Jury

Chair:	Prof. Dr. Karin Coninx Hasselt University
Promoter:	Prof. Dr. Wanda Guedens Hasselt University, IMO-IMOMEC
Co-promoters:	Prof. Dr. Peter Adriaensens Hasselt University, IMO-IMOMEC
	Dr. Ties Steen Redeker Hasselt University, IMO-IMOMEC
Members of the jury:	Prof. Dr. Wouter Maes Hasselt University, IMO-IMOMEC
	Prof. Dr. Serge Muyldermans Vrije Universiteit Brussel
	Prof. Dr. Nick Devoogdt Vrije Universiteit Brussel
	Prof. Dr. Johan Robben Katholieke Universiteit Leuven
	Prof. Dr. Sylvia Dewilde University of Antwerp
	Prof. Dr. Karen Smeets Hasselt University, Biology (CMK)

Table of Contents

List of Figures

Chapter 4

Chapter 5

Chapter 6

List of abbreviations

5-FOA	5-fluoroorotic acid
aaRS	aminoacyl-tRNA synthetase
ADH1	alcohol dehydrogenase
AGT	alkylguanine-DNA alkyltransferase
AP	alkaline phosphatase
APS	ammonium persulphate
bp	base pair
CDR	complementarity-determining region
CH	constant domain
CHO	Chinese Hamster Ovary cells
dNTP	deoxynucleotide triphosphate
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
DTT	dithiothreitol
E. coli	*Escherichia coli*
EcTyrRS	*E. coli* tyrosyl-tRNA synthetase
EDTA	ethylenediaminetetraacetic acid
EIPCR	Enzymatic Inverse PCR
EMEA	European Medicines Agency
Fc domain	fragment crystallisable domain
FDA	Food and Drug Agency
GRAS	Generaly Regarded As Safe
GFP	Green Fluorescent Protein

H-chain	heavy chain
HcAB	heavy chain antibody
His$_6$-tag	polyhistidine tag (6x his)
IMAC	Immobilized Metal ion Affinity Chromatography
IMPACT	Intein Mediated Purification with an Affinity Chitin-binding Tag
IgG	immunoglobulin G antibody
IPTG	isopropyl-beta-D-thiogalactopyranoside
kbp	kilobase pair
kDa	kilodalton
L-chain	light chain
LB	Luria-Bertani
MCS	multiple cloning site
nb	nanobody
nbBCII10_His$_6$	nanobody BCII10 containing a C-terminal His$_6$-tag
NBT-BCIP	5-bromo-4-chloro-3-indolyl phosphate/nitroblue tetrazolium
OMP	oritidine-5'-monophosphate
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
p-azidoPhe	*p*-azidophenylalanine
p-propargyloxyPhe	*p*-propargyloxyphenylalanine
PFTase	farnesyltransferase
PVDF	Polyvinylidene fluoride
RNA	ribonucleic acid

SAM	S-adenosyl methionine
scFv	single-chain variable fragment
S. cerevisiae	*Saccharomyces cerevisiae*
SD	Single drop-out, e.g. SD-URA
SDS	sodium dodecyl sulphate
SDS PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SH group	thiol group
TBST	Tris-Buffered Saline and Tween 20
TCEP	tris(2-carboxyethyl)phosphine
TEF1 promoter	Promoter of the TEF1 gene encoding translational elongation factor EF-1 alpha
TEMED	N,N,N',N'-Tetramethylethane-1,2-diamine
TRP1	Gene encoding phosphoribosylanthranilate isomerase
tRNA	transfer RNA
UMP	uridine-5'-monophosphate
YEp	yeast episomal plasmid
YIp	yeast integrated plasmid

Amino acid	One-letter symbol	Three-letter symbol
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Aspartic acid	D	Asp
Cysteine	C	Cys
Glutamine	Q	Gln
Glutamic acid	E	Glu
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val

Chapter 1

Introduction

In this chapter an overview is given regarding the potential applications and benefits of chemically functionalized proteins and the existing strategies to prepare them. This is followed by the aims as well as the overview of the thesis.

The controlled immobilization of proteins on solid surfaces is an important step in the development and miniaturization of a wide range of applications e.g. biomedical implants and biosensing systems. All these applications demand a high sensitivity and selectivity, combined with a high reproducibility, which can only be obtained with an oriented and covalent coupling to the surface.

The strategies available nowadays only allow the formation of either an oriented or a covalent coupling, but not both simultaneously, and therefore do not meet the requirements that current technologies demand: sensitivity, selectivity and reproducibility. The site-specific modification of proteins with a non-natural functional group, which can serve as a unique chemical "handle" to immobilize proteins in an oriented and covalent way, fulfils these three main requirements.

In this work the site-specific modification of proteins is based on the incorporation of a non-natural amino acid by means of amber suppression. This is done by expanding the genetic code of *Saccharomyces cerevisiae* (*S. cerevisiae)* with the genes of an orthogonal, mutant *E. coli* tyrosyl-tRNA synthetase (EcTyrRS)/tRNA$_{CUA}^{Tyr}$ pair.

1.1 Applications for site-specifically modified proteins

Biosensors. Roughly speaking a biosensor consists out of three parts: i) a receptor, ii) a signal transducing platform and, iii) a data processor[1,2]. Biosensors combine the selectivity of biochemical receptor molecules with sensitive optical or electrical transduction methods to detect concentrations of target molecules as low as possible in various matrices. The biological receptor can be very diverse: single stranded DNA (ssDNA), double stranded DNA (dsDNA), peptides, enzymes, antibodies, *etc*[2]. These receptors are capable of recognizing and binding their target molecules, resulting in physico-chemical changes in the transducer, which in turn will be converted into a measurable signal.

Despite the symbiosis between biotechnology and material sciences, the use of biosensors in routine diagnostics is not common. The main cause is the orientation of receptor molecules on the substrate surface. An optimal sensitivity and reproducibility can only be obtained when the active site remains available for the target molecule. Besides this, a stable (covalent) binding is required between the receptor molecules and the substrate. Figure 1.1 illustrates the difference between random (A) and oriented (B) protein immobilization and its influence on target binding (e.g. antigens)[3]. Despite the progress made for spatially oriented and covalently coupled DNA to numerous surfaces, the orientation of proteins had proven to be more challenging. The difference consists in the simplicity to modify DNA in either its 3' or 5' terminus. On the other hand, proteins contain different amino acids with reactive side chains, which can be present more than once, making it difficult to obtain one unique location for covalent coupling to the substrate. Therefore, the site-specific incorporation of a non-natural amino acid, on a strategically chosen location, could offer a solution for this issue.

