%), protein folding (4.6%), carbohydrate metabolism (9.2%), amino acid metabolism (3.1%), nucleic acid processing (4.6%) and transcription factors (9.2%), which illustrates the very important role of Haa1 at the different physiological levels. [37 42]

Among the acetic acid-responsive genes, *TPO2* and *TPO3, AQR1*, which encode three Multidrug Resistance transporters (MDR) belonging to the Major Facilitator Superfamily (MFS), were described to exert a role in protection against acetic acid, apparently by mediating the active expulsion of acetate from the cell interior. [40 42] These genes are transcriptionally activated in response to acetic acid, dependent on the Haa1 transcription factor, suggesting that the higher acetate accumulation phenotype observed in the deletion mutants for *HAA1* might be due to lack of transcriptional activation of these transporters in response to acetic acid. [37 40] The mapping of the promoter region of *TPO3* allowed the identification of a functional binding site for Haa1 [5'- (G/C)(A/C)GG(G/C)G-3'], designated as Haa1-responsive element (HRE), in which it was found that Haa1p binds to *TPO3* promoter region, proving that *TPO3* is a Haa1-direct target gene. [13]. A list of genes presumed to be directly regulated by Haa1 is indicated in figure 3.

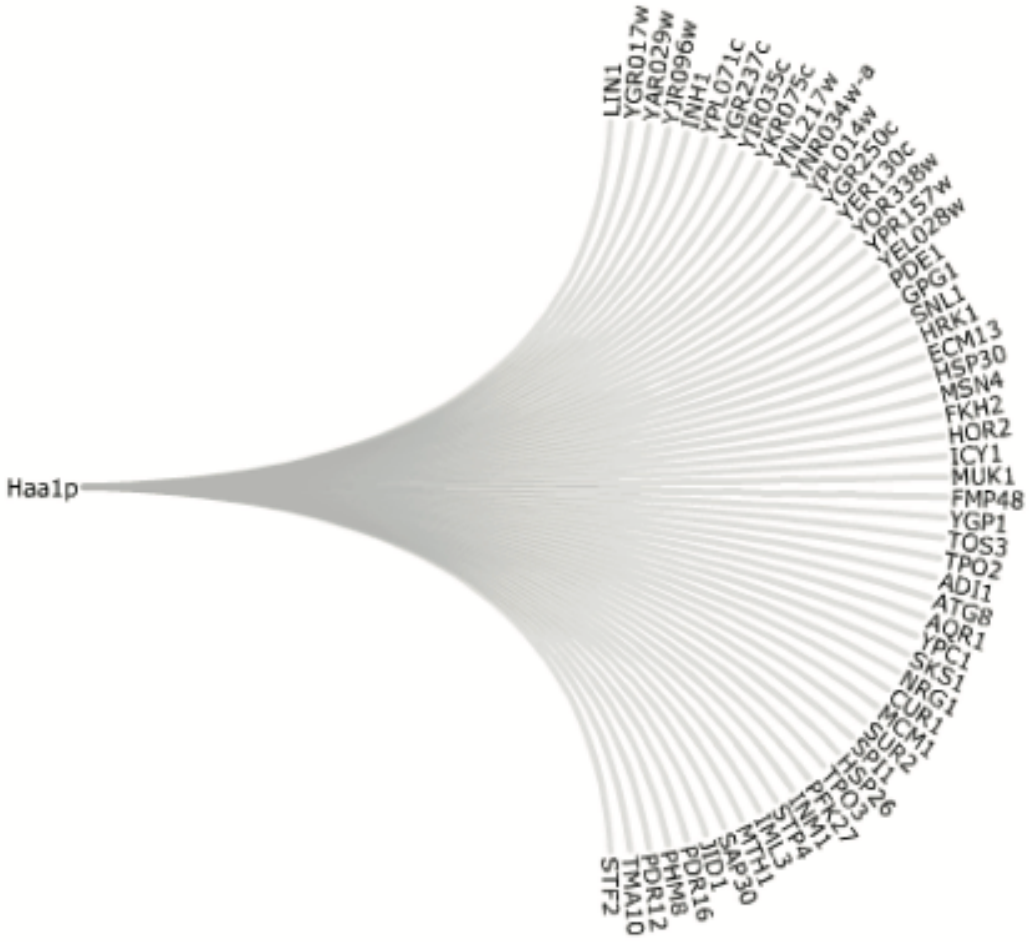

Figure 3 - Acetic acid-induced genes presumably directly regulated by Haa1 transcription factor, having the minimal functional Haa1-responsive element [5'- (G/C)(A/C)GG(G/C)G-3')] in their promoter region. Mira *et al.* 2011 [13]

Additionally, among the genes that are regulated by Haa1 transcription factor, the expression of *SAP30* and *HRK1* genes are the most relevant upon acetic acid stress resistance. *HRK1* encodes a protein kinase belonging to a subfamily of protein kinases implicated in activation of the plasma membrane H^+-ATPase Pma1 in response to glucose metabolism. [37] [43] Acetic acid challenged Δ*hrk* and Δ*haa1* mutant yeast cells showed increased accumulation of acetic acid. This may lead to speculate that the protein kinase is involved in the Hrk1 regulation by phosphorylation of plasma membrane acetate transporters. [35] Another Hrk1p target is Gph1p, a cytosolic enzyme involved in the catabolism of glycogen, whose expression seems to be protective against acetic acid stress.

Other relevant Haa1 target*SAP30* encodes a subunit of the Rpd3L histone deacetylase complex, playing a pivotal role in the regulation of transcription. Elimination of *SAP30* gene led to a strong susceptibility phenotype to acetic acid. [35] *HSP30* encodes a hydrophobic plasma membrane protein

whose expression is activated by heat shock, ethanol and weak acid stress. [23] The transcription of *HSP30* gene is highly induced upon acetic acid stress in a Haa1-dependent manner. [35] Cell wall proteins encoded by *SPI1* and *YGP1* genes also have a protective role against the diffusional entry of the acid. [40] *SPI1* is involved in the decrease of cell porosity, whose transcription is activated in dependence of Haa1 transcription factor in the presence of acetic acid. [44] *YGP1* is a cell wall-related secretory glycoprotein. [40 45 46]

Haa1 is also involved the adaptative response to lactic acid. This response involves Haa1 transcription factor translocation from the cytosol to the nucleus to enable its action, promoting the upregulation of target genes. [47] The extent of phosphorylation of Haa1 was found to be important for its nuclear translocation that is mediated by Msn5. Msn5 is a member of the importin β family of nuclear transport receptors that was reported to interact with Haa1 and its phosphorylated form is favourably exported from nucleus to cytosol. [48 49] Under lactic acid stress Haa1 appears to be essentially located mainly in the nucleus, exhibiting a lower degree of phosphorylation. In the absence of stress, Haa1 appears to be phosphorylated. [47] Therefore, it was suggested that nuclear accumulation of Haa1 that happens under lactic acid stress might be related with its phosphorylation state, particularly with a reduction in the degree of phosphorylation. [47] Recently, a similar mechanism, underlying Haa1 activation under acetic acid stress through the translocation from cytoplasm to nucleus leading to a transcriptional activation of Haa1-target genes was observed in our laboratory. [50]

Additionally, based on a quantitative proteomic approach, it was found that lethal concentrations of acetic acid causes severe intracellular amino-acid starvation, and the adaptive response involves the general amino-acid control system through the TOR-signalling pathway, leading to apoptotic cell death. [51]

Interestingly, on the highly resistant *Z.bailli* species, that can catabolize acetate even in the presence of glucose, acetic acid resistance seems to involve an increased activity of different carbohydrate metabolic pathways (Mdh1p, Aco1p, Cit1p, Idh2p, Lpd1p) together with an increased ATP production (Atp1p, Atp2p) in order to assure cell detoxification. Other mechanisms such as activation of oxidative and general responses (Sod2p, Dak2p, Omp2p) were also reported. [52]

Furthermore, a lipidomic profiling of the major lipid species in the plasma membrane revealed that acetic acid induce changes in the cellular lipid content of both *S.cerevisiae* and *Z.bailii* acetic acid-challenged cells. The level of complex sphingolipids was found to be higher in *Z.bailii* than in *S.cerevisiae* cells, suggesting a link between high sphingolipid content and acetic acid tolerance. [53]

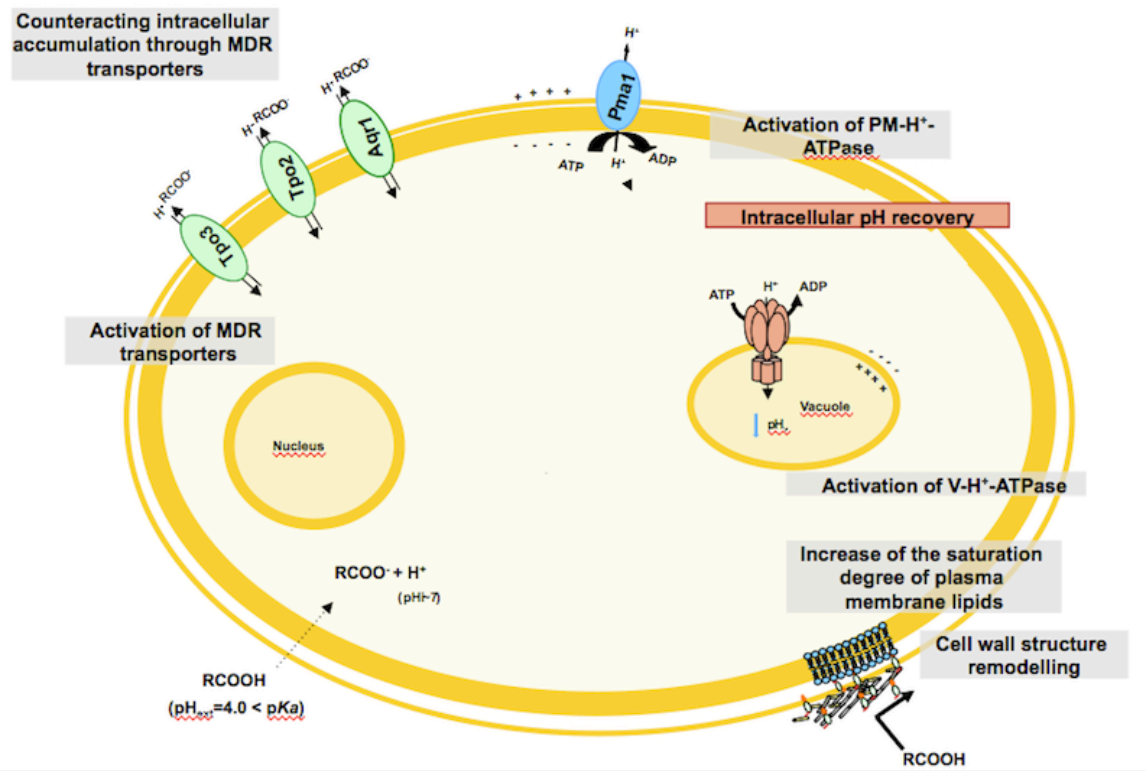

Figure 4 - Mechanistic model for the adaptive yeast response to weak acid stress. The low pH of the external environments favours the entrance of the undissociated form of the weak acid (RCOOH) into the cell. Once in the cytosol, the weak acid dissociates, leading to RCOO- accumulation in the cytosol. To counteract intracellular acidification H+-ATPases present in plasma membrane and vacuolar membranes are activated. Cell wall remodelling contributes to reduce the diffusion rate of undissociated weak acids and reduce weak acid-induced plasma membrane damage. Adapted from Mira et al. (2010) [42]

1.4 Overexpression and purification of recombinant proteins: host cell/ expression systems

Recombinant protein production requires the selection of an appropriate host/expression system, dependent of the characteristics of the protein and intended application. [54]

Yeast expression systems, unlike *E.coli* prokaryotic systems, combine the advantages of unicellular organisms, ease genetic manipulation and growth, with the additional benefits of a eukaryote, including improved folding, assembly, disulphide bond formation and most post-translational modifications. [55] [56] The two most utilized yeast strains are *Saccharomyces cerevisiae and Pichia pastoris*. *S.cerevisiae* has the first yeast routinely used for recombinant expression due to deep knowledge about its genetics, biochemistry and physiology. Furthermore, this study involves the overproduction of Haa1, a *S.cerevisiae* transcription factor, and a yeast host/expression system,

contrary to *E.coli* bacterium, is capable of carrying out the post- transcriptional modifications of proteins.

However, overexpressing an *S.cerevisiae* transcriptional factor using *S.cerevisiae* as the cell host organism may impose an excessive metabolic burden to the cell. In addition, heterologous protein secretion in *S.cerevisiae* is often limited to relatively low levels due to hyperglycolisation and mis-folding of proteins inside the cell. [57] [58] Comparing to *S.cerevisiae, P.pastoris* may have an advantage in the glycosylation of secreted proteins because it is not as prone as the budding yeast to hyperglycosylation. In *P.pastoris,* N-linked oligosaccharides are usually no more than 20, while in *S.cerevisiae* are 50-150 residues per side chain. [59] [60]

One of the widely used *P.pastoris* commercial protein expression systems is PichiaPink™ Expression system (Invitrogen), declaring the production of high protein yields, scalable production and all the advantages of a eukaryotic expression system. [61] In addition, this system allows easy selection of expression clones using *ADE2* complementation of adenine auxotrophy as a selection marker to select *P.pastoris* cells containing the gene of interest, rather than using antibiotic resistance as a selection marker. [61] A number of works were successfully developed using this overexpression system. [62] [63] *P.pastoris*, unlike bacterial expression systems, is a eukaryotic system that has the ability to perform the post-translational modifications occurring in *S.cerevisiae* and in more complex eukaryotes: correct folding and assembly, disulphide bond formation, O- and N- linked glycosylation of signal sequences. [55] [56] Additionally, the preference of this yeast species for respiratory growth is a key physiological trait in this context to obtain high-cell density cultures without generating significant amounts of toxic metabolites as it is the case of ethanol. [55] Furthermore, the choice to use inducible an *AOX1* promoter that is tightly regulated holds advantage for heterologous gene expression, since it is strongly repressed by glycerol or glucose and, upon depletion of these carbon sources, the promoter is de-repressed and fully induced in the presence of methanol, allowing to uncouple the phase of growth from the protein production phase. In that sense, cells are not stressed by the accumulation of recombinant protein during growth phase. [64] In addition, *P.pastoris* has the ability to secrete properly folded, pos-translationally processed and functional recombinant proteins in the supernatant together with the fact that *P.pastoris* secrete low levels of native proteins, meaning that the secreted heterologous protein constitutes the vast majority of the total protein in the culture supernatant, which is convenient for recovery and purification steps.[65] [66] Vectors from the commercial PichiaPink™ Expression system (Invitrogen) also allow the N-terminal α factor signal sequence to allow protein secretion, tagging of the protein with a six Histidine tag to allow subsequent purification and detection, and a TEV cleavage sequence to allow removal of the affinity tag from purified recombinant fusion protein. [61]

Although *P.pastoris* has been successfully used to produce various recombinant heterologous proteins, the variability in efficiency in protein secretion is problematic. In fact, translocation determined by the secretion signal peptide, processing and folding in the Endoplasmic Reticulum (ER) and Golgi, and secretion out of the cell are reported as physiological limitations that compromise protein production. [67] [68] [69] [70] [71] Concerning protein synthesis and translocation, the *S.cerevisiae* α mating sequence used in this work, fused with the recombinant protein, mediates the entry in the

secretory pathway and is one of the most commonly secretion signal peptides used to drive secretion in *P.pastoris.* The *S.cerevisiae* α mating sequence is composed by a pre-region and a pro-region. The pre region is a 19-amino acid sequence, consisting of a amino-terminus, a hydrophobic core and a carboxy terminal polar region ending with a cleavage site, being responsible for directing the nascent protein post-translationally in the ER and is cleaved off subsequently by a signal peptidase. The 67-residue pro sequence is thought to be important for transferring the protein from ER to Golgi compartment, and is cleaved at the dibasic Kex2p site. [72 73 74 75 76] However, it is problematic when processing at the α mating sequence *Kex2* site does not occur, leaving a N-terminus with various portions of the α mating sequence still attached. Additionally, the N-terminus of some recombinant proteins fold on a structure that prevents processing enzymes from reaching their processing site. [55] To tackle these drawbacks, multiple strategies are reported to enhance α mating sequence secretory potential. Deletion site directed mutagenesis on a predicted α mating sequence resulted in 50% increased secretion of horseradish peroxidase in *P.pastoris.* [64 77] Other strategies such as directed evolution or the insertion of spacers have also been reported. [77 78]

Aside of protein retention, overproduction of recombinant proteins has repeatedly been shown to overload endoplasmic reticulum folding (ER), leading to accumulation of unfolded or misfolded proteins that may start to aggregate, triggering the unfolded protein response (UPR) pathway. UPR pathway is responsible for induction of genes that are involved in protein folding, which aims to reduce ER stress conditions. In parallel, UPR pathway is connected to protein degradation, since it is also responsible for triggering Endoplasmic reticulums associated protein degradation (ERAD) pathway, in which detected misfolded or unfolded proteins in the ER are re-translocated to the cytosol, where they are subjected to proteosomal degradation. [79 80 81 82] Also, it is reported that UPR pathway is induced upon changes in environmental conditions, namely upon methanol induction for protein production. [83]

Another system widely used for protein overproduction consists in using *E.coli* has the host cell, although some problems may arise when expressing a heterologous protein like *S.cerevisiae* Haa1 transcription factor, namely the lack of post-translational modifications. Also, yeast and bacteria use different codons as part of the translation machinery, which may lead to translational stalling, premature translation termination, translation or frameshifting of a eukaryotic protein in *E. coli* due to lack of the necessary tRNA molecules. [84] To overcome the latter problem, BL21-CodonPlus(DE3)-RIL *E.coli* cells is a suitable choice for heterologous protein expression, engineered to contain extra copies of genes that encode the tRNAs that most frequently limit translation of heterologous proteins in *E. coli*. BL21-CodonPlus(DE3)-RIL *E.coli* cells have extra copies of the argU, ileY, and leuW tRNA genes. These genes encode tRNAs that recognize the arginine codons AGA and AGG, the isoleucine codon AUA, and the leucine codon CUA, respectively.

In pET expression system (Novagen), one of the most widely and efficient systems to express heterologous protein in *E.coli*, the gene to be overexpressed is under the control of T7 promoter, which is strongly induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to the growth medium. [84] IPTG induces transcription of the T7 RNA polymerase gene, already present in the

genome of the bacteria used for overexpression, which then activates the T7 promoter and allows the expression of the recombinant gene, as schematized in figure 5.