The controlled immobilization of proteins to solid surfaces is of great importance in biotechnology, especially for the development and miniaturization of bioactive materials, imaging tools and advanced biosensors. These biosensors can be used in e.g. the healthcare sector for the detection of early disease markers, allowing an early start of a therapy. The same biosensors can also be used to evaluate the success of a certain treatment. Besides the healthcare sector, these improved biosensors can also be a benefit for the environmental and food sector. Small handheld devices can be used on scene to detect the presence of toxins or microorganisms in the environment or in the food products.
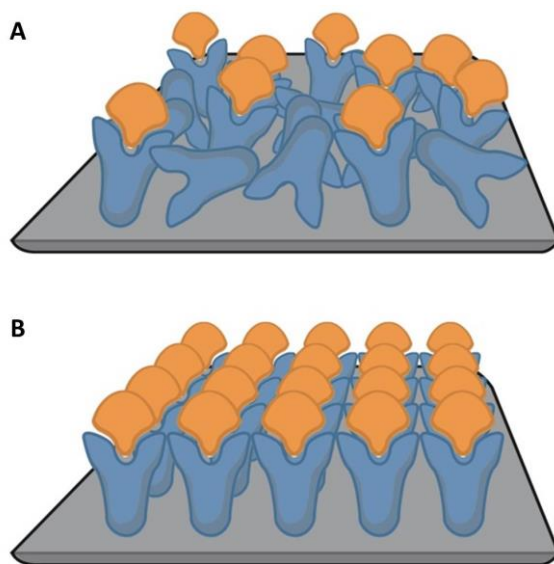
Figure 1.1. The difference between random (A) and oriented (B) protein immobilization and the influence on target binding (adapted from Steen Redeker *et al*[3]).

Generating new protein properties. In every organism, each protein is built by means of 20 naturally occurring amino acids. This, together with the addition of genetically encoded non-natural amino acids, could result in evolutionary advantages to modern organisms as well as in an extended ability to manipulate the biological and physicochemical properties of proteins. For example, engineering metalloproteins with protein-bound metal ions, involved in catalytic and electron transfer reactions in cells, is currently challenging, since it is difficult to control and predict the amino acid shell surrounding the metal ion. This could be facilitated by using multidentate, metal-chelating amino acids with coordinating atoms preoriented in the correct conformation, as shown by Schultz *et al.*[4,5]. The use of non-natural amino acids, *i.e.* amino acids with side chains corresponding to products of post-translational modification, allows the production of these proteins in simple organisms, and can lift the sequence restrictions to which post-translational modifications must adhere. For example, a phosphorylation mimic resistant to phosphatase activity can be used to study the role of a specific phosphorylated residue. This can be used to overcome the complexities seen with removal of this post-translational modification by phosphatases[6].

Protein modification for therapeutics and diagnostics. The site-specific introduction of non-natural amino acids also finds its use in the generation of protein-based therapeutics. For example in cancer, inflammation and protein-misfolding diseases, the organisms own proteins (self-protein) are involved in the disease, to which the body does not have an immune response. Immunogenic amino acids can be used to break this immunological tolerance and generate therapeutic vaccines. It has been shown that the incorporation of e.g. *p*-nitrophenylalanine in a target protein results in immunogenic epitopes that after immunization induce high-titer, long-lived antibodies cross-reactive with native proteins. When *p*-nitrophenylalanine was incorporated into murine TNF-α, it induced antibodies cross-reactive with native murine TNF-α and protected against lipopolysaccharide-induced death[7,8]. This strategy has also been applied against other self-antigens, such as C5a (involved in inflammation), PCSK9 (involved in cardiovascular disease) and EGF and HER2 (involved in cancer), as well as towards conserved epitopes of human immunodeficiency virus (HIV). The fact that a single *p*-nitrophenylalanine residue can break the immunological tolerance might support the theory that enzymatic posttranslational generation of nitrotyrosine in proteins, stimulated by local inflammation and cytokine release, could create immunogenic epitopes potentially initiating many autoimmune diseases[9].

Another therapeutic application is the production of proteins containing a site-specifically attached radioisotope, toxin, drug, polyethylene glycol (PEG) or even another protein (bispecific therapeutics). PEGylation can be used for different purposes, e.g. the shielding of antigenic and immunogenic epitopes, shielding receptor-mediated uptake by the reticuloendothelial system (RES) and preventing recognition and degradation by proteolytic enzymes[10]. It has also been shown that PEGylation successfully increases the serum half-life of proteins[11].

Protein modification has also applications in diagnostics. Antibodies targeted against tumour-associated proteins can be labelled with a radionuclide and be used as tracers for non-invasive *in vivo* tumour imaging. For example, radiolabelled antibodies (e.g. nanobodies) have been developed to assess breast cancer diagnosis in a preclinical setting[12], to improve prognosis of cancer therapy[13] and to evaluate changes on the status of inflammatory responses by visualising dendritic cells[14]. The site-specific modification of proteins can also be a valuable

tool for the field of theranostics. In this case a specific antibody can be labelled with an imaging probe and be conjugated to a drug, allowing for the simultaneous diagnosis and treatment of diseases such as cancer[15].

Study of the structure and function of proteins. The site-specific incorporation of biochemical probes into proteins makes it more feasible to study the relationship between structure and function in an accurate way. The analysis of the structure of proteins and large complexes with NMR is very complicated. For this reason, the introduction of NMR labels at defined locations in proteins would simplify the complexity and signal assignment. For example, ^{15}N-labeled *o*-nitrobenzyltyrosine was used as an NMR probe by Cellitti *et al.*[16]. Also non-natural amino acids with infrared (IR) or X-ray diffraction probes have been used to study the structure and dynamics of proteins. One of these probes is *p*-cyanophenylalanine, used for infrared spectroscopy, due to its characteristic cyano group that absorbs at a clear spectral window at ~2200 cm^{-1} and is sensitive to small environmental changes. It was incorporated into myoglobin by Schultz *et al.*[17], to probe the binding between ligand and metal ion. Another example of an infrared probe is *p*-azidophenylalanine, with an absorbance in a clear spectral window at ~2100 cm^{-1}. It showed to be sensitive to changes in an electrostatic environment when it was site-specifically incorporated at various sites in rhodopsin[18]. For phase determination in X-ray crystallography, the most critical step is the introduction of heavy atoms into protein crystals. An easy way to do this is e.g. to genetically incorporate *p*-iodophenylalanine into proteins in yeast or *E. coli*[19]. In addition, determining the X-ray crystal structures of *O*-acetylserine sulfhydrylase and *N*-acetyltransferase was facilitated by the site-specific incorporation of non-natural amino acids containing heavy atoms[20,21]. The possibility of incorporation at multiple sites, the strong anomalous signal and the use of an internal X-ray source should make it easier to solve protein structures in a high-throughput manner[22].