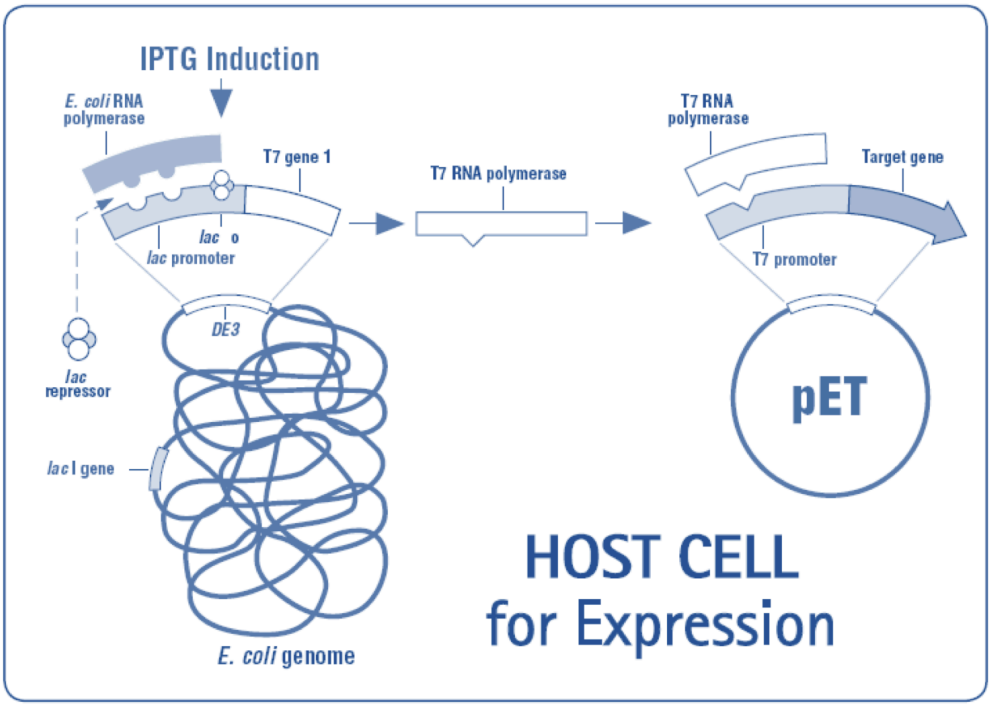

Figure 5 - Overexpression of target proteins using pET system (Novagen). Transcription of the gene of interest is controlled by the T7 promoter, activated by T7 RNA polymerase. This protein is expressed from the T7 gene *1*, present in the λDE3 prophage of *E. coli* BL21-CodonPlus(DE3)-RIL cells, and is under the control of the *lac* promoter. In the absence of IPTG, the *lac* repressor (expressed from the *lac* I gene, also present in the λDE3 prophage) binds to the *lac* operator (*lac o*) and represses transcription of T7 RNA polymerase. When IPTG is present, it is able to bind to *lac* repressor, allowing the transcription of T7 RNA polymerase, which further activates transcription of the target gene. In Novy & Morris, Novagen, Inc. [84]

The pET vector also allows a coding sequence for a C-terminal Histidine-tag fused to the recombinant protein, facilitating protein purification and detection. Histidine-tagged proteins have a high selective affinity for Ni^{2+} and several other metal ions that can be immobilized on chromatographic media using chelating ligands. Consequently, histidine-tagged proteins will be selectively bound to metal-ion-charged supports, while the other cellular proteins, that have a lower histidine content will bind weakly or not bind at all. A schematic purification procedure is shown in figure 6. When a higher degree of purity is required for either tagged or untagged recombinant proteins, a multistep purification may be necessary. [85] However, addition of purification steps will increase purity at the cost of decreased yield of active protein. [86]

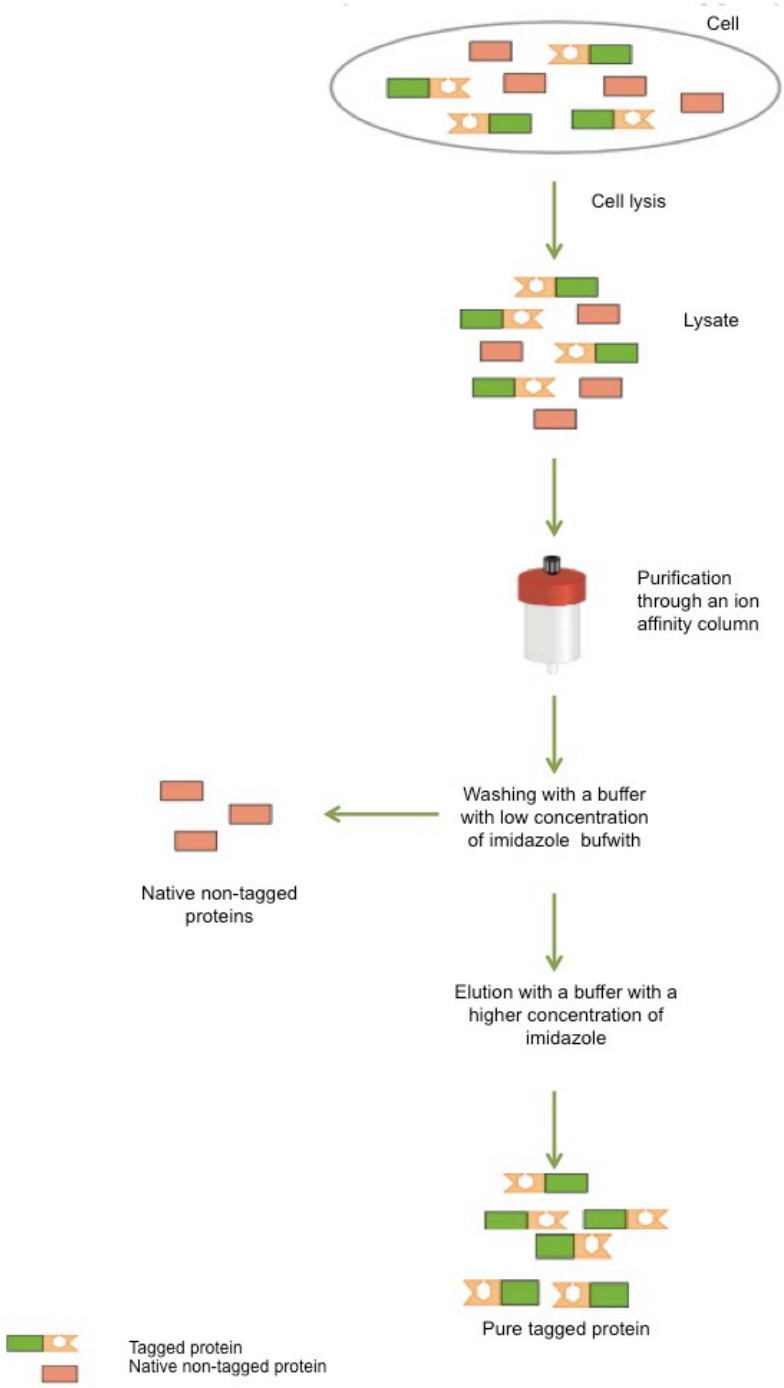

Figure 6 - General purification workflow for histidine-tagged proteins (assumes the use of Ni^{2+}-charged affinity supports or other metal ions that can be immobilized on chromatographic matrix). Adapted from GE Healthcare [85]

Another purification technique used in this work is Size Exclusion Chromatography (SEC). SEC separates molecules according to differences in size as they pass through a SEC medium packed in a column. SEC can be used in protein purification schemes, as a main purification procedure or as polishing step to increase protein purity, or for desalting and buffer exchange. To perform the separation of proteins of different sizes, the resin is packed into a column to form a packed bed. SEC media consist of a porous matrix of spherical particles with chemical and physical stability and inertness (lack of reactivity and adsorptive properties). The packed bed is equilibrated with buffer, which fills the pores of the matrix and the space between the particles. The liquid inside the pores - stationary phase - is in equilibrium with the liquid outside the particles - mobile phase. Samples are eluted isocratically without the need to use different buffers during the separation. A washing step using the running buffer is performed at the end in order to remove molecules that might have been retained on the column. The success of SEC depends mainly on choosing conditions that give sufficient selectivity during the separation. The selectivity of a SEC medium depends only on its pore size distribution. [85 86 87]

1.5 Application of Atomic Force Microscopy in yeast research

Atomic Force Microscopy (AFM) has several capabilities regarding not only the characterization of topographic details of surfaces from the submolecular level to the cellular level, but also the monitoring of dynamic processes of single molecules in physiological relevant solutions. [88 89 90]

AFM is a promising technique for the advance of science in many areas of biological research. A core component of the equipment is a micro-machined probe mounted at the end of a cantilever. A voltage-driven piezoelectric (PZT) transducer enables spatial positioning of the cantilever probe in the *x, y* and *z* directions with precision, converting a electrical stimuli in mechanical motion. The AFM head contains an optical deflection system based on a laser diode and a photodetector. The principle of AFM operation is based on the scanning probe in the *x-y* plane over the sample surface. Moreover, the interaction force between the probe and the sample surface is monitored by measuring the deflection (vertical bending) of the cantilever, detected by a laser beam and reflected to a photodiode. By recording every pixel scanned in the *x-y* plane it is possible to reconstruct the surface topography of the sample. [88 89 91]

A crucial factor for successful imaging using AFM is the specimen preparation using appropriate non-destructive methods. AFM imaging can be in air or in fluid. Flatness and biocompatibility are important criteria to select a substrate. Indeed, to observe biological structures in their native state, there must be well attached to a smooth and solid subtract to resist the lateral forces exerted by the scanning tip. [92] Muscovite mica ($KAl2_2(OH)_2AlSi_3O_{10}$), which is atomically flat with a negatively charged surface, is the most commonly used substrate for AFM imaging for double-stranded DNA, DNA-protein complexes, lipid films or animal cells. [93] Mica surface can be cleaved easily before sample deposition. Additionally, divalent cations, such Mg^{2+} or Ni^{2+} can be included in

deposition buffers, functioning as salt bridges to absorb negatively charged molecules, such as DNA, to the surface. The commonest methods for introducing the molecules on the substrate are the following: by placing a drop of solution in the surface, forming a drop of solution on a hydrophobic surface and then allow contact with mica surface or spraying the solution as an aerosol in the surface. For imaging in air, samples can be incubated on the surface for up to several minutes before rinsing and drying. [88 92 93]

Currently, AFM has been used in biological research to determine the function of membrane proteins, to explore and quantify interactions between DNA and ligands such as peptides or proteins or even to analyse the biomechanical and biochemical properties of cell surfaces at a nanometer resolution. [92 94 95]

Focusing on yeast studies, AFM can be used as a complementary tool to tackle some challenges as to visualise what is occurring at the cell surface and how a stress can have impact on the biophysical properties of the cell wall or other cell components or study proteins present in cell surface that have a pivotal role in adhesion and microbial infection. [96 97] Some of the relevant cell biology questions that can be addressed by AFM are schematized in figure 7.

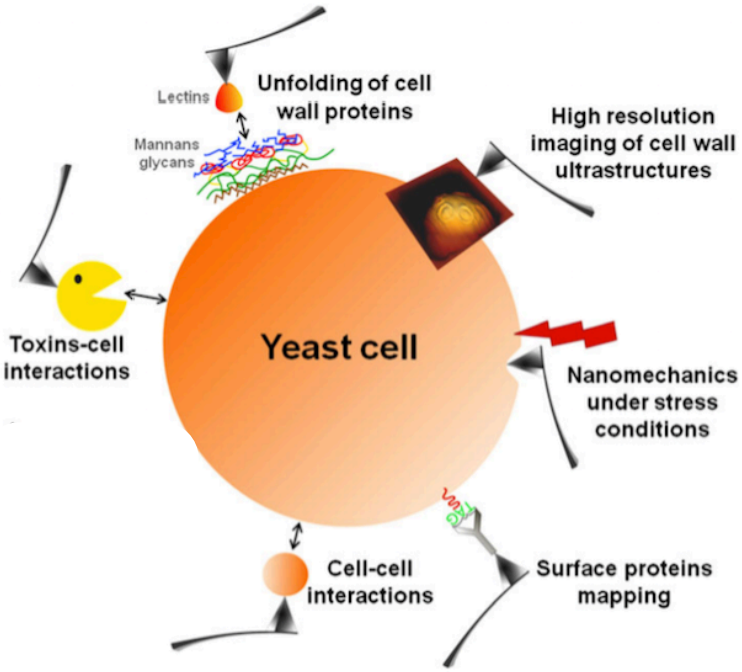

Figure 7 - Schematic presentation of a number of questions that can be addressed by AFM in cell biology research. Adapted from Francois et al., 2013 [96]

For instance, some work has been developed to study properties and response to stress in *Saccharomyces cerevisiae*. [97] [98] [99] [100] Studying the effect of heat stress on nanomechanical properties, AFM images shown that a temperature shift of the yeast culture from 30 to 42 °C induced in less than 1 h the formation of a circular structure that takes its origin at a particular location on the cell surface and evolved as concentric rings at the cell surface and cell elasticity increased by two-fold. This morphological process, taking place at the cell surface, was found to be dependent on genes required for the budding process, suggesting that these rings arise from a defective bud scar or bud emergence site during heat stress. [100]

AFM has been also used to study morphological changes in yeast cell surfaces caused by oxidative stress. For instance, by applying varying concentrations of hydrogen peroxide to *S.cerevisiae* and *Schizosaccharomyces pombe* cells, AFM revealed that oxidative stress caused cell compression in both *S.cerevisiae* and *S.pombe* and an increase of surface roughness. Also, by analyzing the yeast bud scar profile via AFM scanning, it was possible to see that the higher the hydrogen peroxide concentration, the larger was the number of aged yeasts. [101] Alsteens *et al.*(2008) compared the physical proprieties of mannoproteins of *Saccharomyces carlsbergensis* and *S. cerevisiae* using a functionalized AFM tip with concanavalin A, a tetrameric protein that has strong affinity to mannose residues. By analysis of elongation forces and rupture distances, it was found that, on *S. carlsbergensis,* only mannan oligosaccharides were stretched, while on *S.cerevisiae* both oligosaccharide and polypeptide chains of mannoproteins were stretched. According to the authors, the different physical properties of the mannoproteins between the two species may explain why the surface of *S.cerevisiae* is more hydrophobic than that of *S. carlsbergensis*.[102]

Studies focused on the interactions between antifungal drugs and yeast cell wall were also developed, using AFM as a complementary tool. For instance, by using AFM to study the caspofungin effects on the yeasts *Candida albicans* and *S. cerevisiae*, it was found that administration of this antifungal drug induced a deep cell wall remodeling in both yeast species. [103] [104] Quantitatively, the changes in cell wall composition were more pronounced in *C. albicans* cells, with a remarkable rise in chitin as a function of the caspofungin dose administrated to the *C. albicans* cells and with an increase in cell elasticity. [103] Also it was reported, using AFM as a complementary tool, that the exposure of *S.cerevisiae* cells to ethanol stress provoked a dramatic decrease of the cell wall rigidity (stiffness). [97]

Despite its remarkable benefits in cell biology, AFM still is a technology under development with some drawbacks, suffering from a poor temporal resolution (registering an image takes 10-20 minutes) and bears artifacts related to sample fixation and tip's geometry. [95] One major artifact is associated with tip-sample convolution, since any AFM image is a convolution of the shape of the probe and of the sample, having the effect of making protruding features appear wide, and holes appear smaller. [95]

2 Thesis outline

S.cerevisiae Haa1 transcription factor is the major player in the regulation of the yeast transcriptome in response to acetic acid stress, modulating the transcriptional activation of 80% of the acetic acid-responsive genes. Recently, a more acetic acid tolerant mutant, harbouring the S135F mutation (amino acid exchange at position 135 of a serine to phenylalanine), was selected from an *HAA1* library generated by error-prone PCR and enriched in acetic acid containing medium. This work was done in collaboration with the group of Alke Nevoight at Jacobs University Bremen gGmbH, in Germany. This mutation in Haa1 leads to increased transcriptional activation of Haa1-target genes, presumably due to the altered $Haa1_{S135F}$ structure in binding interactions with promoter binding sites.

The goal of this thesis work was to overexpress and purify Haa1 and its derived mutant $Haa1_{S135F}$ to allow the planned X-ray structural analysis of the corresponding DNA:protein complexes. This also allowed to conduct preliminary Atomic Force Microscopy experiments during this thesis work, aiming the visualization of Haa1 or the derived mutant protein complexed with *TPO3* target gene promoter sequence. An attempt to overproduce Haa1 using PichiaPink™ Expression system (Invitrogen)**,** using *Pichia Pastoris* as host cell was carried out, since *P.pastoris* has the ability to secrete properly folded, pos-translationally processed and functional recombinant proteins in the supernatant. Moreover, *P.pastoris* secrete low levels of native proteins, meaning that the secreted heterologous protein constitutes the vast majority of the total protein in the culture supernatant, a convenient trait for subsequent recovery and purification steps. However, the attempt of overproduce and secrete in the supernatant the his-tagged Haa1 protein using *P.pastoris* was not achieved and overproduction of the his-tagged Haa1 protein using *E.coli* as host cell was performed, using pET expression system (Novagen), using two constructs harbouring $HAA1_{WT}$ coding sequence and the derived mutant $HAA1_{S135F}$ coding sequence that were already available in lab (Silvia Henriques, unpublished results) The conditions for overexpression were also optimized (Mira *et al.* (2008). The overproduction of Haa1 and $Haa1_{S135F}$ his-tagged proteins using *E.coli* as cell host was achieved, and this was followed by purification using Ion Metal Affinity Chromatography and polishing, using Fast Protein Liquid Chromatography. Moreover, the purification of Haa1 and $Haa1_{S135F}$ allowed to conduct preliminary Atomic Force Microscopy experiments during this thesis work.

3 Materials and Methods

3.1 Strains, growth conditions and plasmids

The methylotrophic *Pichia pastoris* PichiaPink™ strain (*ade2*, *prb1*, *pep4*) was used for plasmid maintenance and to express the recombinant protein, using Pichia Pink™ Expression System from Invitrogen (figure 8).
For overexpression of the recombinant protein Histidine-tagged Haa1 and the derived mutant using pET expression system (Novagen), *Escherichia coli BL21-CodonPlus (DE3)-RIL* (*E. coli* B F– *ompT hsdS*(rB– mB–) *dcm*+ Tet r *gal* λ(DE3) *end A* Hte [*argU ileY leuW* Cam^r]) (Stratagene) was chosen.
The *Escherichia coli XL1-Blue* strain (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacI*q*Z_M15* Tn*10* (Tet)]) was the bacterial strain used for plasmid maintenance and general molecular cloning procedures.
Both *E.coli* and *P.pastoris* strains were preserved at -80°C in proper media enriched with 15%(v/v) glycerol. Prior to use, frozen cells were cultured in appropriate fresh media agar plates (20g/L agar). Bacterial cells were incubated at 37°C *overnight,* while yeast cells were incubated at 30°C until a generous amount of culture was reached.