Besides the *in vitro* applications mentioned above, the genetic encoding of non-natural amino acids can also be used for *in vivo* applications. For example, small fluorescent non-natural amino acids (e.g dansylalanine) can be genetically encoded at surface sites in a protein with minimal structural perturbation. Thus, it is more convenient than traditional fluorescent tags such as Green Fluorescent Protein (GFP) and derivatives, which are limited by their relatively large size. A recent application was the incorporation of a fluorescent non-natural amino

acid into histones[9]. This was done to visualise the nuclear localisation of histones in yeast and mammalian cells, by means of fluorescence microscopy. Since the fluorescent properties of some non-natural amino acids are sensitive to changes in their environment and these amino acids can be incorporated at or near a site of interest, it is possible to use them for the characterisation of protein folding, biomolecular interactions and local protein conformational changes[21,23–25]. Besides fluorescent probes, photocaged and photoisomerisable non-natural amino acids have been used to study protein function. They have been used *in vitro* to site-specifically cleave proteins[26], photoregulate ligand-protein binding[27], probe protein–nucleic acid interactions[28] and *in vivo* to study cellular processes[29]. Non-natural amino acids capable of photocross-linking have also been used to characterise protein-protein and protein-nucleic acid interactions *in vitro* and *in vivo*[30]. In addition, most biological interactions are transient or unstable in *in vitro* isolation conditions. Therefore, covalent cross-linking of the interacting molecules can be useful to isolate relevant complexes.

1.2 Protein immobilization and modification strategies

Nowadays a large variety of strategies is available for the modification and/or immobilization of proteins. Some strategies are only useful for weak protein immobilization onto surfaces (e.g. physical adsorption), while others can be used for both protein modification as well as immobilization to a surface (e.g. the manipulation of the amino acid's functional groups present on the protein surface). Later on, a distinction can be made between *in vitro* and *in vivo* protein engineering strategies[3].

1.2.1 Non-oriented, noncovalent immobilization

Physical adsorption. Many protocols for protein-immobilization are based on noncovalent, non-oriented adsorption[31-33]. A frequently used method for the immobilization of proteins on solid surfaces is physical adsorption. This method involves the immobilization of proteins via weak interactions between the protein and the surface (*i.e.* hydrogen bonds, electrostatic interactions, hydrophobic interactions and van der Waals interactions)[34]. It is a relatively simple method, where the substrate is "dipped" in a protein solution and incubated for a time period to allow the physical adsorption to occur. Another way is to add a solution of the protein to the surface and allow it to dry, after which the

unbound proteins are rinsed away[35]. Despite the large number of applications for this technique, its main disadvantage consists of the random orientation of the proteins when adsorbed onto the substrate surface. This results in: i) reduced accessibility of the functional sites[34] and, ii) conformational changes influencing the proteins activity. The extent of conformational changes depends on the protein (size, intrinsic stability, surface charge, H-bonds, …), the type and chemistry of the surface, the substrate's surface hydrophobicity or hydrophilicity and the coupling conditions such as temperature, pH, and ionic strength of the solution. All these factors have a large influence on the 3D conformation, therefore making it difficult to immobilize proteins in a reproducible way[36]. An additional disadvantage is that the adsorbed proteins might leach (wash away), resulting in a partial or complete loss of activity of the bioactive surface over time[37].

1.2.2 Non-oriented, covalent immobilization

The problem of protein detachment from bioactive surfaces can be solved by means of covalent coupling to immobilize proteins. This more robust way to create biofunctionalized surfaces requires the presence of two mutually reactive chemical groups on the protein and substrate surface. In order to avoid protein denaturation, the reaction should work under physiological conditions (i.e aqueous buffers and neutral pH). Most methods found in literature describe a covalent coupling strategy that uses the reactivity of endogenous functional groups present in the amino acid's side chain. Obviously, these methods are not restricted to couple proteins on solid surfaces but can also be used to modify proteins via post-translational modifications with biochemical probes, tags or groups influencing the protein stability or activity. Amines, thiols and carboxylic acid groups are most commonly used. Figure 1.2 gives an overview of the functional groups frequently used for covalent immobilization reactions.

Amine and Carboxyl Chemistry. Covalent immobilization of proteins is mostly done using the amine groups present on the surface of the protein. Lysine is present in most proteins at about 6% to over 10% of the overall amino acid sequence and these moieties are often located at the surface of the protein[38,39]. Lysines provide good stability and are very reactive towards electrophilic agents without the need to be activated[37,40,41]. A commonly used tool is *N*-hydroxysuccinimide (NHS) ester which reacts with protein amine groups forming stable peptide

bonds. This method has for example been used to immobilize the highly toxic protein ricin on a gold surface[42]. Ricin has nine lysine residues at its surface, which all can react with NHS to immobilize the protein. However, different protein orientations appeared at the substrate as a result of the multiple lysine residues accessible on the protein surface. Generally, NHS is used to activate carboxylic acids. Consequently, proteins can be immobilized on carboxyl-functionalized surfaces, as demonstrated by Pei *et al.*[43]. Here, they studied the effect of different activation reagents on carboxylic acids and the coupling efficiency with monoclonal antimyoglobin antibody at pH 4. The reagents were:

   i.   1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) resulting in a positively charged surface,
   ii.   EDC/NHS resulting in a neutral surface,
   iii.   EDC/sulfo-NHS resulting in a negatively charged surface.

Since the antibody has a pI of 7-8, it will be positively charged at pH = 4, thus rendering the EDC/sulfo-NHS reaction the most efficient at these conditions. This clearly shows that electrostatic attraction can be essential for successful immobilization, which is dependent on the pK_a of the surface functional groups, the pH of the reaction buffer and the pI of the protein.