For obtain suitable transformants, *Pichia* cells were cultured in Yeast Peptone Dextrose (YPD) medium, containing 20 g glucose (Merck), 20 g bactopeptone (Difco) and 10 g yeast extract (Difco) and selected on Pichia Adenine Dropout™ (PAD) agar lacking adenine. To express the recombinant protein, *Pichia* cells were cultured in Buffered Glycerol-complex (BMGY) medium (1%(m/v) yeast extract, 2%(m/v) peptone, 100mM potassium phosphate pH 6.0, 1.34%(m/v) YNB, 0.00004%(m/v)biotin and 1%(v/v) glycerol) and Buffered Methanol-complex (BMMY) medium (1%(m/v) yeast extract, 2%(m/v) peptone, 100mM potassium phosphate pH 6.0, 1.34%(m/v) YNB, 0.00004%(m/v)biotin and 1%(v/v) methanol). BMGY and BMMY media were both prepared from stock solutions Yeast Nitrogen Base with ammonium sulphate (YNB) without amino acids (13.4%(m/v) YNB), 0.02%(m/v) biotin, 5%(v/v) methanol, 10%(v/v) glycerol and 1M potassium phosphate buffer pH 6.0.
E.coli cells were cultured in rich solid or liquid Luria Bertani (LB) growth medium. For *E.coli* strain used for plasmid maintenance and cloning procedures, LB medium was supplemented with ampicillin (150mg/L), while for *E.coli* strain used for overexpression of the recombinant proteins, LB medium was supplemented with ampicillin (150mg/L) and chloramphenicol (30mg/L) to maintain selection during growth.

All plasmids used in this study are listed in table 1. The overexpression in *P.pastoris* cells of the Haa1 transcription factor was carried out, using pPINK-HC as expression vector, containing *S.cerevisiae* Alpha mating factor secretion signal sequence. Plasmid pPINK-HC(+)-$HAA1_{WT}$ –his6 contains the coding sequence of *HAA1* fused in frame with a N-terminal α factor signal sequence to allow protein secretion, a six Histidine tag to allow subsequent purification by Ion Metal Affinity Chromatography

(IMAC) using His-Trap® columns, and a TEV cleavage sequence to allow removal of the affinity tag from purified recombinant fusion protein. Plasmid pPINK-HC(+) has a selection marker for ampicillin resistance (*AmpR* gene) for *E.coli* and *ADE2* gene as a selection marker for *P.pastoris ade2* auxotroph strain used in this study. On the Pichia Pink™ Expression System (Invitrogen), *P.pastoris* grown in excess glycerol medium to repress protein expression, followed by methanol induction to initiate heterologous protein expression, since the gene to be overexpressed is under control of *AOX1* methanol inducible strong promoter.

The expression of Haa1 and the derived mutant (S135F) on *E.coli* was carried out using the pET expression system from Novagen (Figure 6). Plasmid pET23a(+)-$HAA1_{WT}$-His6 and pET23a(+)-$HAA1_{S135F}$-His6, containing the coding sequence of *HAA1* or the coding sequence of the derived fused in frame with a C-terminal six Histidine tail, was already available in the laboratory. (Henriques SH unpublished results).

The pET23a(+) expression vector has a selection mark for ampicillin resistance to maintain selective pressure in *E. coli* and is inducible by Isopropyl β-D-1-thiogalactopyranoside (IPTG). The temperature of bacterial growth, the concentration of IPTG and the time of cell harvesting after IPTG induction that maximize the production of soluble proteins in *E. coli* cells had been optimized before (Mira, 2008).

Table 1 - List of the plasmids used in this work

Plasmid	Description	Reference
pPINK-HC(+)-$HAA1_{WT}$	Encodes the Sc Haa1 transcription factor C-terminally fused in frame with a N-terminal α factor signal, a six-histidine tag and a TEV cleavage sequence	This study
pET23a(+)-$HAA1_{WT}$-His6	Encodes the Sc Haa1 transcription factor C-terminally fused to a six-histidine tag	Henriques,SH (unpublished results)
pET23a(+)-$HAA1_{S135F}$-His6	Encodes the Sc Haa1 transcription factor with a point mutation in residue 135 [Serine(S)--> phenylalanine (F) C-terminally fused to a six-histidine tag	Henriques SH (unpublished results)

3.2 Overexpression of Haa1 in *Pichia pastoris*

3.2.1 Amplification of *HAA1* from *S.cerevisiae* S288C genomic DNA

The amplification of *HAA1* ORF from *S.cerevisiae* S288C was carried out using a polymerase chain reaction (PCR). Each amplification by PCR was done containing 0.125 µl of genomic DNA, 5 µl of Phusion Buffer 10X (Thermo Fisher Scientific), 0.5 µl of 10mM dNTPs, 0.125 µl of each 0.1mM

primer solution , 1 µl of $MgCl_2$ (50mM) solution (Thermo Fisher Scientific), 0.75 µl of DMSO (100%) (Thermo Fisher Scientific), 17.125 µl of RNAfree bidistillate de-ionized water and 0.25 µl of Phusion polymerase (Thermo Fisher Scientific). PCR amplification cycling parameters were optimized and were as following: the preincubation at 98°C for 1 minute, 35 cycles of denaturation at 98°C for 10 seconds, followed by annealing step at 59°C for 20 seconds and an extension step at 72°C for 1 minute and 10 seconds. A final extension step was carried out at 72°C for 7 minutes. Amplification procedure was performed in a Thermal Cycler Block (BioRad). The oligonucleotide specific primers sequences used for the amplification of the selected gene were designed using OligoAnalyser Tool. Primers used are listed in Table 2. The *HAA1* gene sequence from *S.cerevisiae* S288C was retrieved from *Saccharomyces* Genome Database (SGD). The DNA amplification product was confirmed using a 0.8% agarose gel electrophoresis, carried out at 100W during 50 minutes, using *Kbplus* (Invitrogen) as ladder.

The amplified purified product was used as insert in further ligation and cloning procedures. Reaction mixtures with the amplification product were purified using DNA Clean and Concentrator Kit (Zymo Research) and eluted in 30 µl of RNAfree bidistillate de-ionized water.

Table 2 - Primers selected for the amplification of *S.cerevisiae* S288C *HAA1* gene. For primer design the following features were considered: grey - protection nucleotides; green – 6xHis-tag; orange – TEV cleavage site; purple – *Mly*I recognition site added to *HAA1* upstream fragment; blue – *Kpn*I recognition site added to *HAA1* downstream fragment

Primer	Amplification sequence	Length(nt)	Length of fragment (Kbp)
Forward	5'GAGTCCTTACCACCACCACCACCAC CACGAGAATCTTTACTTTCAAGGTATG GTCTTGATAAATGGCA 3'	68	2.1
Reverse	5'GGGGTACCTCATAACGAAGACAT GAAATTATCC 3'	33	

3.2.2 Restriction and ligation

For the *HAA1* gene sequence from *S.cerevisiae* S288C, both purified *HAA1* insert and pPinkα-HC (Figure 6, from Invitrogen included in the Pichia Pink™ Expression System Kit) were digested using restriction enzymes. For the cloning procedure, the insert was obtained from a PCR amplification experiment, using primers listed in Table 2, digested using *Kpn*I (from Takara – Clontech Laboratories, Inc.) and *Mly*I (from New England Biolabs) and cloned into the pPinkα-HC vector (Figure 8, from Invitrogen, part of the Pichia Pink™ Expression System Kit), previously digested with *Kpn*I and

*Stu*I. Digestion products were precipitated. The insert and pPinkα-HC vector were quantified using *Nanodrop* (ND-1000, Spectrophotometer). After digestion, the insert and pPinkα-HC vector were ligated, using T4 ligase (NEB). For this process, various ligations mixtures with different conditions were performed, using 20ng and 40ng of the vector, both with a molar ratio of vector: insert of 1:5 and 1:3. The control mixture was prepared containing only the vector. DNA ligation mixtures were incubated 2h at 22°C, and 16°C *overnight.* An additional incubation step was carried out for T4 ligase inactivation at 70°C during 5 minutes

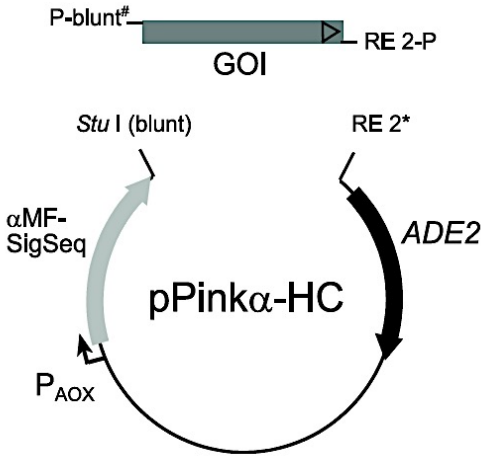

Figure 8 - pPinkα-HC ((from Invitrogen, included in the Pichia Pink Expression System) vector schematic view. In *PichiaPink ™ Expression System Guide.* [61]

The genetically engineered constructs were pPinkα-HC(+)-*HAA1*$_{WT}$, used for transformation of *E. coli XL1-Blue* (Stratagene) strain (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacI*q*Z_M15* Tn*10* (Tet)] using Classical Transformation technique. For this, all the ligation volume and 50 µl of TCM solution (10 mM $CaCl_2$, 10 mM $MgCl_2$, 10 mM Tris HCl pH 7.5) were added to 150 µl of competent *E. coli XL1-Blue* (Stratagene) cells, previously slowly defrosted on ice. The same procedure was applied to the control, but without the construct. The mixture was incubated on ice for 20 minutes, followed by heat shock at 42°C for 3 minutes. After incubation on ice for 5 minutes, 800 µl of liquid LB were added to each cell suspension and incubated for 1 hour, at 37 °C, with orbital agitation of 250 rpm. Cell suspension was centrifuged at 8000g for 3 minutes and resuspended in 100 µl for spreading in LB+ampicillin. Plates were incubated at 37°C *overnight* and the obtained transformants were selected for ampicillin resistance. Plasmid DNA was extracted from the obtained transformants using the alkaline lysis method QIAprep® Spin Miniprep Kit (QIAGEN) and quantified using *Nanodrop* (ND-1000, Spectrophotometer)

3.2.3 Testing *E.coli* candidate transformants

To test *E.coli* candidate transformants, the previously extracted plasmid DNA from *E.coli* transformants was used as DNA template and analysed by PCR, using primers listed in Table 3, to determine if the gene of interest was present. Amplification products were digested, using *Hind*III. Digestion products were confirmed using 0.8% agarose gel electrophoresis, carried out at 100V during 60 minutes, using *Kbplus* (Invitrogen) as ladder. Linearized vector was used as control, digested with *Kpn*I. To confirm digestion results, sequencing was carried out by STAB Vida Lda., using primers listed in Table 4

Table 3 - Primers used for the amplification of *S.cerevisiae* S288C *HAA1* gene, to test *E.coli* candidate transformants

Primer	Amplification sequence	Length(nt)	Length of the amplified fragment (Kbp)
Forward	5' GGAATTCGCCACCATGGTCTTGATA AATGGCATAAAG 3'	37	2.1
Reverse	5'GGGGTACCTCATAACGAAGACAT GAAATTATCC 3'	35	

Table 4 – Primers used for *S.cerevisiae* S288C *HAA1* gene sequencing

Primer	Sequence	Length(nt)
pHLsecHaa1-up	5' GGAATTCGCCACCATGGTCTTGATA AATGGCATAAAG 3'	37
HAA1-Sc-seq-interm.	5'TATACCACAAAGCCCGCCGT3'	20

3.2.4 Site directed mutagenesis to introduce a missing nucleotide

After sequencing the construct containing the *HAA1* gene (described in 3.2.2 section), it was found that there was a missing nucleotide Guanine (G) in position 9 within *HAA1* sequence. Taking this fact in consideration, site directed mutagenesis was carried out in order to obtain the correct sequence of $HAA1_{WT}$ within pPinkα-HC(+)-$HAA1_{WT}$, by incorporating the missing nucleotide. The primers were designed in order to be lined up back-to-back on either side of the area, where the missing nucleotide would be inserted. One of the primers contains the additional nucleotide that would

be inserted. Mutagenic primers are listed in Table 5. The mutagenic primers were previously phosphorylated to facilitate ligation and plasmid recircularization following amplification. The phosphorylation mixture contained: 36 µl of RNAfree bidistillate de-ionized water, 5 µl of Kinase buffer (NEB), 2 µl of 25mM $MgSO_4$ solution (Roche), 5 µl of each 100 µM primer stock, 1 µl of 100mM ATP and 1 µl of PNK (NEB). The mixture was incubated at 37°C for 45 minutes, followed by heat inactivation at 95°C for 5 minutes. Each mutagenic PCR was done containing: 33 µl of RNAfree bidistillate de-ionized water, 4 µl of Phusion Buffer 10X (Thermo Fisher Scientific), 0.6 µl of each 0.1 mM primer solution, 0.4 µl of 10mM dNTPs, 0,6 µl of DMSO (100%) (Thermo Fisher Scientific), 0,4 µl of plasmid DNA and 0.2 µl of Phusion polymerase (Thermo Fisher Scientific). PCR amplification cycling parameters were optimized and were as following: the preincubation at 98°C for 30 seconds, 25 cycles of denaturation at 98°C for 10 seconds, followed by annealing step at 62°C for 20 seconds and an extension step at 72°C for 8 minutes . Amplification procedure was performed in a Thermal Cycler Block (BioRad). 1 µl of *Dpn*I was added to the mixture in order to digest template methylated DNA, followed by circularization using T4 ligase (NEB). DNA ligation mixtures were incubated 2h at 22°C.

Table 5 - Mutagenic primers used in site-directed mutagenesis to replace the missing nucleotide Guanine (G) in *HAA1* gene. The replacing nucleotide is highlighted in blue.

Primer	Amplification sequence	Length(nt)	Length of the amplified fragment (Kbp)
Forward	5' GAT AAA TGG CAT AAA GTA TGC CTG TG 3'	26	10
Reverse	5' AAG ACC ATA CCT TGA AAG TAA AGA TTC TC3'	29	

3.2.5 Obtaining PichiaPink™ transformants by electroporation

To prepare electrocompetent cells, PichiaPink™ (from Invitrogen included in the Pichia Pink™ Expression System Kit) strain 4 (*ade2, prb1, pep4*) cells were grown at 30°C for 3 days and a single colony was used to pre-inoculate 20ml of YPD media in a sterile100ml flask at 30 °C at 250 rpm for 2 days. The starter culture was used to inoculate 100 ml of YPD media in a 1L flask in order to start a growth with 0.2 OD_{600nm}, at 30 °C at 250 rpm during 1 day. When the inoculum reached 1.3-1.5 OD_{600nm}, cell suspension was centrifuged at 2000g at 4°C during 5 minutes. Pellet was ressuspended in 20 ml YPD with 4ml HEPES buffer 1M and 250 µl DTT, followed by incubation at 30°C during 15

minutes without agitation. Cell suspension were bringing to 100 ml with cold bidistillate sterile water and centrifuged at 1500g at 4°C for 5 minutes. Pellet was washed by centrifugation at 1500G at 4°C during 5 minutes in 50 ml cold bidistillate sterile water and the process was repeated by washing with 4 ml cold sorbitol 1M and ressuspended in 300 µl cold sorbitol 1M.

For electroporation, 80 µl of electrocompetent PichiaPink™ cells were mixed with 10 µg of linearized pPinkα-HC(+)–*HAA1*$_{WT}$ construct in an ice-cold 0.2cm electroporation cuvette and pulsed according to the instrument manufacturer's instructions for yeast. Pulsed cells were pour in 1ml of cold sorbitol 1M and incubated at 30°C for up to 12 hours. After incubation, 300 µl of the cell mixture were spread on PAD selection plates and incubated at 30°C for 3-10 days until distinct white colonies were formed.

Total genomic DNA of PichiaPink™ integrants candidates was extracted and analysed by PCR to determine if the gene of interest has integrated into the genome for further use. Primers used are listed in Table 6. Each amplification by PCR was done containing 1 µl of genomic DNA, 10 µl of NZYTaq 2× Green Mater Mix (Nzytech), 1 µl of each 0.1mM primer solution and 7 µl of RNAfree bidistillate de-ionized water. PCR amplification cycling parameters were optimized and were as following: the preincubation at 95°C for 1 minute, 29 cycles of denaturation at 95°C for 30 seconds, followed by annealing step at 59°C for 30 seconds and an extension step at 72°C for 2 minute and 30 seconds. A final extension step was carried out at 72°C for 10 minutes.

Table 6 - Primers selected for the amplification of *S.cerevisiae* S288C *HAA1* gene to test *P.pastoris* integrants.

Primer	Amplification sequence	Length(nt)	Length of the amplified fragment (Kbp)
Forward	5' TCCAACAGCACAAATAACGGG 3'	21	2.3
Reverse	5' GGCGTGAATGTAAGCGTGAC 3'	20	

3.2.6 Overexpression of Haa1$_{WT}$ in *Pichia pastoris*

The overexpression of Haa1$_{WT}$ was carried out using PichiaPink™ integrants. 2 sets of 400ml BMGY medium in 2L flask were inoculated with a single colony and grown at 30°C, 250rpm for 48h. Cells were harvest by centrifugation at 1500g for 5 minutes at room temperature and resuspended in 20ml of BMMY medium in 100ml flask to induce recombinant protein expression. These cultures were either grown at 25°C or 30°C, at 250rpm for 96 hours. Every 24 hours 0.5% methanol was added repeatedly. To assess expression levels, 100 µl culture were centrifuged at 1500g for 10 minutes at room temperature and both pellet and supernatant were collected separately and frozen at -80°C until

further use. The established time course to collect samples was the following: 0,6,12,24,36,48,72 and 96 hours. Growth was followed by measuring culture OD_{600nm} .

To test the purity of the eluted fractions of $Haa1_{WT}$ recombinant protein 15ul of culture supernatants or cell free extracts were supplemented with 3ul of 5× SDS sample buffer without reducing agent, and loaded onto a 8% non-reducing SDS polyacrylamide gel. PageRuler®plus PreStained Protein Ladder (Thermo Fisher Scientific) was used as ladder. After electrophoresis, the proteins were transferred to a nitrocellulose membrane using Trans-Blot Turbo Transfer System (Biorad) at 115V for 10min. The membrane was incubated for 1 hour with 15mL of blocking solution (80 g/l NaCl, 2 g/l KCl, 6.1 g/l Na_2HPO_4 and 2 g/l KH_2PO_4, pH 7.4, 0.05 %v/v Twin-20, 5.0 %w/v non-fat dry milk). The membrane was then incubated with 10 µl of his-probe mouse monoclonal IgG_1 (Santa Cruz Biotechnology) and left at 4°C *overnight* with gentle agitation (35-55rpm). After incubation, membrane was washed 3 times, each time 10 minutes with 15ml PBST buffer (80 g/l NaCl, 2 g/l KCl, 6.1 g/l Na_2HPO_4 and 2 g/l KH_2PO_4, pH 7.4, 0.05 %v/v Twin-20) and then incubated 1 hour with 8µl of secondary antibody goat anti-mouse IgG (Santa Cruz Biotechnology). After that time, the membrane was washed 3 times with 20ml PBST (10 minutes each wash). The signal was detected by chemiluminescence using Western Lighting® Ultra kit (Perquin Elmas)

3.3 Overexpression of $Haa1_{WT}$ and the derived mutant $Haa1_{S135F}$ in *E.coli*

The overexpression of Haa1 and the derived mutant harbouring the mutation S135F was carried out using E*. coli BL21-CodonPlus (DE3)-RIL* (Agilent Technologies) as host cells to express the recombinant proteins. Since Haa1 protein is a eukaryotic protein that harbors a large number of arginines in its sequence, and to avoid codon usage that might lead to translational stalling or premature translation termination, *E.coli BL21-CodonPlus (DE3)-RIL* (Agilent Technologies) was a suitable choice as host cell, since this strain contains extra copies of genes enconding tRNAs which recognize arginine, leucine and isoleucine codons.[105]

An aliquot of E*. coli BL21-CodonPlus (DE3)-RIL RIL* (Agilent Technologies) cells, previously slowly defrosted on ice, was transformed with 2µl of pET23a(+)-$HAA1_{WT}$-His6 and or with the pET23a(+)-$HAA1_{S135F}$-His6 plasmid, using Classical Transformation technique. The mixture was incubated on ice for 30 minutes, followed by heat shock at 42°C for 1 minute. After incubation on ice for 50 minutes, 1ml of liquid LB was added to each cell suspension and incubated for 1 hour, at 37°C, with a orbital agitation of 250 rpm. Cell suspensions were then centrifuged at 8000G for 1 minute and resuspended in 100µl of LB for spreading in LB+ampicillin plates. Plates were incubated at 37°C *overnight* and the obtained transformants were selected for ampicillin resistance.