Another study making use of the amine group of lysine was performed by Chen *et al.*[44]. In this study they combined the covalent coupling using an amine group present on the surface of the protein, with the noncovalent coupling of a bifunctional linker molecule to a surface. They were able to demonstrate the immobilization of ferritin, streptavidin and biotinyl-3,6-dioxaoctanediamine (biotin-PEO-amine) on the sidewalls of single-walled carbon nanotubes (SWNTs). In this case the protein was first covalently coupled to biotin and the complex was then attached to the SWNTs in a non-covalent way, in order to preserve the sp^2 nanotube structure and thus its electronic characteristics. For the non-covalent reaction, the bifunctional molecule 1-pyrenebutanoic acid, succinimidyl ester is irreversibly adsorbed on the hydrophobic surface of the SWNTs. This highly stable adsorption leads to a surface of SWNTs functionalized with succinimidyl ester groups, which are highly reactive towards nucleophilic substitution by primary and secondary amines located on the protein surface. Using amines to make covalent bonds will result in stable bonds, therefore preventing the leaching out of the proteins during use.

Nevertheless, the disadvantage is that there is no control on the orientation of how the protein is coupled. When multiple amine groups are available, this will result in multiple orientations in which the protein is coupled to the surface. This will result in a heterogeneous surface and will make it difficult to produce bio-active surfaces in a reproducible way.

Thiol Chemistry. Free cysteines are quite rare in proteins and contain a reactive thiol (SH) group that mostly appears in oxidized disulfide bridges[45]. Because of its low natural abundance, cysteine is an interesting tag for immobilization reactions since random orientations or multiple contact points are less probable. In case no free cysteines are present on the surface of the protein, they can easily be introduced at a site of interest by means of site-directed mutagenesis. Maleimides are commonly used to selectively modify surfaces towards covalent protein coupling via the free thiols. Reactions based on maleimide-cysteine coupling require that the protein (cysteine) remains in a reduced form. In order to prevent the inactivation of cysteine by the formation of disulfide bridges, reducing agents such as dithiothreitol (DTT) and tris(2-carboxyethyl)phosphine (TCEP) can be used[46]. Ménard *et al*.[47] reported a specific maleimide-cysteine coupling reaction. A human cytochrome P450 enzyme was immobilized on maleimide functionalized agarose beads and silica microspheres. Three out of four cysteines were replaced by other amino acids, by means of site-directed mutagenesis. This allowed the protein to be site-specifically immobilized through the remaining cysteine, and the enzyme was proven to be still active after immobilization. Besides maleimides it is also possible to functionalize surfaces with e.g. vinyl-sulfone groups[48] or disulfide probes[45].

It can be concluded that several endogenous chemical groups present in amino acids can be used to covalently couple proteins on complementary functionalized surfaces. Since in general multiple copies of amino acids are present in proteins, these chemical groups are not unique and can result in multiple protein orientations on the substrate surface. The orientation is difficult to predict or alter if it is not optimal. Therefore, for prospective applications where sensitivity and reproducibility are important, more site-directed immobilization techniques are needed, whereby a controllable and homogeneous orientation of the proteins can be achieved.
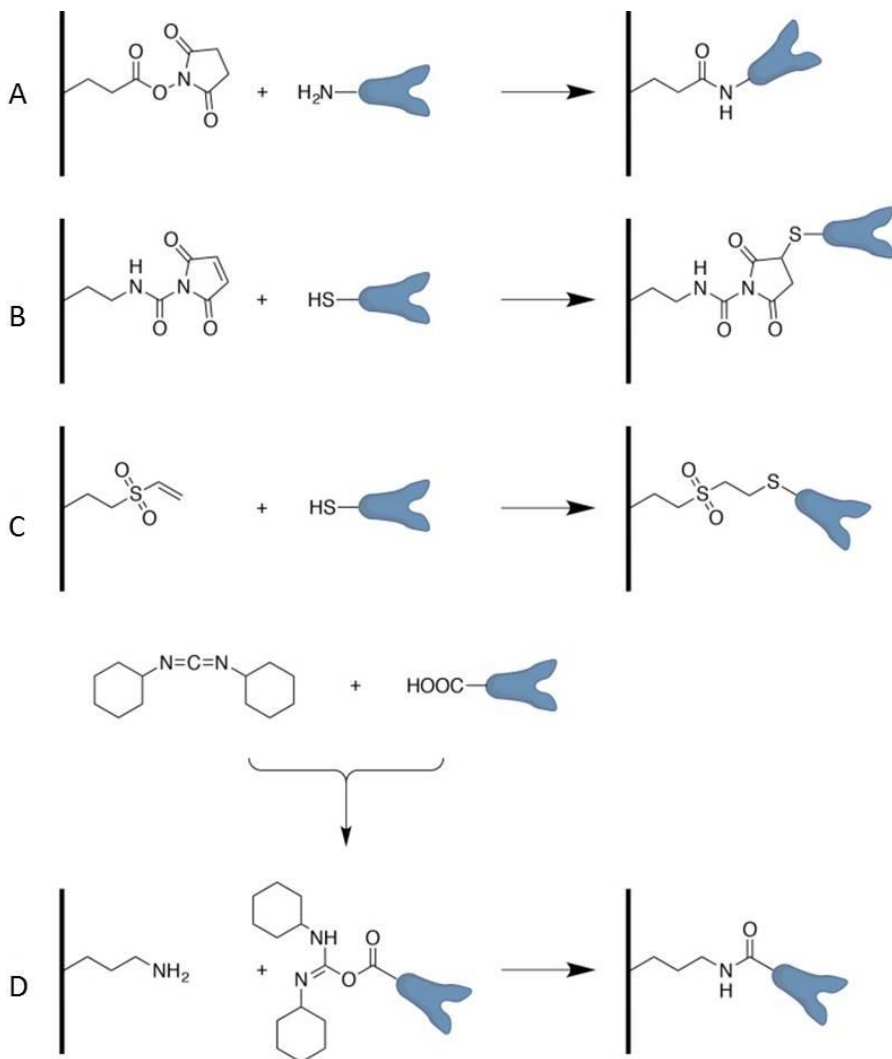
Figure 1.2. Reactions of the main functional groups of endogenous amino acids commonly used in immobilization chemistry (adapted from Steen Redeker *et al.*[3]): (A) immobilization via the amine group on *N*-hydroxysuccinimide (NHS) modified substrates, (B) immobilization via thiols on maleimide, (C) immobilization via vinyl sulfone, (D) immobilization via *N,N'*-dicyclohexylcarbodiimide-activated (DCC) carboxylic acid groups on amine-modified substrates.

1.2.3 Oriented, noncovalent immobilization

Since it is difficult to control protein immobilization solely relying on the endogenous functional groups, much effort has been put in the development of alternative strategies to generate homogenous and reproducible substrate surfaces. This controlled, site-specific conjugation approach can be divided in a noncovalent and a covalent strategy. First the noncovalent, site-specific approach is discussed, which is mostly based on complementary affinity interactions between biomolecules. The high selectivity of these interactions allows the control of the protein alignment on substrates, functionalized with the respective counterpart when the selective biological tag is site-specifically introduced in the protein. A way of introducing such a biological tag in a protein is by use of recombination techniques. Thus, genetic engineering of the protein of interest is required.

Polyhistidine. A commonly used affinity tag is the polyhistidine or His-tag. This approach is based on the interaction of histidine or histidine-rich regions in proteins with divalent metal ions, such as Ni^{2+}, Cu^{2+} and Zn^{2+}. The use of His-tags was inspired by immobilized metal ion chromatography (IMAC)[49]. The histidine tag is recombinantly added to either the N- or C-terminus of the protein. When incubated on a metal treated surface (through a chelating agent like nitrilotriacetic acid (NTA)) the polyhistidine-tail of the protein interacts selectively with the metal. Such polyhistidine–metal interactions are reversible and can be undone by the addition of competitive ligands (e.g. imidazole or histidine) or by metal chelators such as ethylenediaminetetraacetic acid (EDTA). The advantage of this technique is its reversibility, which can be useful for applications where surface reusability is important. Recent applications of His_6-tagged proteins have been reported by Ganesena *et al.*[50] and by Oshige *et al.*[51]. Ganesena *et al.* described the site-specific affinity immobilization of His_6-tagged acetylcholinesterase onto Ni/NiO nanoparticles for the development of electrochemical screen-printed biosensors. Moreover, Oshige *et al.* describes a method involving the modification of chitosan with nitrilotriacetic acid (NTA) to achieve a larger amount of His-tagged protein immobilization. However, reversibility could also be a disadvantage for applications that require stability and a larger storage time (long shelf-life).

Peptide epitope tags. Another common strategy is the use of peptide epitope tags. An epitope is the part of an antigen that is recognized by

an antibody. With recombinant techniques, the genetic sequence coding for the epitope tag can be fused to the protein of interest. Although a peptide epitope tag can be placed anywhere in the protein, it is typically placed either on the N- or the C-terminus. In theory any short peptide known to bind an antibody could be used as an epitope tag, but in reality there are a few commonly used epitopes e.g. the FLAG-tag and the myc-tag.

The *FLAG-tag* is a short, highly charged and therefore water soluble octopeptide with the amino acid sequence of Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys[52]. Because of its hydrophilic character, the FLAG-tag is most likely situated on the surface of the recombinant protein, which makes it interesting for immobilization. Currently, commercial vectors with the FLAG-tag up- or downstream of multiple cloning sites are available, which facilitates the construction of proteins with a N- or C-terminal FLAG-tag[53]. The FLAG-tag contains the binding site for several highly specific anti-FLAG poly- and monoclonal antibodies.

The *myc-tag* is an epitope consisting of eleven amino acids (Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu-Asn) and is derived from the c-myc gene. The anti-c-myc antibody was developed by Evan *et al*.[54]. It has been shown that the fusion of the myc-tag sequence to the N- or C-terminus of target proteins does not affect its affinity towards the anti-c-myc antibody[55]. The c-myc tag has been used for protein immobilization by Wingren *et al*.[56]. Here a proof-of-principle was reported for the design of a protein microarray based on myc-tagged single-chain antibody fragments.

Biotine-(Strept)avidin interaction The strong interaction between (strept)avidine, biotine and biotinylated proteins is also an efficient affinity based immobilization technique. Holland *et al*.[57] used this method to immobilize a biotinylated aldo/keto reductase on streptavidin-coated surfaces. It was demonstrated that the activity of the site-specifically immobilized enzyme was 60- to 300-fold greater than that of the randomly immobilized enzyme and was comparable to the activity of the wild-type enzyme in solution. Yu *et al*.[58] managed to immobilize a membrane-bound sialyltransferase via a biotine-labeled C-terminal cysteine on streptavidin functionalized magnetic nanoparticles. Using this biotin-streptavidin interaction it was possible to immobilize the protein under mild conditions, with an approximate twofold increase in activity compared to other conjugation techniques. Another

interesting application based on the binding of biotine to streptavidin, is the so-called Strep-tag® (IBA, Goettingen, Germany). This small peptide tag (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) was originally developed for protein purification and it is also used for protein immobilization, e.g. the immobilization of fusion proteins on the commercially available Biacore sensor chip CM5 (IBA).

DNA-mediated protein immobilization. Protein-DNA conjugates have also been used for the immobilization of proteins. For example, Fruk *et al*.[59] described DNA-directed immobilization of horseradish peroxidase on electrochemically active surfaces. Immobilization by means of DNA-mediators is interesting because with DNA, the sequence design as well as the base pairing are easy to perform. Although many methods exist to prepare these protein-DNA conjugates[60], there is still no method to site-specifically couple the DNA to the protein.

Besides the ones mentioned above, other affinity based methods exist e.g. the binding of Glutathione-S-transferase (GST) to glutathione, the binding of the maltose binding protein with maltose, the binding of antibodies to protein A or protein G via the Fc-region or chitin-binding protein with chitin[36,61–64]. For some applications, the noncovalent type of immobilization based on affinity tags is beneficial due to the possible regeneration and reusability of the surface, while this could be a disadvantage for other applications. Besides oriented, noncovalent immobilization, these affinity tags can also be used in a two-step process in which the protein is first directed and oriented to the surface by the affinity tag, after which a covalent coupling can take place between the oriented protein and the surface. This two-step method will result in an oriented and covalent coupling of the protein[65].