A single colony of the transformants obtained was cultivated in 25ml of LB medium in a 100ml flask, supplemented with ampicillin (150 mg/L) and chloramphenicol (30mg/L) *overnight* with orbital agitation of 250 rpm. These cells were reinoculated in 500ml of LB medium with ampicillin (150 mg/L) and chloramphenicol (30mg/L) in order to start a growth at 37°C with 0.05 OD_{640nm}, with orbital

agitation of 250 rpm, until cell suspension reached an initial 0.4 OD_{640nm}. Isopropyl β-D-1-thiogalactopyranoside (0.3mM) was added to the culture to induce recombinant protein expression, at 20°C to prevent protein aggregation, for 8.30 hours, with orbital agitation of 250 rpm. The final cell culture was centrifuged at 10000g (Eppendorf Centrifuge 5804) for 8 minutes at 4°C and cell pellets were frozen at -80°C until further use.

3.3.1 Purification by Ion Metal Affinity Chromatography and Fast Protein Liquid Chromatography of $Haa1_{WT}$ and the derived mutant $Haa1_{S135F}$

To each millilitre of cell culture volume, 50 µl of lysis buffer (10 mM imidazole, 10mM Tris, 0.625 mM NaCl) and 4 µl of protease inhibitor (100mM PMFS) were added and the cell suspension was kept in ice. Cell suspensions were sonicated to promote cell disruption using Sonifier® 250 sonicator, which were subjected to 4 sonication cycles for 30 seconds interspersed with 30 seconds on ice, using Duty cycle of 50% and an output control of 4.5. Cell free extracts were centrifuged at 25.000g for 30 minutes at 4°C and supernatant was collected and frozen at -80°C until further use.

Since Haa1 and the derived mutant included a C– terminal 6×His tag, the purification was achieved by Ion Metal Affinity Chromatography (IMAC), using a nickel coated column (HisTrap ® HP). The column was first washed with water and equilibrated with buffer A (10 mM imidazole, 10mM Tris, 0.625 mMNaCl). The supernatant was loaded on the column after being clarified by centrifugation at 25.000g for 30 minutes at 4°C, and filtered through a Millipore™ syringe filter. The column was again washed with buffer A, and with 5% buffer B (500 mM imidazole, 10mM Tris, 0.625 mM NaCl) for 10 minutes to prevent contamination by any other protein that may have remained on the column. Protein elution was carried out using a stepwise increasing gradient of imidazole that ranged from 10mM to 500mM. Under these conditions, the recombinant protein $Haa1_{WT}$ and the derived mutant started to elute when imidazole concentration was above 150mM.

The purity of each collected protein fraction was assessed by SDS-PAGE with GelCode™ Blue Safe Stain (ThermoFisher Scientific) and Western Blot analysis.

The purest fractions were mixed and concentrated using Amicon® Ultra-15 Centrifugal Filters 10kDa cutoff (Merck Millipore), according to the manufacturer's instructions. The concentrated fractions of the recombinant protein Haa1 and the derived mutant were subjected to Fast Protein Liquid Chromatography (FPLC), using a 10/300 GL column (Healthcare®) containing Superdex® 75 matrix. The column was equilibrated with 10mM Tris-HCl, 200 mM NaCl, pH 7.4, at a constant flow rate of 1.2 ml/min. After sample injection, elution was performed with the same buffer used to equilibrate the column. Protein elution was followed by accompanying the increase in Abs_{280nm}. The purest fractions were mixed and concentrated using Amicon® Ultra-15 Centrifugal Filters 10kDa cutoff (Merck Millipore), according to the manufacturer's instructions and further analysed by SDS-PAGE with GelCode™ Blue Safe Stain (Thermo Fisher Scientific).

3.3.2 SDS-PAGE and Western blot Assay

To test the purity of the eluted fractions of Haa1 recombinant protein and its derived mutant, Western Blot and SDS-PAGE followed by protein staining were performed. Protein staining was carried out in both protein eluted fractions collected from Ion Metal Affinity Chromatography and Fast Protein Liquid Chromatography, while Western Blot was performed only in the protein eluted fractions collected from Ion Metal Affinity Chromatography.

For SDS-PAGE followed by protein staining, 12 µl of each protein eluted fraction were supplemented with 3ul of 5× SDS sample buffer without reducing agent, and boiled for 1 minute at 100°C. The mixture was loaded on a 8% non-reducing SDS polyacrylamide gel. PageRuler®plus PreStained Protein Ladder (Thermo Fisher Scientific) was used as ladder. The electrophoresis was run at 150 Volts for approximately 1.3 hours. The gel was stained with GelCode™ Blue Safe Stain (ThermoFisher Scientific) for 2 hours and destained 3 hours with water.

For Western Blot, 5 µl of each protein fraction were resolved on a 8% non-reducing SDS polyacrylamide gel and transferred to a nitrocellulose membrane using Trans-Blot Turbo Transfer System (Biorad) at 25V for 10min. Membrane was incubated for 1 hour with 15mL of blocking solution (80 g/l NaCl, 2 g/l KCl, 6.1 g/l Na_2HPO_4 and 2 g/l KH_2PO_4, pH 7.4, 0.05 %v/v Twin-20, 5.0 %w/v non-fat dry milk). The membrane was then incubated with 10 µl of his-probe mouse monoclonal IgG_1 (Santa Cruz Biotechnology) and left at 4°C *overnight* with gentle agitation (35-55rpm). After incubation, membrane was washed 3 times, each time 10 minutes with 15ml PBST buffer (80 g/l NaCl, 2 g/l KCl, 6.1 g/l Na_2HPO_4 and 2 g/l KH_2PO_4, pH 7.4, 0.05 %v/v Twin-20) and then incubated 1 hour with 8µl of secondary antibody goat anti-mouse IgG (Santa Cruz Biotechnology). After that time, the membrane was washed 3 times with 20ml PBST (10 minutes each wash). The signal was detected by chemiluminescence using Western Lighting® Ultra kit (Perquin Elmas)

3.4 Atomic force microscopy: visualization of $Haa1_{WT}$ and derived mutated $Haa1_{S135F}$ bound to *TPO3* promoter region

3.4.1 Preparation of *TPO3* promoter region

A 1000bp DNA fragment containing the Haa1 binding site (5′- GAGGGG-3′) was amplified from *S.cerevisiae* S288C by polymerase chain reaction (PCR). Each amplification by PCR was done containing 0.5 µl of genomic DNA, 5 µl of Phusion Buffer 10X (Thermo Fisher Scientific), 0.5 µl of 10mM dNTPs, 0.5 µl of each 0.1 mM primer solution, 2 µl of 50mM $MgCl_2$ (Thermo Fisher Scientific) solution, 0.75 µl of 100% DMSO (Thermo Fisher Scientific), 15 µl of RNAfree bidistillate de-ionized water and 0.25 µl of Phusion polymerase (Thermo Fisher Scientific). PCR amplification cycling parameters were the following: preincubation at 98°C for 1 minute, 35 cycles of denaturation at 98°C for 10 seconds, followed by annealing step at 62°C for 20 seconds and an extension step at 72°C for 1 minute and 10 seconds. A final extension step was carried out at 72°C for 7 minutes. Amplification

procedure was performed in a Thermal Cycler Block (BioRad). The oligonucleotide specific primers sequences used for the amplification of the selected gene were designed using OligoAnalyser Tool. Primers used to amplify *TPO3* promoter region are listed in table 7. The *HAA1* gene sequence from *S.cerevisiae* S288C was retrieved from *Saccharomyces* Genome Database (SGD). The DNA amplification product was confirmed using a 0.8% agarose gel electrophoresis, carried out at 100W during 60 minutes, using *Kbplus* (Invitrogen) as ladder.
The PCR product was gel-purified by using Zymno clean Gel DNA Recovery Kit (Zymno Research) and quantified using ImageJ software, by determination of agarose band intensity.

Table 7 - Primers selected for the amplification of *S.cerevisiae* S288C *TPO3* promoter region.

Primer	Amplification sequence	Length(nt)	Length of amplified fragment (Kbp)
Forward	5'AGCAACATAATTGACTGACCC 3'	21	1.0
Reverse	5'CATTTCTGTTTATCTTTGGCTAG 3'	23	

3.4.2 Atomic force microscopy imaging

The solutions containing DNA and Haa1 protein or the derived mutant were prepared in a ratio of 1:1 and 1:5 (DNA:protein). 1nM of dsDNA was either incubated with 2nM or 10nM of protein for 30 minutes at room temperature. Control solutions were also prepared, containing 1nM of dsDNA or 10nM of Haa1 recombinant protein. After incubation, 5mM of $MgCL_2$ solution, 10mM of HEPES (pH 6.5) were added.10 µl of the solution was adsorved on a freshly cleaved mica for 15 minutes. The drop was rinsed 7 times with 200 µl of RNAfree bidistillate de-ionized water and dried on air for approximately 10 minutes.
To obtain image of the DNA-protein complexes, PicoSPM system (Molecular Imaging, Inc.) was operated in tapping mode was used for these experiments. High resolution Cr-Au cantilevers (MikroMasch) were employed for imaging in air. Cantilever oscilation frequency was tuned to the resonance frequency of the cantilever (160 KHz), with a Force Constant of 5N/m. The 512×512 resolution images were obtained with a scan rate between 0.7-1Hz and a scan range between 1-5µm. Images were taken with based on topography and displayed without modification except for flattening to remove the background curvature of the mica surface. Images were processed for analysis using Gwyddion 2.45.

4 Results

4.1 Cloning procedures to express *HAA1* in *Pichia pastoris*

In an initial stage of this study, a recombinant plasmid pPINK-HC(+)-*HAA1*$_{WT}$ was constructed to allow the overproduction of Haa1 transcription factor in *Pichia pastoris* cells. Plasmid pPINK-HC(+) was used as a vector of expression. Standard genetic engineering techniques were used to generate recombinant pPINK™ vectors, containing the *HAA1* gene into the plasmid pPinkα-HC.
Before transforming *Pichia pastoris*, cloning procedures were performed in *E. coli XL1-Blue* strain to obtain suitable candidates, expressing the construct pPINK-HC(+)-*HAA1*$_{WT}$.

PCR amplification of *HAA1* coding sequence from *S.cerevisiae* S288C genomic DNA was performed. Since the main goal was the overexpression of the *HAA1* gene, the designed primers had to allow the amplification of a fragment with: restriction enzymes recognition sites, a 6 His-tag to allow immunodetection and purification of the overproduced protein and a TEV cleavage site to allow the removal of affinity tags from the purified protein for further studies. An amplified DNA fragment of approximately 2.1 Kbp, corresponding to the expected size of *HAA1* gene coding sequence, was obtained after electrophoresis of the PCR mixture (figure 9)

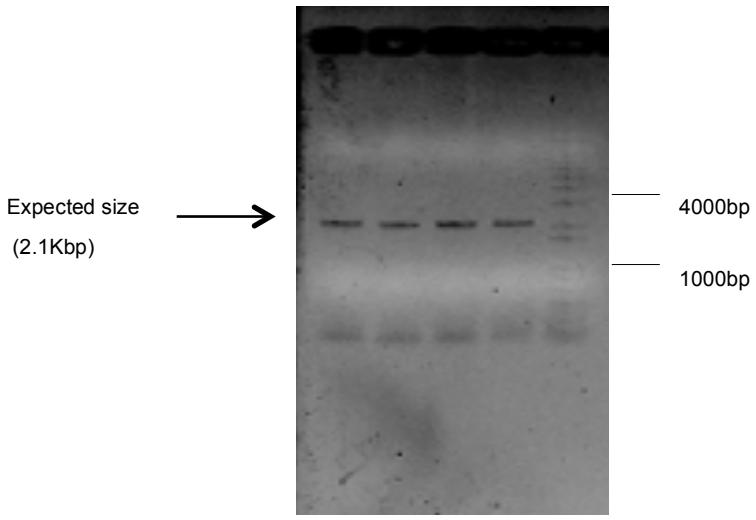

Figure 9 - Result of gel electrophoresis of the mixture obtained following PCR amplification of the *HAA1* gene. The used primers are listed in Table 2

After proceeding with digestion and ligation of the *HAA1* gene and the plasmid pPINK-HC(+) (as described in Materials and Methods section 3.2.2), transformation of *E.coli* XL1-Blue was performed and led to the obtention of 16 *E.coli* transformants that were selected in the presence of ampicillin, which is the vector selective marker for *E.coli.* 16 *E.coli* transformants were obtained from

all the ligation conditions tested. PCR from the obtain *E.coli* transformants plasmid DNA was performed to test if the *HAA1* insert was present in the *E.coli* transformants (as described in Materials and Methods section 3.2.3). The results were obtained after the electrophoresis of the PCR amplification mixture of all *E.coli* analyzed transformants (Figure 10) Primers for the PCR amplification are listed in Table 3.

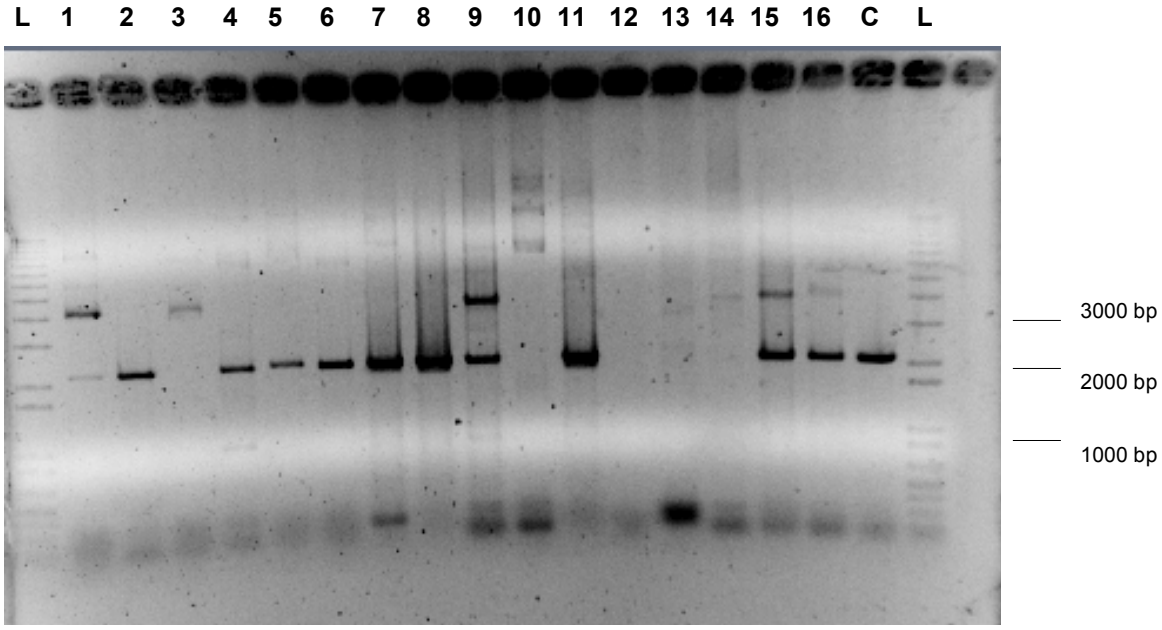

Figure 10 - Result of gel electrophoresis of the mixture obtained following PCR amplification of the *HAA1* gene, in which plasmid DNA of the obtained *E.coli* transformants was used as DNA template. *E.coli* candidate transformants were tested searching for the *HAA1* gene product. C) - *HAA1* gene amplified from *S.cerevisiae* strain S288C genomic DNA. L) - *Kbplus* (Invitrogen) ladder

The presence of the *HAA1* coding sequence within the construct was also confirmed in candidates 6 and 11 with subsequent analysis of the restriction profiles of the extracted plasmid DNA from these transformants were selected to proceed to enzymatic digestion, using *Hind*III (Figure 11). Results shown that the clones had the expected restriction profile after *Hind*III digestion. These results are in line with simulation using NEBcutter (New England Biolabs), in which the DNA sequence of the pPINK-HC(+)-*HAA1*$_{WT}$ construct was used as template to simulate enzymatic digestion using *Hind*III (Annex 1).

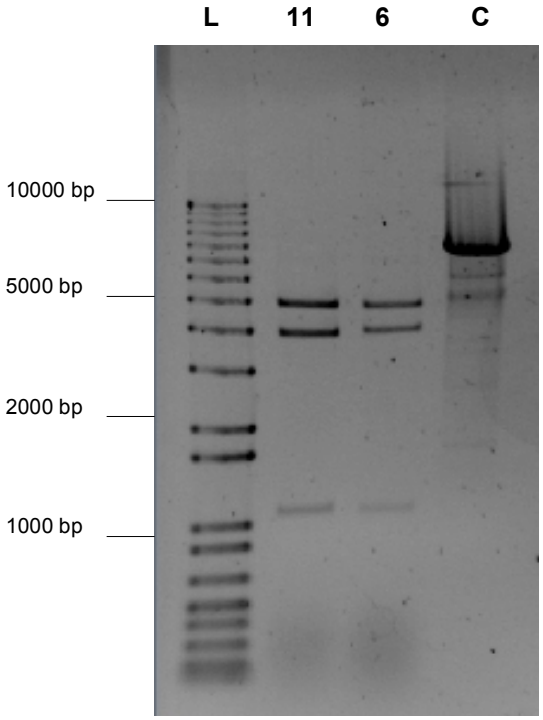

Figure 11 - *Hind*III digestion profile of transformants 6 and 11 plasmid DNA, using *Hind*III to confirm the presence of suitable construct pPINK-HC(+)-*HAA1*$_{WT}$ C) - Linearized empty vector pPINK-HC(+) digested with *Kpn*I. L) - *Kbplus* (Invitrogen) ladder

After the confirmation of the presence of the *HAA1* coding sequence within the construct following analysis of the restriction profiles of the plasmid DNA extracted from the transformants 6 and 11, the transformant 11 was selected for DNA sequencing and within *HAA1* coding sequence a nucleotide Guanine (G) in position 9 was missing, compromising the expression of the correct amino acid sequence. In order to correct $HAA1_{WT}$ coding sequence within pPinkα-HC(+)-$HAA1_{WT}$, site directed mutagenesis was carried out (according to the procedures in Materials and Methods section 3.2.4) and the amplified product was verified by gel electrophoresis (Figure 12).

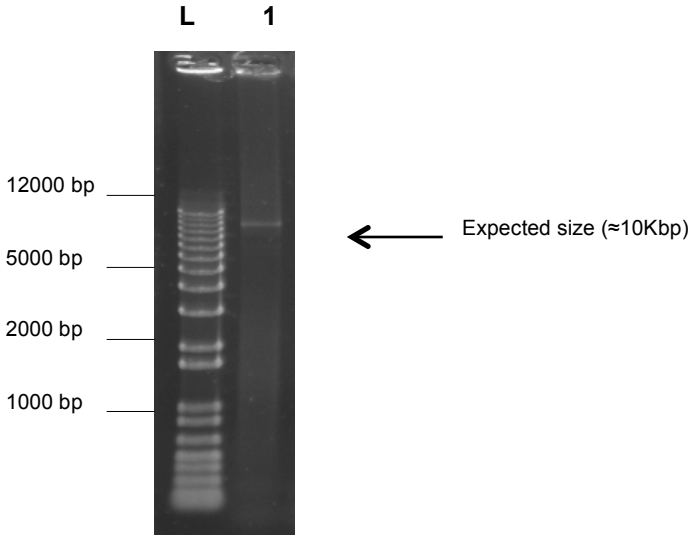

Figure 12 - Result of gel electrophoresis of the mixture obtained following site directed mutagenesis PCR.1) - Amplification product, using pPinkα-HC(+)-$HAA1_{WT}$ as DNA template, where the *HAA1* coding sequence miss a Guanine (G) nucleotide in position 9. L) - *Kbplus* (Invitrogen) ladder

After the correction of pPinkα-HC(+)-*HAA1*$_{WT}$ construct by site directed mutagenesis and sequencing of the *HAA1* coding sequence within pPinkα-HC(+)-*HAA1*$_{WT}$ construct , *P.pastoris* cells were electroporated in order to integrate pPinkα-HC(+)-*HAA1*$_{WT}$ into the *P.pastoris* genome. To allow integration of the pPinkα-HC(+)-*HAA1*$_{WT}$ into the *P.pastoris* genome, the construct was linearized following *Spe*I digestion that cuts in the *TRP2* region of the pPink vector. Homologous recombination of the linearized vector of the free DNA termini resulting in single-crossover integration can only occur into the *TRP2* genomic locus of *P.pastoris.* pPink™ vectors contain *ADE2* gene as a selection marker for *P.pastoris.* The integrant candidates were selected based on adenine auxotrophy complementation phenotype. Genomic DNA from Ade$^+$ *P.pastoris* cells was extracted and analysed by PCR (according to Materials and Methods section 3.2.5), where *HAA1* amplification was confirmed by gel electrophoresis (Figure 13)

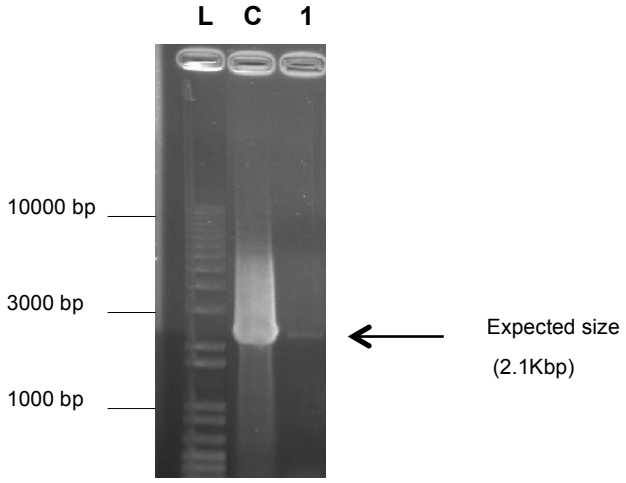

Figure 13 - Result of gel electrophoresis of the mixture obtained following PCR amplification of the *HAA1* coding sequence 1) - *HAA1* coding sequence amplification using Ade$^+$ *P.pastoris* integrant's genomic DNA as DNA template. C) - *HAA1* coding sequence amplification using pPinkα-HC(+)-*HAA1*$_{WT}$ from *E.coli* extracted plasmid DNA L) - *Kbplus* (Invitrogen) ladder

Moreover, the expression of *HAA1* $_{WT}$ with *P.pastoris* integrant was carried out. Since the production of recombinant proteins in *P.pastoris* is growth-related, the culture was grown up to very high cell densities in order to obtain large quantities of the desired protein [106]. Glycerol was selected as carbon source, recommended by the manufacturers' instructions of the commercial PichiaPink™ Expression system (Invitrogen). [61] During growth phase in which the culture reaches high cell densities *AOX1* promoter is strongly repressed by glycerol. Upon depletion of this carbon source with subsequent addition of methanol, *AOX1* methanol inducible promoter is fully induced, allowing expression of the desired recombinant protein.

To assess *HAA1* expression levels, a time course experiment was performed to study the induction profile over time, at two different temperatures: 30°C because it is the usually optimal defined growth temperature for *P.pastoris*, and 25°C, since it is reported that a reduction of the

temperature can significantly improve recombinant protein productivity without hampering growth, due to the decrease of proteases activity and a reduced death cell rate. [71] The expression levels were assessed by Western blot, using an anti-His_6 tag specific antibody to confirm the presence of Haa1 within the induced-time course samples, either in cell free extracts or in the supernatant, since it was expected that the recombinant protein would be secreted due to the presence of *S.cerevisiae* α mating sequence linked to the recombinant protein. The results are shown in figure 14 and 15.

Concerning the supernatant analysis during cultivation, It was expected to see a clear signal detected by immunoassay corresponding to Haa1 protein with a molecular with approximately 80 KDa, the predicted molecular weight of the His-labelled Haa1 protein with the fused TEV cleavage sequence, assuming that the *S.cerevisiae* α mating sequence would be cleaved transversing secretion pathway. Unfortunately, no bands with approximately 80 KDa, corresponding to the estimated size of the His-labelled Haa1 protein fused with the TEV cleavage sequence were detected in the supernatant samples obtained during the cultivation, neither at 25°C nor at 30°C induction temperatures (lanes 3-10 in Figure 14; lanes 3-7 in figure 15).

Regarding the cell free extracts analysed during cultivation, contrary to our expectations, the results show two specific bands that were detected by immunoassay with approximately 70 KDa and approximately 100KDa. The results also show unspecific multiple bands with different molecular weights, apparently. The band detected by immunoassay with approximately 100KDa may presumably corresponds to the His-labelled Haa1 protein with the fused TEV cleavage sequence and with the *S.cerevisiae* α mating sequence that remained retained intracellularly, in which the α mating sequence wasn't properly processed and cleaved. The reason for the presence of a band detected by immunoassay with approximately 70 KDa remains unclear. However, it is likely that this may correspond to proteolytic degradation of the recombinant protein.

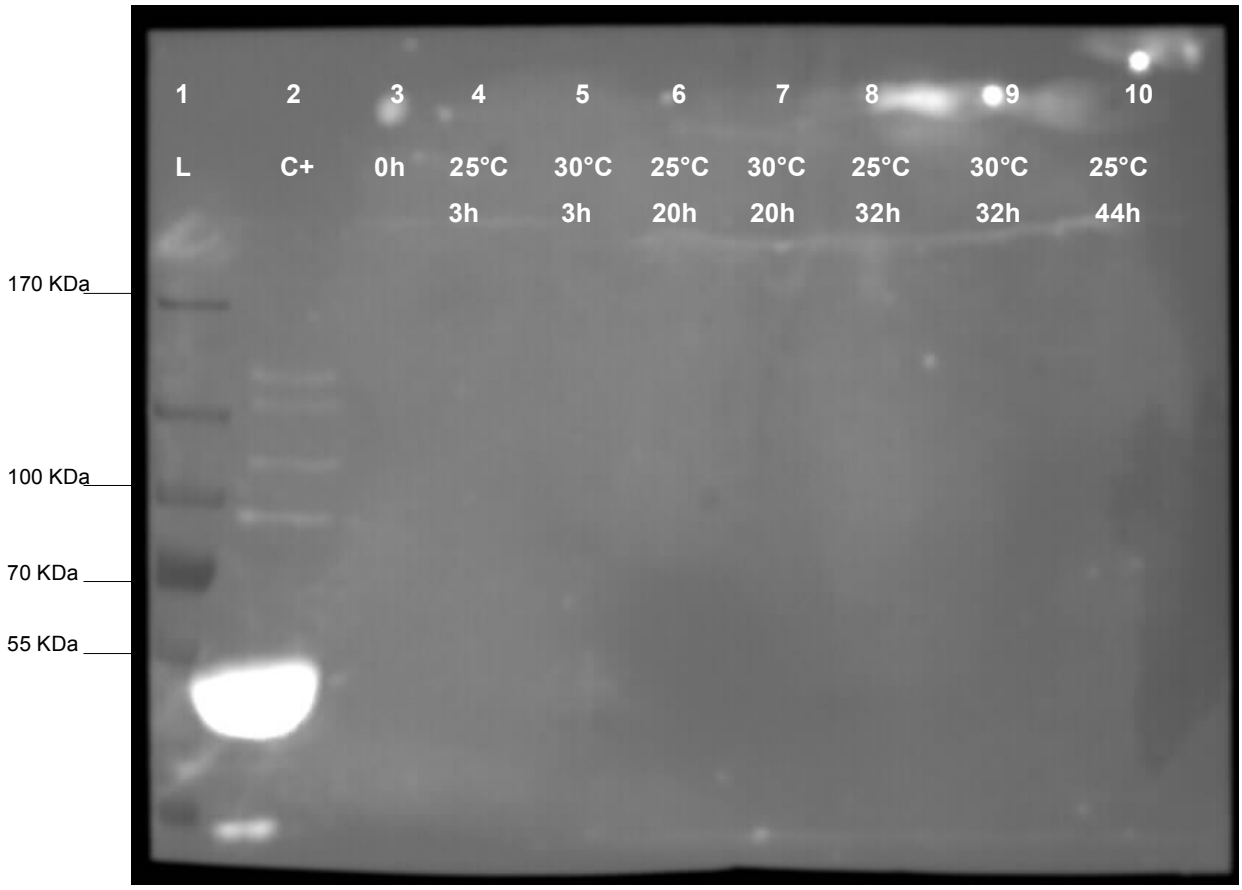

Figure 14 - Western blot analysis of the His-tag Haa1 overexpression in *P.pastoris*, using *AOX1* methanol inducible promoter. Haa1 was detected using anti-His_6 tag specific antibody. . Lane 3-10 - *P.pastoris* culture supernatants, collected at different time points during cultivation after methanol induction, at either 25°C or 30°C. Lane 2 – Recombinant BceC protein containing a C-terminal $His6_6$ tag used as a positive control. . Lane 1 - PageRuler®plus PreStained (Thermo Fisher Scientific) Protein Ladder

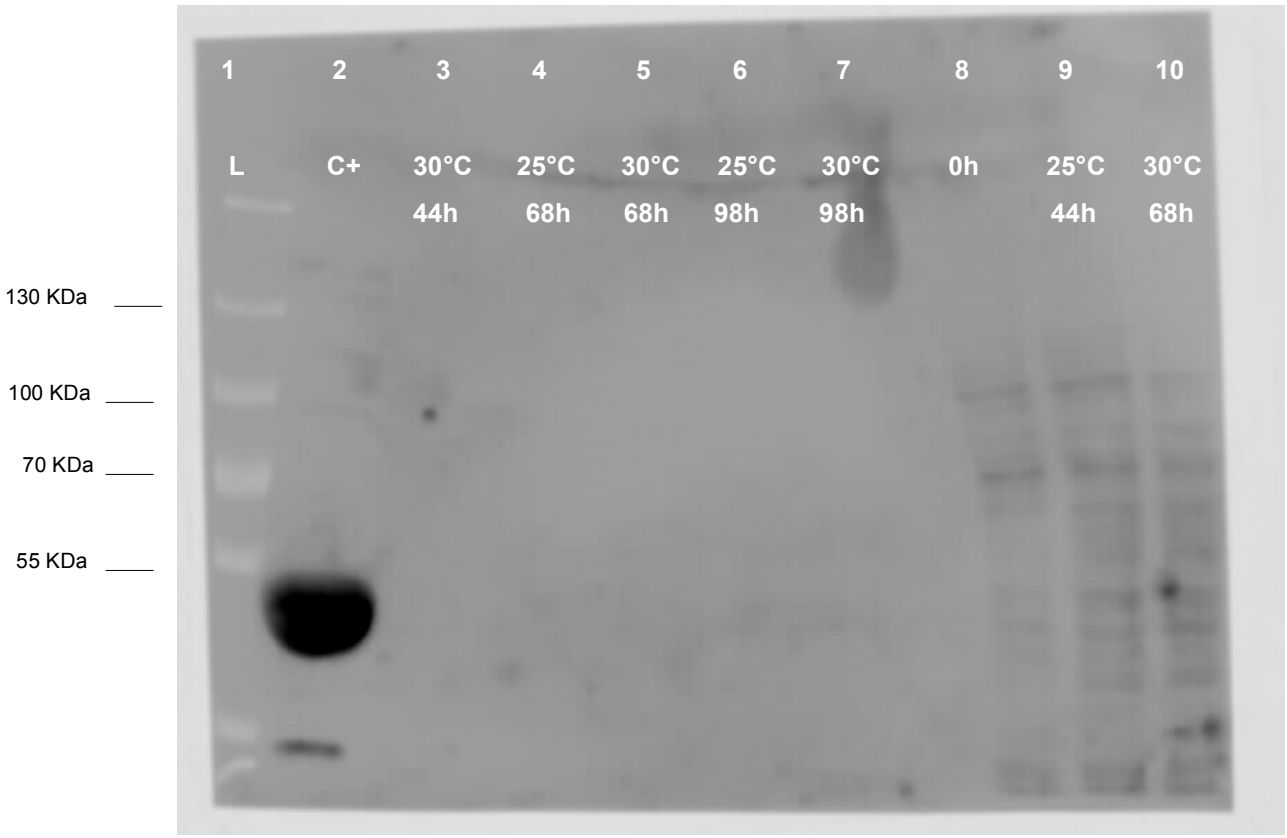

Figure 15 - Western blot analysis of the His-tag Haa1 overexpression in *P.pastoris*, using *AOX1* methanol inducible promoter. Haa1 was detected using anti-His_6 tag specific antibody. . Lane 3-7 - *P.pastoris* culture supernatants, collected at different time points during cultivation after methanol induction, at either 25°C or 30°C. Lane 8-10 - *P.pastoris* cell free extracts collected at different time points during cultivation after methanol induction, at either 25°C or 30°C. Lane 2 – Recombinant BceC protein containing a C-terminal $His6_6$ tag used as a positive control. . Lane 1 - PageRuler®plus PreStained (Thermo Fisher Scientific) Protein Ladder

The literature reports that a decrease in pH and cultivation temperature improves protein secretion using *P.pastoris* as host for protein overproduction, as the result of the decrease protease activity and by increase of protein stability. [71] Taking this in consideration, an additional experiment was carried out, using the growth medium containing glycerol and the inducer methanol at a pH of 4, and temperatures 22°C or 25 °C. The different samples obtained during cultivation time-course were analysed by Western blot, but no bands with approximately 80 KDa, corresponding to the estimated size of the His-labelled Haa1 protein fused with the TEV cleavage sequence. (data not shown) Concerning the cell free extracts analysed during cultivation time-course, the results obtained were equal compared to the results obtained from the samples during cultivation time course at a pH of 6, either at 25°C or 30°C (data not shown).

A specific band with approximately 100KDa presumably corresponding to the His-labelled Haa1 protein with the fused TEV cleavage sequence and with the *S.cerevisiae* α mating sequence was been detected by Western blot in the cell free extracts. However, one of the major goals of overproducing the Haa1 protein using PichiaPink™ Expression system (Invitrogen) was to secrete the

protein properly folded, pos-translationally processed and functional recombinant proteins in the supernatant. If successful, this strategy would be convenient for recombinant protein recovery and purification, given that the vast majority of the total protein in the culture supernatant is the secreted heterologous protein. Since the attempt of overproduce and secrete to the supernatant the his-tagged Haa1 protein using *P.pastoris* was not achieved, overproduction of the Haa1 using *E.coli* as host cell was performed.

4.2 Overexpression in *E.coli* and purification of $Haa1_{WT}$ and the derived mutant $Haa1_{S135F}$

The overproduction of the recombinant Haa1 and $Haa1_{S135F}$ proteins was carried out using *E.coli* as host cell, using pET expression system (Novagen). The genes to be expressed from the recombinant plasmids have a fused C-terminal 6×His-tag, allowing the desired protein purification by Ion Metal Affinity Chromatography (IMAC). For this strategy, both plasmids pET23a(+)-$HAA1_{WT}$-His6 and pET23a(+)-$HAA1_{S135F}$-His6 harbouring $HAA1_{WT}$ coding sequence and the derived mutant $HAA1_{S135}$ coding sequence ,respectively, were already available in lab, and the expression conditions were already been optimized (Mira *et al.,*2008). The desired genes were overexpressed in *E.coli* by addition of Isopropyl β-D-1-ThioGalactoPyranoside (IPTG) to the bacterial cell culture.

To assess the purity of the eluted fractions of both Haa1 recombinant protein and its derived mutant eluted with different imidazole concentrations, a SDS-PAGE polyacrylamide gel followed by protein staining was carried out. Additionally, a Western blot using an anti-His_6 tag specific antibody was performed to confirm the expression of the desired protein. A representative SDS-PAGE polyacrylamide gel of the 1ml eluted fractions and of the respective Western blot with increasing concentrations of imidazole of both Haa1 and its derived mutant are shown in figures 16 and 17, respectively.

Regarding Haa1 purification by IMAC, from the analysis of the elution profile of the different fractions it is possible to see that, although some protein was eluted at imidazole concentrations of 120mM (lane 5 figure 16 A) and B)), most of the recombinant protein started to be eluted at concentrations of 130mM imidazole and above (lane 6-17 figure 16 A) and B). Furthermore, co-purification of *E.coli* native contaminant proteins exhibiting affinity for Ni^{2+} is evident, particularly with imidazole elution concentrations between 120 mM and 150mM (Figure 16 A and B). A similar methodology was used to purify the derived Haa1 mutant by IMAC. In general, the elution profile of the derived Haa1 mutant is similar to the elution profile of $Haa1_{WT}$. Most of the recombinant protein was eluted at a concentration of, approximately, 135mM imidazole and above (lane 5 figure 17 A) and B)). As seen for his-tagged $Haa1_{WT}$ purification, co-purification of *E.coli* native contaminant proteins is evident also in the derived mutant purification. (Figure 17 A) and B)). Also, in the elution profile of the derived Haa1 mutant, other bands with lower molecular weights that the one expected for Haa1 protein (with 78KDa) were detected by the anti-His antibody in Western blot analysis on the fractions eluted with 108-135mM imidazole (lane 4 and 5 figure 17 B)), suggesting the possible proteolytic degradation of Haa1 derived mutant in these fractions. To avoid the more contaminated fractions to assure higher purity, only the

Haa1 protein fractions eluted with a imidazole concentration of approximately 160mM and above (lane 9-17 figure 16 A) and B), were gathered and concentrated for polishing step by performing Fast Protein Liquid Chromatography (FPLC). The same was applied to the derived mutant (lane 8 figure 17 A) and B))

Previous results obtained in our lab suggested that Haa1 behaves as a dimer in solution and its elution volume was established using the column, used for the FPLC polishing step. For that reason, the purified fractions of Haa1 (78KDa) were collected at an apparent elution volume equivalent to a 156 KDa molecule. However, by comparing the protein extracts of Haa1 protein before and after the FPLC polishing step, no major improvements in purification were found since protein contaminants with a lower molecular weight than the one expected for Haa1 protein (78KDa) were still visible after this step (Figure 18). Also, the amount of the recovered protein after the FPLC polishing step was lower (Figure 18). A similar methodology was used for the derived Haa1 mutant by performing a FPLC polishing step with similar results (data not shown).