1.2.4 Site-directed, covalent conjugation

In order to have an oriented and covalent coupling to a substrate, a unique chemical group or sequence has to be site-specifically introduced in the protein. Coupling reactions are often done in complex mixtures of biomolecules. This means that even if a unique functional group is present on the target protein, the same functional group could be present in other (unwanted) biomolecules within the reaction mixtures. As a consequence, these unwanted proteins could also couple to the reactive surface. Therefore, the site-specific incorporation of a single, unique (non-natural) bioorthogonal functional group would improve the

selective coupling of the target protein to the surface. This bioorthogonal functional group does not appear on nor does it cross-react with endogenous amino acids. The functional group should be strategically introduced on a location with a minimal influence on the target binding or active site. This bioorthogonal group can then be used as a chemical handle to immobilize the protein on a complementary functionalized surface. A number of biorthogonal chemistries have been reported, but it are the so-called 'click' reactions that have become very popular in both industry and academic research. The 'click' chemistry concept[66–68] consist of a set of highly selective reactions in organic chemistry. It contains a set of characteristics including i) mild reaction conditions, ii) insensitivity to oxygen and water and, iii) the formation of a stable product under physiological conditions (aqueous environment and pH~7). Figure 1.3 gives an overview of the biorthogonal chemistries used for protein immobilization. When adding the complementary functionality to the substrate, these 'click'-functionalized proteins can be coupled in a covalent and oriented way, resulting in a high sensitivity, selectivity and reproducibility.

Cycloaddition. The copper-catalyzed Huisgen 1,3-dipolar cycloaddition of azides and alkynes, resulting in a 1,2,3-triazole, is the archetype of 'click' chemistry[69,70]. Although several Cu(I) sources can be used directly, in many cases the catalyst is prepared *in situ* by the reduction of Cu(II) salts by e.g. tris(2-carboxyethyl)phosphine (TCEP) or ascorbic acid. Unfortunately, for *in vivo* reactions the copper catalyst can be toxic to cells. Another disadvantage is that copper might cause the precipitation of proteins, when being used to immobilize proteins on a surface[71,72].

A copper-free Huisgen cycloaddition has been reported using the release of ring strain energy of a cyclooctyne group to enable the Huisgen 1,3-dipolar azide-alkyne cycloaddition to proceed without the need of an additional catalyst[73,74]. It has been shown that 'click' chemistry can be used for protein modifications[75], making the search for efficient methods for the introduction of azides and alkynes in proteins an interesting research topic. Both azides and alkynes have been successfully introduced without influencing structure and function of the protein[76–79].

Staudinger ligation. In the Staudinger ligation reaction, an azide reacts with a triphenylphosphine to form an aza-ylide. However, in the presence of water this aza-ylide hydrolyzes spontaneously to form a

primary amine and a phosphine oxide. To solve this issue, a modified Staudinger reaction was reported, where an ester group was placed on one of the phosphine's aryl substituents[80]. This results in a stable amide bond, instead of an amine bond.

Diels-Alder reaction The Diels-Alder cycloaddition occurs between a dienophile and a conjugated diene, leading to a six-membered unsaturated ring. These types of cycloadditions proceed in a higher rate in water at room temperature than in organic solvents. For this reason, the Diels-Alder cycloaddition has also been used for protein immobilization[81,82].

Thiol-Ene reaction. In this strategy, a thiol is added to an "ene" group via a radical mechanism. This reaction can be initiated either by a radical mechanism or by light. Although it was mostly used for polymer synthesis, it has also attracted researchers from other fields, because of its biocompatible 'click' characteristics. A benefit of this reaction is its photoactivatable character, allowing the possibility to create patterns of immobilized proteins on a surface. Moreover, this surface patterning can be achieved by using photo masks or by laser irradiation[83–85]. The main disadvantage of the Thiol-Ene reaction is that thiols are not bioorthogonal, which might result in potential cross-reactions with other proteins.

Oxime ligation. The oxime ligation is a condensation reaction between aldehydes or ketones with oxyamines or hydrazides, forming a stable oxime linkage. These reactions are generally slow and normally proceed around pH 4-5, making them not very useful in biological applications[86]. However, this reaction recently became of interest for biological systems, with the discovery of aniline as a catalyst making the reaction possible at neutral pH[87,88]. Neither oxyamine/hydrazide nor aldehyde/ketone functionalities are present in proteins, making them suited to be used for the site-specific immobilization of proteins on surfaces[86,89].
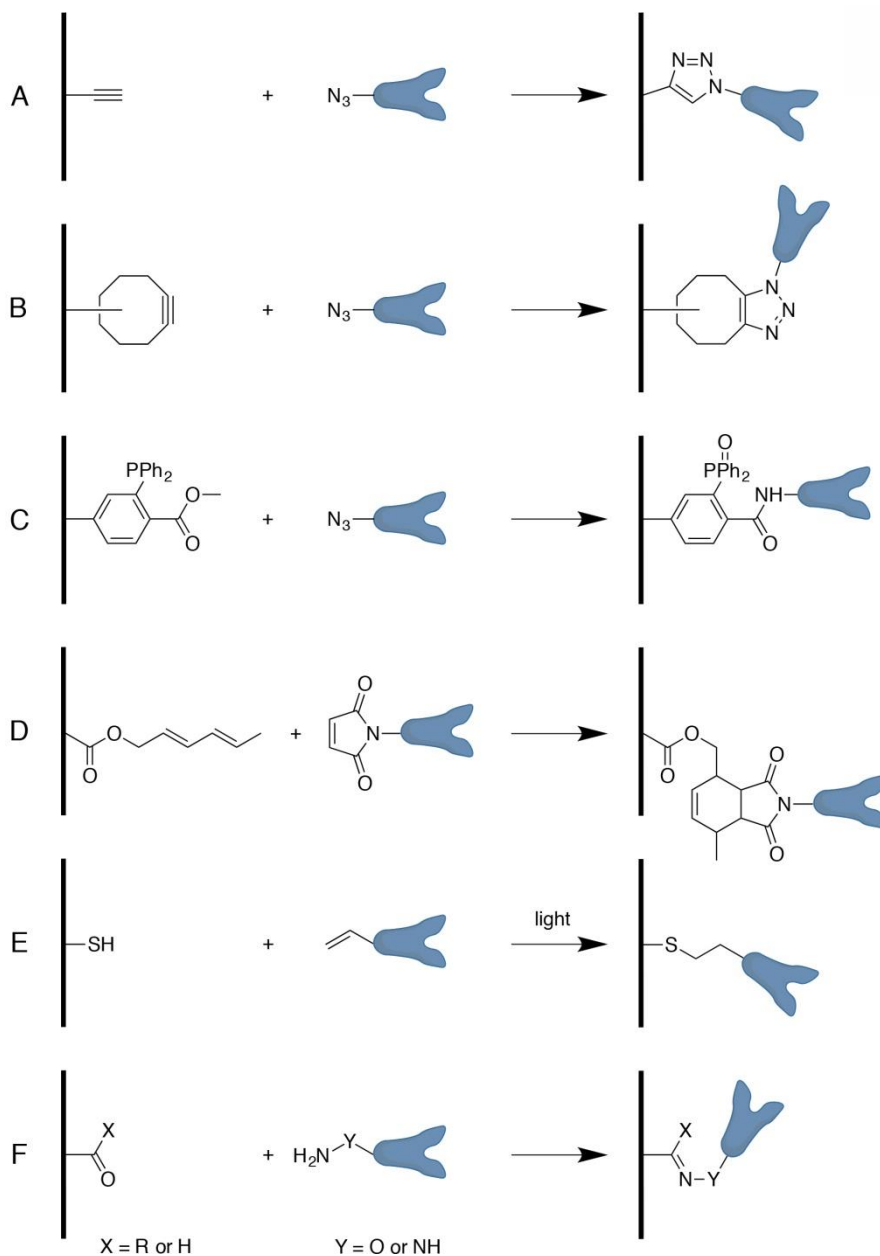
Figure 1.3. Bioorthogonal chemistries used for protein immobilization: (A) copper catalyzed cycloaddition, (B) ring-strain catalyzed cycloaddition, (C) modified Staudinger ligation, (D) Diels−Alder cycloaddition, (E) thiol−ene additions, (F) oxime ligation(adapted from Steen Redeker *et al.*[3]).