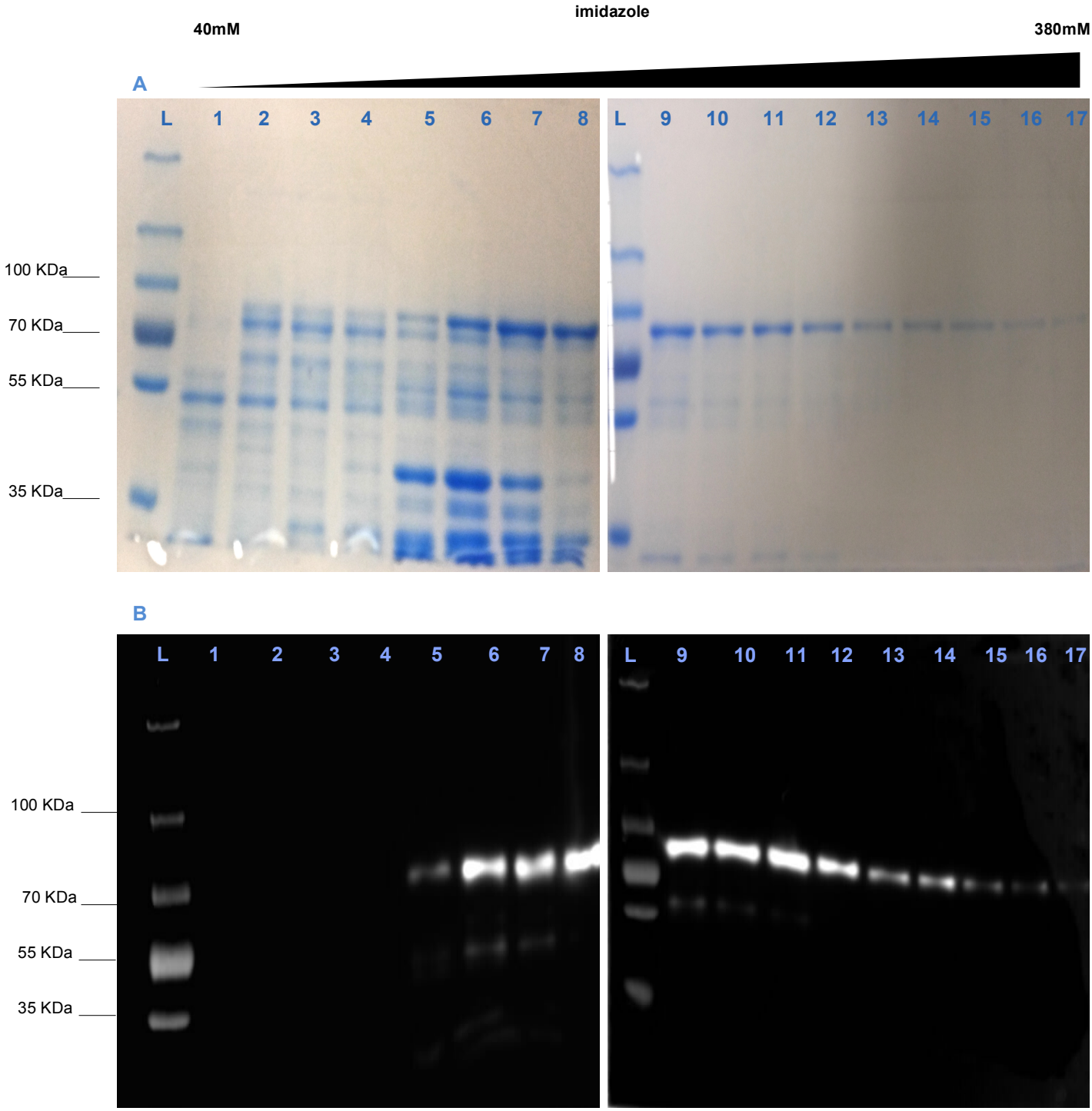

Figure 16 - Analysis of the elution profile of the recombinant his-tagged Haa1 protein, purified by Ni^{2+} -based Ion Metal Affinity Chromatography. A) SDS-PAGE analysis of the elution profile with an increased gradient of imidazole. Gel was stained with GelCode™ Blue Safe Stain (Thermo Fisher Scientific) B) Western blot analysis of the elution profile with an increased gradient of imidazole, using Anti-His_6 tag specific antibody to confirm the presence of Haa1 (with a molecular weight of 78 kDa). The fractions collected at imidazole concentrations above 160mM (lane 9 and above) were gathered and concentrated. L) - PageRuler®plus PreStained Protein Ladder (ThermoFisher Scientific) used as ladder.

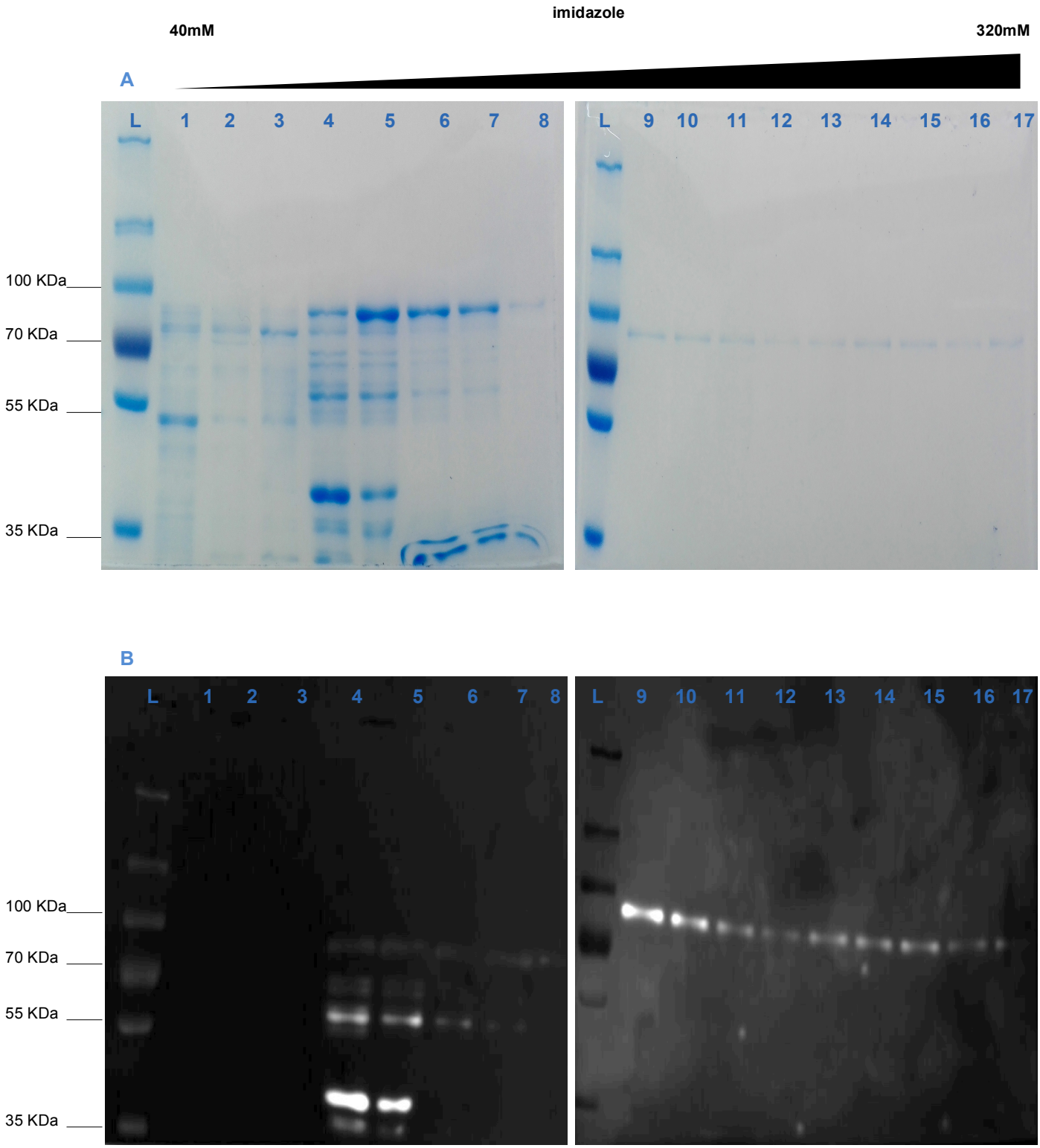

Figure 17 - Analysis of the elution profile of the recombinant his-tagged $Haa1_{S135F}$ protein, purified by Ni^{2+} -based Ion Metal Affinity Chromatography. A) SDS-PAGE analysis of the elution profile with an increased gradient of imidazole. Gel was stained with GelCode™ Blue Safe Stain (Thermo Fisher Scientific) B) Western blot analysis of the elution profile with an increased gradient of imidazole, using Anti-His_6 tag specific antibody to confirm the presence of Haa1 derived mutant (with a molecular weight of 78 kDa). The fractions collected at imidazole concentrations above 190mM (lane 8 and above) were gathered and concentrated. L) PageRuler®plus PreStained Protein Ladder (Thermo Fisher Scientific) used as ladder

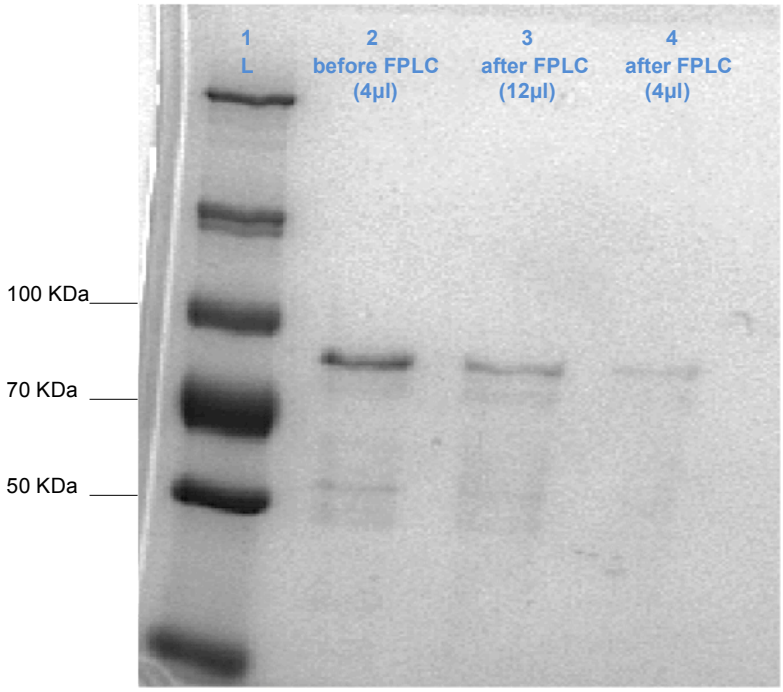

Figure 18 - An example of the various SDS-PAGE gels for separation of Haa1 protein fractions purified by Ion Metal Affinity Chromatography. Lane 2 - Purified Haa1 solution without FPLC polishing step. Lane 3 and 4 – Different volumes of the purified Haa1 solution after FPLC polishing step. L) - PageRuler®plus PreStained Protein Ladder (Thermo Fisher Scientific) used as ladder.

4.3 Atomic force microscopy : visualization of $Haa1_{WT}$ and derived mutated $Haa1_{S135F}$ bound to *TPO3* promoter DNA region

Before imaging DNA-Haa1 protein complexes, two control experiments were performed using only DNA or protein (as described in Materials and Methods section 3.4.2). The DNA control experiment is important to know if the DNA fragment can be properly adsorbed to the mica and to measure its length. The protein control experiment was performed to estimate free protein height profile. The*TPO3* promoter region used has 1000bp, upstream to the initiation ATG codon. Since 1bp corresponds to approximately 0.3nm, *TPO3* promoter DNA region has, as expected, approximately 300nm of length (Figure 19). The measured heights of the free protein performed in the protein control experiment varied between 0.3nm and 1.1nm (Figure 20). This variation might be due to the electrostatic interactions between the proteins and the tip and the different orientation of the proteins, since the proteins can adhere to the mica surface at different sites, and therefore, different parts contribute to the different protein height measurements.

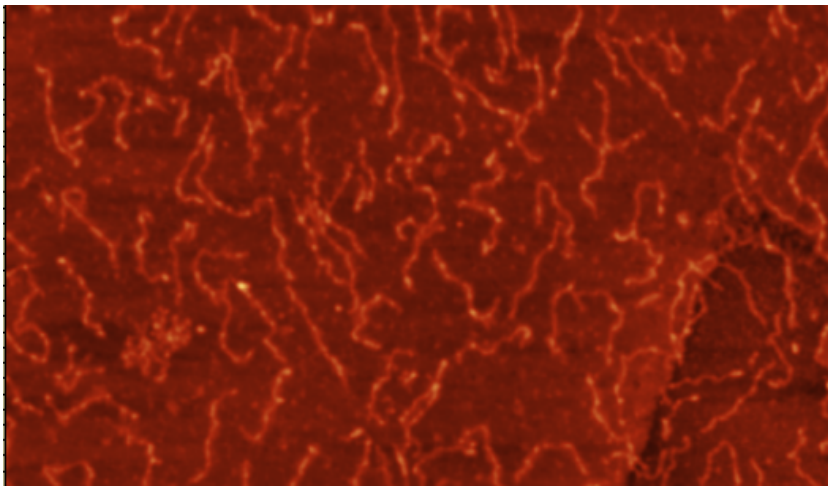

Figure 19 - Atomic Force Microscopy Imaging. AFM imaging of *TPO3* promoter region (1nM DNA solution). The scan size corresponds to 5μm^2

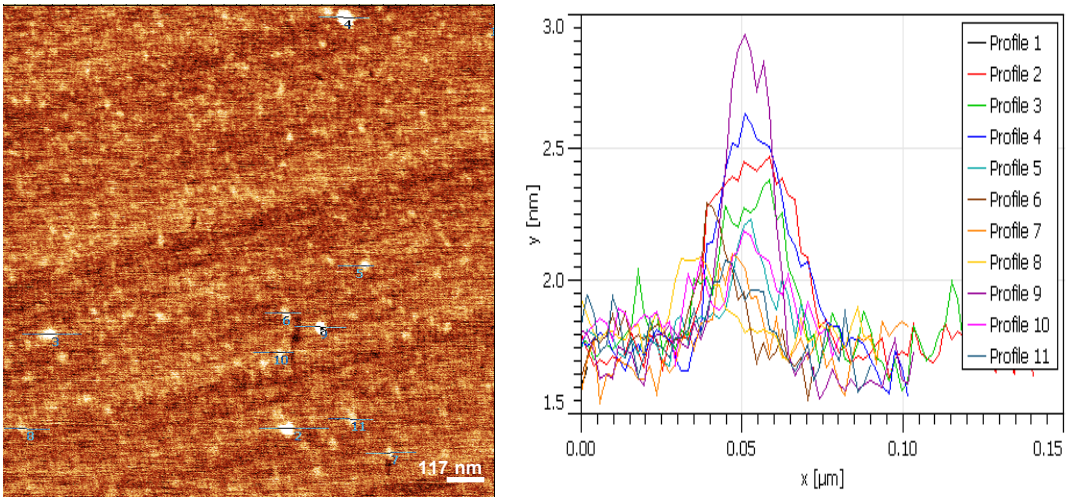

Figure 20 - Atomic Force Microscopy Imaging and processing for analysis using Gwyddion software 2.45. A) AFM imaging of the Haa1 transcription factor at 10nM. The image shows proper surface coverage for height profile analysis. Proteins selected for height measurement are spotted with a blue dot. The scan size corresponds to 1μm^2. B) Height profile data of the Haa1 transcription factor. Flattening to remove mica height background was not applied in this data, corresponding approximately to 1.9nm. Protein height varies between 0.3nm and 1.1nm, approximately

Regarding DNA-protein complexes, different ratios of DNA:protein were used, containing the *TPO3* promoter region and Haa1 protein or the derived mutant in a ratio of 1:1 and 1:5 to ensure reproducibility and to see if the formation of complexes was more favourable by increasing concentrations of protein, for the same concentration of DNA. However, this was not possible to confirm because the number of suitable individual measurements was scarce. Clearer results of protein:DNA complexes were obtained at a 1:5 molar coupling stoichiometry and, for that reason, only those are shown in this thesis.

Protein:DNA complexes were selected based on a end-to-end distance of the DNA molecule with approximately 300nm of length expected for *TPO3* promoter region and the location of Haa1 bound to *TPO3* promoter region, approximately at 1/3 of the DNA chain, concomitant with the location of Haa1-responsive element between nucleotides -690 and -590 of the *TPO3* promoter.[13]

AFM images indicating the formation of $Haa1_{WT}$:DNA complexes is shown in figure 21 A) and B). In figure 21 A) the length of the DNA fragment has approximately 300nm and in figure 21 B) the DNA fragment has approximately 400nm. A globular shape presumably corresponding to the Haa1 protein is found attached to the DNA in both cases, at 1/3 of the DNA chain, at 100nm position approximately. The measured height of the attached protein to the DNA chain is approximately between 2.2nm and 3nm (Figure 21 A) and B)).

Similar images indicating the formation of complexes between $Haa1_{S135F}$ and *TPO3* promoter region by AFM imaging are also shown in Figure 21 C) and D). In both cases the length of the DNA fragment corresponding to *TPO3* promoter region is approximately 300nm. A globular shape corresponding to the Haa1 derived mutant is found in both images. In figure 21 C), the protein seems to be attached at 1/3 of the DNA chain, at 100nm position approximately. However, in figure 21 D) the protein seems to be attached at 1/2 of the DNA chain approximately. Also, uneven measured height variations in some regions of the DNA:protein complex are shown in figure 21 D), possibly due to the presence of holes in the mica surface near and/or above the protein:DNA complex, corresponding to black spots visible in the image. Moreover, by comparing with the $Haa1_{WT}$:DNA complexes, height measurements of the attached protein that presumably corresponds to $Haa1_{S135F}$ are lower, being approximately between 0.8nm and 1.1nm (Figure 21 C) and D)).

.

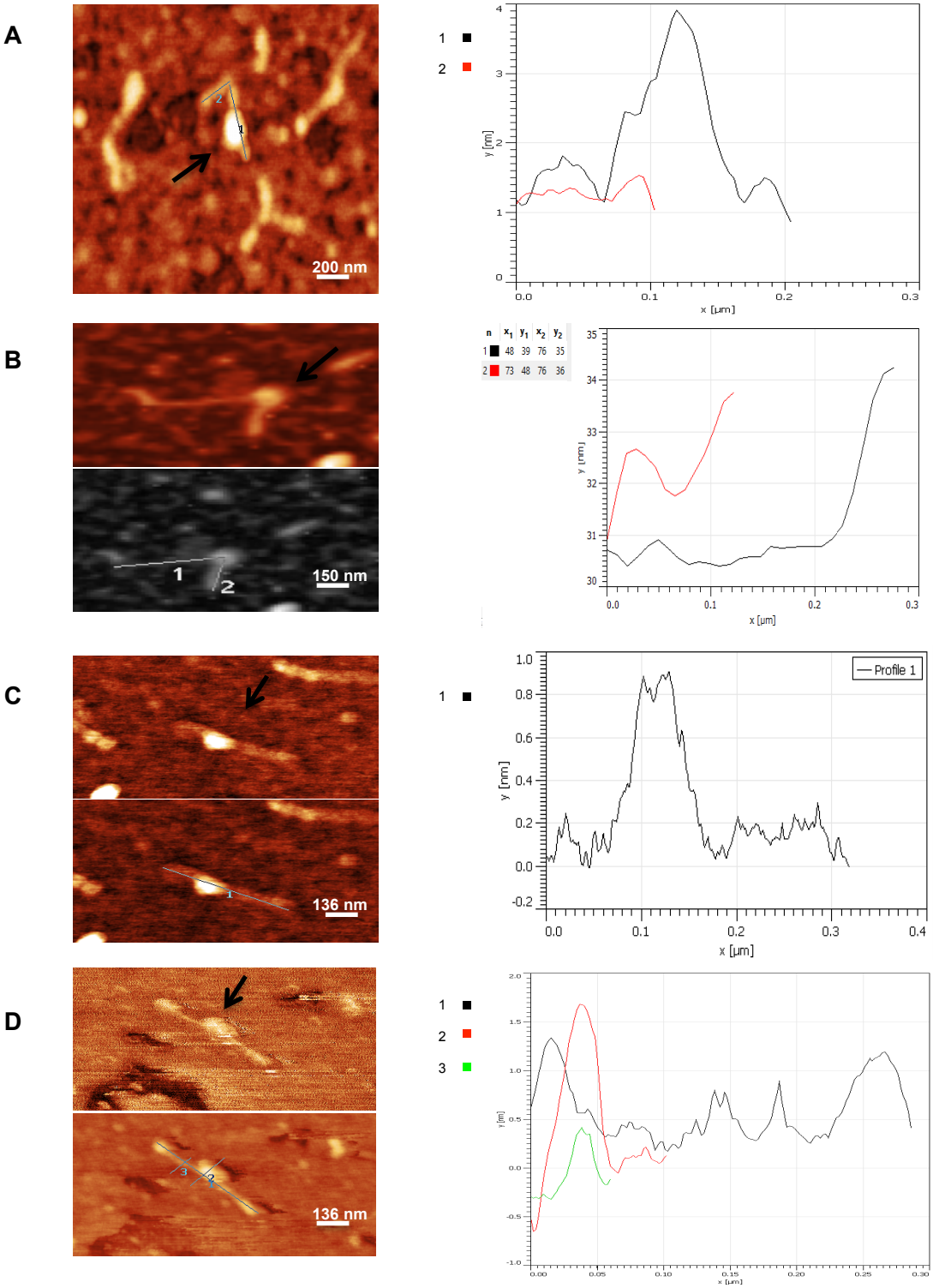

Figure 21 - Atomic Force Microscopy Imaging and processing for analysis using Gwyddion software 2.45. Protein:DNA complexes were selected based on the end-to-end distance of the DNA molecule with approximately 300nm of length expected for *TPO3* promoter region and the location of Haa1 bound to *TPO3* promoter region, approximately at 1/3 of the DNA chain, concomitant with the location of Haa1-responsive element. A and B) - AFM image of $Haa1_{WT}$:DNA complex at a 1:5 molar coupling stoichiometry and the respective measurement based on a end-to-end distance of the DNA, on the right. The putative $Haa1_{WT}$ bound to DNA has 2.2nm of height in A and 3nm of height in B. The scan size corresponds to 2µm^2 in A and 5 µm^2 in B. C and D) - AFM image of $Haa1_{S135F}$:DNA complex at a 1:5 molar coupling stoichiometry and the respective measurement based on the end-to-end distance of the DNA, on the right. The putative $Haa1_{S135F}$ bound to DNA has 0.8nm of height in C and 1.1nm of height in D. The scan size corresponds to 1µm^2. The protein:DNA complexes are indicated with a black arrow.

5 Discussion

The aim of the present work was to overproduce and purify the *S.cerevisiae* Haa1 transcription factor and S135F mutated form that is known to confer increased tolerance to acetic acid. The final objective to obtain a suitable amount of the protein for further functional and structural studies of this transcription factor involved in the adaptive response of yeast to acetic acid. To fulfil this objective, the first experimental strategy was to overexpress the *HAA1* gene using the commercial PichiaPink™ Expression system using *Pichia pastoris* as cell host.

The attempt to overproducing and secrete the Haa1 transcription factor was not achieved It was expected that the recombinant Haa1 protein would be secreted and detected in the culture supernatant, as a 80KDa protein detected by Western blot, corresponding to the His-labelled Haa1 protein with the fused TEV cleavage sequence, this assuming that the fused *S.cerevisiae* α mating sequence would be cleaved through the secretion pathway. However, results indicate that the recombinant Haa1 transcription factor was produced and retained intracellularly and that a fraction of the retained recombinant protein was possibly degradated. Several studies suggest that one of the most critical bottlenecks that influence the productivity of an expression system is the recombinant protein translocation and secretion out of the cell. [67] [68] [69] [70] [71] The secretion limiting step of the recombinant Haa1 protein may be associated with the processing of the secretion peptide signal (the *S.cerevisiae* α mating sequence fused to the Haa1 protein, used in this work), since processing at the α mating sequence *Kex*2 site is often incomplete or it does not occur, leaving a N-terminus with various portions of the α mating sequence still attached and compromising secretion out of the cell. [72] [73] [74] [75] [76] Additionally, the N-terminus of the recombinant Haa1 protein may fold on a structure that prevents processing enzymes from reaching their processing site. [55] Moreover, the fraction of the retained recombinant Haa1 protein that seems degradated may be related to the activation of Endoplasmic Reticulum Associated Protein Degradation (ERAD) pathway, since it was reported that overproduction of recombinant proteins leads to endoplasmic reticulum (ER) overloading and consequent accumulation of misfolded or unfolded proteins, triggering the Unfolded Protein Response (UPR) pathway. [79] [80] [81] [82] As consequence, the UPR pathway triggers the ERAD pathway, in which unfolded or misfolded proteins are subjected to proteosomal degradation. [79] [80] [81] [82] Additionally, the recombinant Haa1 may impose a significant metabolic burden during the overproduction of the recombinant Haa1 protein. Since Haa1 is a yeast transcription factor and *P.pastoris* also have a *HAA1* homologue gene, this may also contribute to the decreased expression levels observed in this study. To overcome these problems in protein overproduction some strategies could be applied. It is reported that a co-feeding of glycerol and methanol over a certain time, whereby the glycerol feeding is decreased slowly and methanol feeding is increased, can be beneficial for *P.pastoris* cells to adapt better for protein overproduction. [107] [108] Also, the Haa1 transcription factor is not secreted in their native host. For this reason it could be beneficial to overproduce the recombinant Haa1 intracellularly, without the secretion leader signal peptide, to mimic a more natural environment similar to its production in their native host.

Since the attempt to overproduce extracellularly the $Haa1_{WT}$ transcription factor in *P.pastoris* was unsuccessful, the overexpression of Haa1 and the derived mutant harbouring the S135F mutation was carried out using *E.coli* as host cell using pET expression system. Both recombinant Haa1 and the derived mutant his-tagged proteins were successfully overexpressed. However, the amount of recovered protein was low, and no major improvements in purification were found using Fast Protein Liquid Chromatography (FPLC) polishing step after performing Ion Metal Affinity Chromatography (IMAC). In order to obtain more pure protein fractions and to assure higher amounts of recovered protein, polishing step must be optimized after IMAC purification. For instance, the use of heparin chromatography could be a suitable strategy. The structure and negative charge of heparin enable it to mimic DNA in its overall binding properties and it can confer higher selectivity in purification of the transcription factor Haa1 [109] It is reported that heparin is able to bind to a range of biomolecules, including transcription factors. [109] [110]

Remarkably, the very preliminary experiments carried out using Atomic Force Microscopy (AFM) allowed the visualization of the complexes structures, in which Haa1 transcription factor is apparently bound to the *TPO3* promoter region (1000bp upstream the *TPO3* start codon), with approximately 300nm of length. Apparently, similar complex structures were also visualized between the Haa1 derived mutant and *TPO3* promoter region. Interestingly, by focusing on the position distribution of the protein on the DNA sequence, it is possible to hypothesise a specific recognition site for this protein in the *TPO3* promoter region. In fact, the protein:DNA complexes observed mostly appears to be located at 1/3 of the DNA chain, consistent with the location of the Haa1-responsive element (HRE) between nucleotides -690 and -590 of *TPO3* promoter.[13]

Another interesting indication was obtained by comparing of the results from height measurements between the putative $Haa1_{WT}$:DNA complexes and the putative $Haa1_{S135F}$:DNA complexes, since the differences in height protein preliminary measurements between the Haa1 and the derived mutant S135F bound to DNA appear to suggest a conformational change upon binding. Proteins and DNA can undergo conformational changes in order to form functional complexes to carry out the required function. [111] [112] [113] Co-relations, using AFM approach, between conformational changes and the average protein height measurement of bound and unbound forms, were already been reported.[114] The Haa1 derived mutant S135F harbours a single point mutation by substitution of a serine for a phenylalanine in 404 nucleotide position, near the DNA binding domain, and leading to an increase in acetic acid tolerance. Single point mutations, such as the S135F mutation in the Haa1 protein, may be crucial for local conformational changes that may impact protein function. [115] Hydrophobic and aromatic residues were found to be critical for trans-activation domain target contacts, which are essential for efficient transactivation activity. [116] [117] [118] [119] [120] [121] Moreover, the S135F substitution may be also affecting the trans activation potential of Haa1, for instance by improving the recruitment of RNA transcription machinery. [50] Additionally, nonpolar residues with an aromatic nature such as phenylalanine can contribute to protein function by forming hydrophobic cores that are relevant for global protein, geometrical compatibility and efficient packing, folding and stability. [112] [122] [123] [124] Given that the Haa1 protein is a multi-phosphorylated protein (Sugiyama *et al,. 2014,*

phosphoGRID), the elimination of a serine residue that can be prone to phosphorylation and its impact during the binding and/or in the stability of Haa1 protein:DNA complexes remains to be determined. Also, additional experiments to increase the number of Haa1 protein:DNA complexes found must be carried out in order to ensure reproducibility of the results obtained from AFM.

6 Bibliographic References

1. Hohmann, S Mager, W, H. *Topics in Current Genetics Yeast Stress Responses*. (Springer Berlin Heidelberg, 2003).
2. Piper, P., Calderon, C. O. & Hatzixanthis, K. Weak acid adaptation : the stress response that confers yeasts with resistance to organic acid food preservatives. *Microbiology* 2635–2642 (2001).
3. Ruis, H. & Schüller, C. Stress signaling in yeast. *BioEssays* 17, (1995).
4. Mitchel, R. & Morrison, D. Heat-schock induction of ionizing radiation resistance in *Saccharomyces cerevisae*, and correlation with stationart growth phase. *Radiat Res* (1982).
5. Blomberg, A., Larsson, C. & Gustafsson, L. Microcalorimetric monitoring of growth of *Sacharomyces cerevisae*: osmotolerance in relation to physiological state. *J Bacteriol* (1988).
6. Wieser, R. *et al.* Heat shock factor-independent heat control of transcription of the CTT1 gene enconding the cytosolic catalase T of *Saccharomyces cerevisae*. *J Bacteriol* (1991).
7. Lewis, J., Learmonth, R. & Watson, K. No Induction of heat, freezing and salt tolerance by heat and salt shock in *Saccharomyces cerevisae*. *Microbiology* (1995).
8. Flattery-O'Brien, J., Collinson, L. & Dawes, I. *Sacharomyces cerevisae* has an inductible response to menadione which differs from that to hydrogen peroxide. *J Gen Microbiol* (1993).
9. Gasch, A. P. *et al.* Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes □. 11, 4241–4257 (2000).
10. Takagi, H., Kitagaki, H. & Brewing, I. *Stress Biology of Yeasts and Fungi*. (Springer Berlin Heidelberg, 2015).
11. Stratford, M. *et al.* International Journal of Food Microbiology Extreme resistance to weak-acid preservatives in the spoilage yeast *Zygosaccharomyces bailii* ☆. *Int. J. Food Microbiol.* 166, 126–134 (2013).
12. Mollapour, M. & Piper, P. W. Weak Organic Acid Resistance of Spoilage Yeasts. *Elsevier Ltd* 143–155 (2008).
13. Henriques, F. *et al.* Identification of a DNA-binding site for the transcription factor Haa1 , required for *Saccharomyces cerevisiae* response to acetic acid stress. 39, 6896–6907 (2011).
14. Davenport, R. . Forensic microbiology for soft drinks business. *Soft Drink. Manag. Int.* (1996).
15. Davenport, R. . Microbiology of soft drinks. *Chemestry Technol. Soft Drink. Fruit Juices, Sheff. Acad. Press* (1998).
16. Almeida, R. M., Modig, T. & Petersson, A. Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. *J. Chem. Technol. Biotechnol.* 349, 340–349 (2007).
17. Allen, I., Laskin, S. . & Geoffrey, M. *Advances in Applied Microbiology 77*. (Elsevier Inc., 2011).
18. Soares-silva, I., Casal, M. & Paiva, S. Transport of carboxylic acids in yeasts. *Blackwell Publ. Ltd.* 32, 974–994 (2008).
19. Fleet, G. Spoilage yeasts. *Crit. Rev. Biotechnol* 12, 1–44 (1992).
20. Ullah, A., Orij, R., Brul, S. & Smits, G. Quantitative analysis of the modes of growth inhibition by weak organic acids in Saccharomyces cerevisae. *Appl Env. Microbiol* 8377–8387 (2012).
21. Holyoak, C., Stratford, M., McMullin, Z., Cole, M. & Crimmins, K. Activity of the plasma membrane H(+)-ATPase and optimal glycolytic flux are required for rapid adaptation and growth of Saccharomyces cerevisiae in the presence of the weak-acid preservative sorbic acid. *Appl Env. Microbiol* (1996).
22. Viegas, C. . & Sa Correia, I. Activation of plasma membrane ATPase of *Saccharomyces cerevisiae* by octanoic acid. *J. Gen. Microbiol.* 645–651 (1991).

23. Piper, P. W. et al. Hsp30, the integral plasma membrane heat shock protein of *Saccharomyces cerevisiae*, is a stress-inducible regulator of plasma membrane ATPase. *Cell Stress Chaperones* (1997).

24. Guadalupe Cabral, M., Sa-Correia, I. & Viegas, C. Adaptative responses in yeast to the herbicide 2-methyl-4-chlorophenoxyacetic acid at the level of intracellular pH homeostasis. *J Appl Microbiol* 603–612 (2004).

25. Orij, R., Urbanus, M. & Vizeacoumar, F. Genome-wide analysis of intracellular pH reveals quantitative control of cell division rate by pH(c) in *Saccharomyces cerevisae*. *Genome Biol* (2012).

26. Pearce, A., Booth, I. & Brown, A. Genetic manipulation of 6-phosphofruto-1-kinase and fructose 2,6-bisphosphate levels affects theextent to which benzoic acid inhibits the growth of *Saccharomyces cerevisae*. *Microbiology* 403–410 (2001).

27. Dechant, R., Binda, M., Lee, S., Pelete, S. & Winderrickx, J. Cytosolic pH is a second messenger for glucose and regulates the pKa patyway through V-ATPase. *EMBO L* 2515–2526 (2010).

28. Young, B., Shin, J., Orij, R. & Chao, J. Phosphatidic acid is a pH biosensor that links membrane biogenisis to metabolism. *Science (80-.).* (2010).

29. Piper, P. W. Yeast superoxide dismutase mutants reveal a prooxidant action of weak organic acid food preservatives. *Free Radic. Biol. Med.* 27, 1219–1227 (1999).

30. Bracey, D. Comparison of the inhibitory effect of sorbic acid and amphotericin B on *Saccharomyces cerevisiae*: Is growth inhibition dependent on reduced intracellular pH? *J. Appl. Microbiol.* 85, 123–126 (1998).

31. Stratford, M. & Anslow, P. Evidence that sorbic acid does not inhibit yeast as a classic 'weak acid preservative'. *Lett. Appl. Microbiol* 142, 203–206 (1998).

32. Mollapour, M. & Piper, P. W. Hog1p MAP kinase determines acetic acid resistance in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 6, 1274–1280 (2006).

33. Bauer, B. ., Rossington, D., Mollapour, M., Kuchler, K. & Piper, P. W. Weak organic acid stress inhibits aromatic amino acid uptake by yeast, causing a strong influence of amino acid auxotrophies on the phenotypes of membrane transporter mutants. *Eur. J. Biochem.* 270, 3189–3195 (2003).

34. Mollapour, M. & Piper, P. Hog1 mitogen-activated protein kinase phosphorylation targets the yeast Fps1 aquaglyceroporin for endocytosis, thereby rendering cells resistant to acetic acid. *Mol Cell Biol* 27, 6446–6456 (2007).

35. Mira, N. P., Teixeira, M. C. & Sa, I. Adaptive Response and Tolerance to Weak Acids in Sac*charomyces cerevisiae* : A Genome-Wide View. *OMICS* 14, 525–540 (2010).

36. Makrantoni, V., Dennison, P., Stark, M. & Coote, P. a novel role for the yeast protein kinase Dbf2p in vacuolar H+-ATPase function and sorbic acid stress tolerance. *Microbiology* 12, 4016–4026 (2007).

37. Mira, N. P., Becker, J. D. & Sa, I. Genomic Expression Program Involving the Haa1p-Regulon in *Saccharomyces cerevisiae* Response to Acetic Acid. *Omi. A J. Integr. Biol.* 14, (2010).

38. Latchman, D. S. Transcription Factors: An Overview. *Int. J. Biochem. Cell Biol* 29, 1305–1312 (1997).

39. Keller, G., Ray, E., Brown, P. O. & Winge, D. R. Haa1 , a Protein Homologous to the Copper-regulated Transcription Factor Ace1 , Is a Novel Transcriptional Activator *. 276, 38697–38702 (2001).