1.2.5 Protein Engineering

There is a large number of biological and chemical techniques that allow the introduction of bioorthogonal functionalities into proteins. The easiest approach is the use of bifunctional cross-linkers that can be coupled to the side chains of amino acids, mostly lysine or cysteine. Nevertheless, this method presents the same disadvantage as the endogenous functional groups of amino acids, used for covalent immobilization.

Since these functional groups are present on multiple locations, the possibility to attach these bifunctional linkers in a site-specific manner is very limited. This can result in modifications at unfavorable position, e.g. on the antigen-binding site or on the active site, and in mixtures of modified proteins, which are difficult to purify or characterize with biochemical tools[90,91]. For this reason, other techniques for site-specific protein modification are needed.

1.2.5.1   *In vitro* protein modification

Enzymatic modification. Enzymes that are involved in post-translational modifications are also interesting for introducing site-specific modifications. These enzymes normally recognize a short peptide sequence and covalently modify proteins with small molecules. Enzymatic modification can be done by a three-components system containing:

1. The target protein fused to a specific tag,
2. The modifying enzyme recognizing the tag and,
3. The substrate of the enzyme, that has to be coupled to the target protein.

An example of such an enzyme is the *E. coli* enzyme biotin ligase (BirA)[92,93]. BirA catalyzes the biotinylation of a lysine side chain within a consensus 'acceptor peptide'. The *E. coli* BirA is orthogonal to the peptides that are recognized by mammalian biotin ligase, meaning that mammalian proteins tagged with the BirA recognition sequence can be selectively biotinylated and coupled to streptavidin-functionalized surfaces[94]. It was shown that the tolerance of *E. coli* for unnatural substrates is limited to modified biotin isoesters. However, it was discovered that the yeast biotin ligase accepts an alkyne derivative of

biotine[92], while the *Pyrococcus horikoshii* biotin ligase is capable of using both alkyne and azide biotin analogues[95].

Sortase A (SrtA) has also been extensively studied for potential protein immobilization. This transpeptidase from *Staphylococcus aureus* is located in its plasma membrane and catalyzes a cell wall sorting reaction, attaching surface proteins to the cell wall envelope. It cleaves proteins between a threonine and a glycine residue within the recognition motif LPXTG (with X being any amino acid). This cleavage occurs near the C-terminus thereby generating a covalent enzyme intermediate. Then the N-terminus of an oligoglycine nucleophile attacks the threonine carboxyl group of the thioester intermediate, resulting in the formation of a covalent peptide bond[96]. In recent years, this reaction has been exploited for a range of biotechnological applications, including the covalent immobilization of proteins on solid supports[97]. Recently, Jiang *et al.* reported the successful immobilization of recombinant human thrombomodulin onto an N-terminal diglycine-functionalized glass slide surface via SrtA-mediated ligation[98]. Besides biotin ligase or Sortase A, many more enzymes can be used for the immobilization of proteins. For example, protein farnesyltransferase (PFTase) catalyzes the transfer of a farnesyl isoprenoid group from farnesyl diphosphate to the sulfur atom of cysteine present in the tetrapeptide sequence CAAX positioned at the C-terminus of the target protein. PFTase can modify proteins with substrates that incorporate biorthogonal groups, such as alkynes and azides[99]. Other enzymes include transglutaminase, iopoic acid ligase, formylglycine-generating enzyme and phosphopantetheinyl transferases[100].

Self-labeling Protein Tags. Enzymatic protein modification can also be done by so-called self-labeling protein tags. These tags, mostly derived from enzymes[101], react covalently with a substrate. In this system the auxiliary enzyme is the tag itself and is directly fused to the target protein, making it a two-component system[102].

One of these self-labelling tags is the SNAP-tag[103]. It has been used as a tool for the labelling as well as for the immobilization of proteins[104]. The SNAP-tag is 20 kDa protein based on the human DNA repair protein O^6-alkylguanine-DNA alkyltransferase (hAGT). This enzyme transfers the alkyl group from its substrate O^6-alkylguanine-DNA to an internal cysteine residue. As the labelling is very specific towards the hAGT, but promiscuous towards its substrate, derivatives of the substrate can be

used[105]. The SNAP-tag has successfully been used for the direct immobilization of proteins on benzylguanine-modified surfaces[106–108]. Another AGT-based tag was developed by Johnson *et al.*[103,109], namely the so-called CLIP-tag, which reacts specifically with O^2 benzylcytosine derivatives[110]. The CLIP-tag and SNAP-tag possess orthogonal substrate specificities, meaning that CLIP and SNAP fusion proteins can be used simultaneously and specifically with different molecular probes. Another enzyme tag is the so-called HaloTag, developed by Promega[111]. Even though these self-labeling protein tags are very useful for the modification of proteins, the large size of these tags (~20 kDa for SNAP and CLIP-tag and 33 kDa for the HaloTag) makes them less suited for protein immobilization.