40. Fernandes, A., Mira, N., Vargas, R., Canelhas, I. & Sa-Correia, I. *Saccharomyces cerevisiae* adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes. *Biochem Biophys Res Commun* 337, 95–103 (2005).

41. Mira, N. P., Palma, M., Guerreiro, J. F. & Sá-correia, I. Genome-wide identification of *Saccharomyces cerevisiae* genes required for tolerance to acetic acid. *Microb. Cell Fact.* 9, 1–13 (2010).

42. Mira, N. P., Cacho Teixeira, M. & Sá-Correia, I. Adaptive Response and Tolerance to Weak Acids.pdf. *Omics* 14, (2010).

43. Yeheskely-Hayon, D. *et al.* The Roles of the Catalytic and Noncatalytic Activities of Rpd3L and Rpd3S in the Regulation of Gene Transcription in Yeast. *PLoS One* 8, (2013).

44. Simões, T., Mira, N., Fernandes, A. & Sá-correia, I. The SPI1 gene, encoding a glycosylphosphatidylinositol-anchored cell wall protein, plays a prominent role in the development of yeast resistance to lipophilic weak-acid food preservatives. *Appl Env. Microbiol* 18, 7168–7165 (2006).

45. Destruelle, M., Holzer, H. & DJ, K. Identification and characterization of a novel yeast gene: the YGP1 gene product is a highly glycosylated secreted protein that is synthesized in response to nutrient limitation. *Mol Cell Biol.* 4, 2740–2754 (1994).

46. Minetaka, S. *et al.* Nuclear localization of Haa1, which is linked to its phosphorylation status, mediates lactic acid tolerance in Saccharomyces cerevisiae. *Appl. Environ. Microbiol* (2014). doi:10.1128/AEM.04241-13

47. Sugiyama, M. *et al.* Nuclear localization of Haa1, which is linked to its phosphorylation status, mediates lactic acid tolerance in *Saccharomyces cerevisiae*. *Appl Env. Microbiol* 80, 3488–3495 (2014).

48. DeVit, M. J. & Johnston, M. The nuclear exportin Msn5 is required for nuclear export of the Mig1 glucose repressor of *Saccharomyces cerevisae*. *Curr. Biol.* 9, 1231–1241 (1999).

49. Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623–627 (2000).

50. Swinnen, S. *et al.* Improvement of yeast tolerance to acetic acid through Haa1 transcription factor engineering: towards the underlying mechanisms (submitted).

51. Almeida, B. *et al.* Yeast protein expression profile during acetic acid-induced apoptosis indicates causal involvement of the TOR pathway. *Proteomics* 9, 720–732 (2009).

52. Guerreiro, J., Nuno, P. & Isabel, S. Adaptive response to acetic acid in the highly resistant yeast species *Zygosaccharomyces bailii* revealed by quantitative proteomics. *Proteomics* 12, 2303–2318 (2012).

53. Lindberg, L., Santos, A. X. S., Riezman, H., Olsson, L. & Bettiga, M. Lipidomic Profiling of *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii* Reveals Critical Changes in Lipid Composition in Response to Acetic Acid Stress. *PLoS One* 8, 1–12 (2013).

54. Burgess, R. R. & Deutscher, M. P. *Methods in Enzymology : Guide to protein purification*. (2009).

55. Cregg, J. M. *Pichia Protocols*. 389, (Humana Press, 2007).

56. Lin-Cereghino, G. ., Leung, W. . & Lin-Cereghino, J. *Expression of protein in Pichia pastoris.* (Methods Express. Scion Publishing Limited; Oxfordshire, 2007).

57. Idiris, A., Tohda, H., Kumagai, H. & Takegawa, K. Engineering of protein secretion in yeast: strategies and impact on protein production. 2010;86: 403–17. *Appl Microbiol Biotechnol* 86, 403–17 (2010).

58. Hou, J., Tyo, K., Liu, Z., Petranovic, D. & Nielsen, J. Metabolic engineering of recombinant protein production by *Saccharomyces cerevisiae*. *FEMS Yeast Res* 12, 491–510

59. Daly, R. & MTW, H. Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J Mol Recognit* 18, 119–138 (2005).

60. Dale, C., Allen, A. & Fogerty, S. *Pichia pastoris*: a eukaryotic system for the large-scale production of biopharmaceuticals. *Biopharm* 12, 36–40 (1999).

61. Technologies, L. *PichiaPink ™ Expression System Guide*. (2014).

62. Li, B., Cao, Y., Zhou, L., Liang, C. & Sun, F. A novel protein expression system-PichiaPink ™ - and a protocol for fast and efficient recombinant protein expression. 10, 19464–19472 (2011).

63. Chen, H. *et al.* Expression and Purification of Integral Membrane Fatty Acid Desaturases. 8, 1–

9 (2013).

64. Ahmad, M., Hirz, M. & Pichler, H. Protein expression in *Pichia pastoris :* recent achievements and perspectives for heterologous protein production. (2014). doi:10.1007/s00253-014-5732-5

65. Cregg JM, Vedvick TS, R. W. Recent advances in the expression of foreign genes in *Pichia pastoris*. BioTechnology 11: 905–910. *BioTechnology* 905–910 (1993).

66. Puxbaum, V., Gasser, B. & Mattanovich, D. The bud tip is the cellular hot spot of protein secretion in yeasts. *Appl Microbiol Biotechnol* 100, 8159–8168 (2016).

67. Briand, L. *et al.* Optimization of the production of a honeybee odorant-binding protein by *Pichia pastoris*. *Protein Expr. Purif.* 15, 362–369 (1999).

68. Koganesawa, N. *et al.* Construction of an expression system of insect lysozyme lacking thermal stability: the effect of selection of signal sequence on level of expression in the *Pichia pastoris* expression system. *Protein Eng* 14, 705–710 (2001).

69. Kowalski, J., Parekh, R. N., Mao, J. & Wittrup, K. D. Protein folding stability can determine the efficiency of escape from endoplasmic reticulum quality control. *Biol. Chem.* 273, 19453–19458. (1998).

70. Agaphonov, M. O., Romanova, N. V., Trushkina, P. M., Smirnov, V. N. & Ter-Avanesyan, M. D. Aggregation and retention of human urokinase type plasminogen activator in the yeast endoplasmic reticulum. BMC Mol. Biol. 3, 15. *BMC Mol. Biol.* 3, (2002).

71. Mattanovich, D., Gasser, B., Hohenblum, H. & Sauer, M. Stress in recombinant protein producing yeasts. 113, 121–135 (2004).

72. Waters, M., Evans, E. & Blobel, G. Prepro-alpha-factor has a cleavable signal sequence. *J Biol Chem* 263, 6209–6214 (1988).

73. Julius, D., Brake, A., Blair, L., Kunisawa, R. & J, T. Isolation of the putative structural gene for the lysine–arginine-cleaving endopepti- dase required for processing of yeast prepro-alpha-factor. *Cell* 37, 1075–1089 (1984).

74. Brake, A. *et al.* Alpha- factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 81, 4642–4646 (1984).

75. Kurjan, J. & Herskowitz, I. Structure of a yeast pheromone gene (MF alpha): a putative alpha-factor precursor contains four tandem copies of mature alpha-factor. *Cell* 30, 933–43 (1982).

76. Hegde, R. & Bernstein, H. The surprising complexity of signal sequences. *Trends Biochem Sci.* 31, 563–71 (2006).

77. Lin-Cereghino, G., Stark, C., Kim, D., Chang, J. & Shaheen N, Poerwanto H, Agari K, Moua P, Low LK, Tran N, Huang AD, Nattestad M, Oshiro KT, Chavan A, Tsai JW, L.-C. J. The effect of α-mating factor secretion signal mutations on recombinant protein expression in Pichia pastoris. *Gene* 519, 311– 317 (2013).

78. Rakestraw, J., Sazinsky, S., Piatesi, A. & Antipov E, W. K. Directed evolution of a secretory leader for the improved expression of heterologous proteins andfull-lengthantibodies in *Saccharomyces cerevisiae*. *Biotechnol Bioeng* 103, 1192–1201 (2009).

79. Hoseki, J., Ushioda, R. & Nagata, K. Mechanism and components of endoplasmic reticulum-associated degradation. *J Biochem* 147, 19–25 (2010).

80. Kohno, K. Stress-sensing mechanisms in the unfolded protein response: similarities and differences between yeast and mammals. *J Biochem* 147, 27–33 (2010).

81. Malhotra, J. & Kaufman, R. The endoplasmic reticulum and the unfolded protein response. *Semin Cell Dev Biol* 18, 716–731 (2007).

82. Yu, X., Sha, C., Guo, Y., Xiao, R. & Y, X. High-level expression and characterization of a chimeric lipase from *Rhizopus oryzae* for biodiesel production. *Biotechnol Biofuels* 6, (2013).

83. Dragosits, M. *et al.* The response to unfolded protein is involved in osmotolerance of *Pichia pastoris*. *BMC Genomics* 11, (2010).

84. *pET System Manual Novagen pET System Manual Novagen*. (2003).

85. Healthcare, G. E. Recombinant Protein Purification Handbook.

86. Healthcare, G. E. & Sciences, L. *Design of Experiments in Protein Production and Purification For more information refer to www.gelifesciences.com/handbooks*.

87. Healthcare, G. E. & Sciences, L. *Size Exclusion Chromatography For more information refer to www.gelifesciences.com/handbooks*.

88. Hinterdorfer, P. & Oijen, A. *Handbook of Single-Molecule Biophsics*. (Springer Science+Business Media, 2009).

89. Jalili, N. & Laxminarayana, K. A review of atomic force microscopy imaging systems : application to molecular metrology and biological sciences. 14, 907–945 (2004).

90. Liu, S. & Wang, Y. Application of AFM in Microbiology : A Review. *Willey Period. Inc* 32, 61–73 (2010).

91. Lyubchenko, Y. L., Shlyakhtenko, L. S. & Gall, A. A. *Methods in Molecular Biology, DNA - Protein Interactions*. 543, (Humana Press).

92. Morris, V. J., Kirby, A. R. & Gunning, A. P. *Atomic Force Microscopy for Biologists*. (Imperial College Press, 2009).

93. Kirat, K., Burton, I., Dupres, V. & Dufrene, Y. F. Sample preparation procedures for biological atomic force microscopy. *J. Microsc.* 218, 199–207 (2005).

94. Chang, K., Chiang, Y., Yang, C. & Liou, J. Atomic force microscopy in biology and biomedicine. *Tzu Chi Med. J.* 24, 162–169 (2012).

95. Alsteens, D. & Dufre, Y. F. Atomic force microscopy – looking at mechanosensors on the cell surface. (2009). doi:10.1242/jcs.106005

96. Francois, J. M. *et al.* Use of atomic force microscopy (AFM) to explore cell wall properties and response to stress in the yeast *Saccharomyces cerevisiae*. *Curr Genet* 59, 187–196 (2013).

97. Schiavone, M., Formosa-dague, C., Elsztein, C. & Teste, M. An Atomic Force Microscopy study of yeast response to ethanol stress : Evidence for a role of the plasma membrane in the nanomechanical properties of the cell walls. *Am. Soc. Microbiol.* 33, (2016).

98. Adya, A., Canetta, E. & Walker, G. Atomic force microscopic study of the influence of physical stresses on *Saccharomyces cerevisiae* and *Schizosaccharomyces pomb*e. *FEMS Yeast Res.* 6, 120–128 (2006).

99. Canetta, E., Adya, A. & GM, W. Atomic force microscopic study of the effects of ethanol on yeast cell surface morphology. *FEMS Microbiol Lett* 255, 308–315 (2006).

100. Pillet, F. *et al.* Uncovering by Atomic Force Microscopy of an original circular structure at the yeast cell surface in response to heat shock. 1–11 (2014).

101. Canetta, E., Graeme, M. W. & Ashok, K. A. Nanoscopic Morphological Changes in Yeast Cell Surfaces Caused by Oxidative Stress: An Atomic Force Microscopic Study. *J. Microbiol. Biotechnol* 19, 547–555 (2009).

102. Alsteens, D. *et al.* Structure , cell wall elasticity and polysaccharide properties of living yeast cells , as probed by AFM. *Nanotechnology* 384005, (2008).

103. Kirat-Chatel, S. *et al.* Nanoscale analysis of caspofungin-induced cell surface remodelling in *Candida albicans*. *Nanoscale* 5, 1105–1115 (2013).

104. Formosa, C. *et al.* Nanoscale effects of Caspofungin against two yeast species; *Saccharomyces cerevisiae* and *Candida albicans*. *Antimicrob Agents Chemother* 57, 3498–3506 (2013).

105. Khow, O. & Suntrarachun, S. Strategies for production of active eukaryotic proteins in bacterial expression system. 2, 159–162 (2012).

106. Macauley-patrick, S., Fazenda, M. L., Mcneil, B. & Harvey, L. M. Heterologous protein production using the *Pichia pastoris* expression system. *Wiley Intersci.* 22, 249–270 (2005).

107. Minning, S. *et al.* Optimization of the high-level production of *Rhizopus oryzae* lipase in *Pichia pastoris*. *J Biotechnol* 1, 59–70 (2001).

108. Zhang, W. & Meagher, M. M. Fermentation Strategies for Recombinant Protein Expression in the Methylotrophic Yeast *Pichia pastoris*. *Biotechnol. Bioprocess Eng.* 5, 275–287 (2000).

109. Posch, A. *Methods in Molecular Biology 2D PAGE: Sample Preparation and Fractionation*. 1, (Humana Press, 2008).

110. Hausner, W. & Thomm, M. Purification and Characterization of a General Transcription Factor, aTFB, from the Archaeon *Methanococcus thermolithotrophicus*. *J. Biol. Chem.* 268, 24047–52 (1993).

111. Andrabi, M., Mizuguchi, K. & Ahmad, S. Conformational changes in DNA-binding proteins: Relationships with precomplex features and contributions to specificity and stability Munazah. *Wiley Period. Inc.* 82, 841–857 (2013).

112. Dickerson, R. DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res* 26, 1906–1926 (1998).

113. Matsumoto, A. & Olson, W. Sequence-dependent motions of DNA: a normal mode analysis at the base-pair level. *Biophys J* 83, 22–41 (2002).

114. Trajkovic, S., Zhang, X. & Sylvia, D. Atomic Force Microscopy Study of Conformational Change of Immobilized Calmodulin. *Natl. Institutes Heal.* 27, 10793–10799 (2012).

115. Schaefer, C. & Rost, B. Predict impact of single amino acid change upon protein structure. *BMC Genomics* 13, S4 (2012).

116. Landrieu, I., Verger, A., Baert, J., Rucktooa, P. & Ferreira, E. Characterization of ERM transactivation domain binding to the ACID / PTOV domain of the Mediator subunit MED25. *Nucleic Acids Res.* 1–12 (2015). doi:10.1093/nar/gkv650

117. Brzovic, P. S. *et al.* The acidic transcription activator gcn4 binds the mediator subunit gal11/med15 using a simple protein interface forming a fuzzy complex. *Mol. Cell* 44, 942–953 (2011).

118. Chabot, P. R. *et al.* Structural and functional characterization of a complex between the acidic transactivation domain of EBNA2 and the Tfb1/p62 subunit of TFIIH. *PLoS Pathog.* (2014).

119. Jennings, B. H., Pickles, L. M., Wainwright, S. M., Roe, S. M. & Pearl, L.H. and Ish-Horowicz, D. Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. *Mol. Cell,* 22, (2006).

120. Langlois, C., , Mas, C., Di Lello, P., Jenkins, L. M. & Legault, P. and Omichinski, J. G. NMR structure of the complex between the Tfb1 subunit of TFIIH and the activation domain of VP16: structural similarities between VP16 and p53. *J. Am. Chem. Soc.,* 130, 10596–10604. (2008).

121. Sullivan, S. M. *et al.* Mutational analysis of a transcriptional activation region of the VP16 protein of herpes simplex virus. *Nucleic Acids Res* 26, 4487–4496 (1998).

122. Ahmad, S., Keskin, O., Sarai, A. & Nussinov, R. Protein – DNA interactions : structural , thermodynamic and clustering patterns of conserved residues in DNA-binding proteins. 36, 5922–5932 (2008).

123. Keskin, O., Ma, B., Nussinov, R. & Aviv, T. Hot Regions in Protein – Protein Interactions : The Organization and Contribution of Structurally Conserved Hot Spot Residues. *J. Mol. Biol.* 345, 1281–1294 (2005).

124. Bhardwaj, N. & Gerstein, M. Relating protein conformational changes to packing efficiency and disorder. 18, 1230–1240 (2009).

7 Annexes

7.1 Annex 1 – Digestion simulation of using pPINK-HC(+)-*HAA1*$_{WT}$ construct sequence NEBcutter software (New England Biolabs).

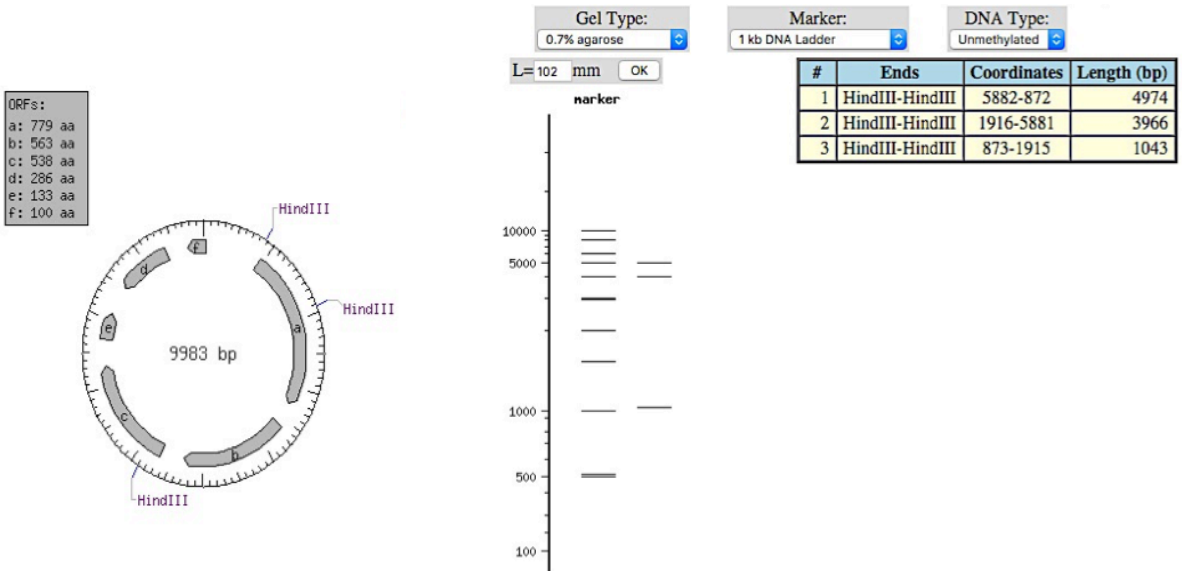

Figure 22 - Digestion simulation using NEBcutter software (New England Biolabs). pPINK-HC(+)-*HAA1*$_{WT}$ construct sequence used for overexpression of *HAA1* in *P.pastoris* was used as template to simulate enzymatic digestion with *Hind*III.