Expressed protein ligation (EPL). This technique is based on self-splicing proteins. In this post-translational process of protein splicing, a precursor protein undergoes a self-catalyzed intramolecular rearrangement. This results in the removal of an internal protein domain (called an intein) and the ligation of the two flanking polypeptides (called the N- and C-exteins)[112]. In EPL, the target protein is fused with an intein, whose catalytic activity will shift the chemical equilibrium from a peptide bond towards a stable C-terminal thioester. The inteins for EPL are mutated in such a way that they are unable to undergo S-N acyl transfer[100]. Often the intein is fused to a chitine-binding domain (CBD), so the expressed fusion protein can be easily isolated using a chitin column. After washing the column, the protein of interest can be cleaved from the intein-CBD fusion by substitution with terminal thiol-containing molecules. When the splicing is done with a thiol-containing peptide or modified cysteine, a rearrangement occurs to form a stable peptide bond. If these molecules also contain a bioorthogonal group, the EPL will result in a protein terminally modified with this bioorthogonal group (shown in Figure 1.4). Both maltose binding protein (MBP) and the enhanced green fluorescent protein (eGFP) have been modified by Lin *et al.*[113], using the commercially available IMPACT (Intein Mediated Purification with an Affinity Chitin-binding Tag) system from New England Biolabs[61]. The drawback of Expressed Protein Ligation is that modifications are restricted to either the N- or C-terminus of the target protein. This makes EPL only applicable for proteins where at least one of the termini is exposed to the surface and is not near, nor part of the active site of the protein. In case the modification interferes with the

activity of the protein or when none of the termini is exposed, other strategies will be required.
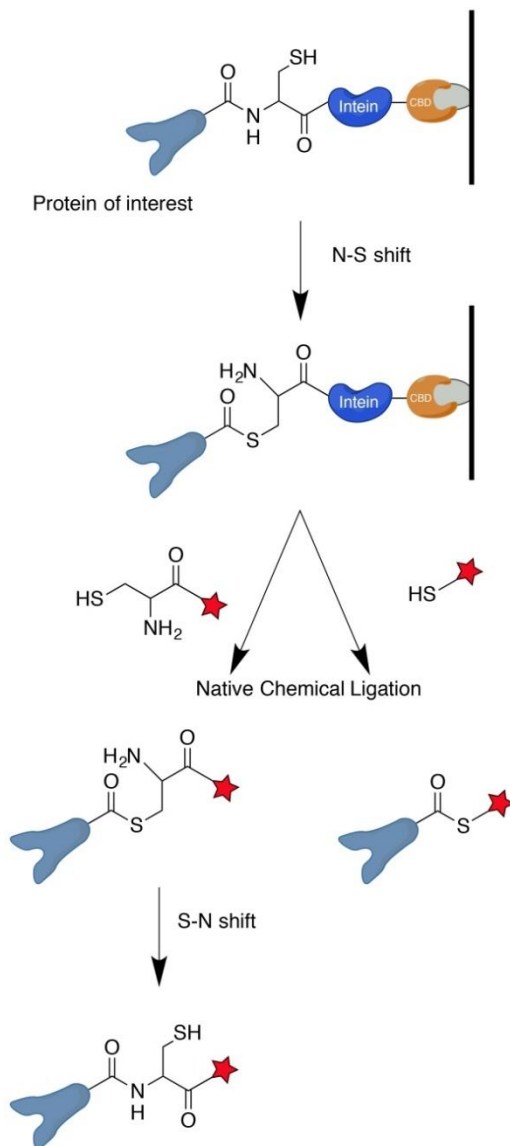
Figure 1.4. Expressed protein ligation (EPL)[3]: In order to facilitate purification on a chitin column, the protein of interest is recombinantly fused to an intein and a chitin-binding domain (CBD). The recombinant protein of interest is selectively cleaved from the immobilized chitin by native chemical ligation (NCL) with a molecule containing a thiol and a bioorthogonal functionality of choice (red star).

1.2.5.2   *In vivo protein modification*

Auxotrophic Expression. In this approach the promiscuity of the wildtype aminoacyl-tRNA synthetases (aaRS) is exploited by replacing one of the twenty natural amino acids with a structural analogue[72,114]. The incorporation of amino acids is controlled by aminoacyl-tRNA synthetases responsible for the coupling of amino acids to their corresponding tRNA. As a consequence of the substrate tolerance of aaRS, a non-natural amino acid analogue can be coupled to the corresponding tRNA. This can only be done if the original amino acid is not present in the growth medium and the host organism is auxotrophic for it (incapable of producing the original amino acid). The amino acid analogue will be incorporated into the protein at specific places where the natural amino acid is genetically encoded[72]. The group of Bertozzi[115] tested two azide-functionalized amino acids, azidoalanine and azidohomoalanine, as *in vitro* substrates for methionine-tRNA synthetase. Both *in vitro* and *in vivo* assays showed azidohomoalanine to be an excellent methionine surrogate. Therefore, auxotrophic expression allows to incorporate functionalities in a protein, orthogonal to the endogenous amino acids. However, the disadvantage it presents is that the incorporation occurs at residue-level rather than at a site-specific level. This means that when a certain amino acid is encoded multiple times in a protein, multiple copies of the analogue will be incorporated. Moreover, when the incorporation efficiency is below 100%, it will result in a mixture with heterogeneously modified proteins. Another drawback of this method is that it does not allow continuous cell growth[9].

Nonsense suppression. The most elegant and controllable way to modify proteins in a site-specific way is by means of nonsense suppression. With this methodology bioorthogonal functionalities are incorporated during the actual protein translation. Different components of the translational machinery are needed in order to genetically encode a non-natural amino acid: i) a unique codon and the corresponding tRNA recognizing this codon, ii) a non-natural amino acid and, iii) a new aminoacyl-tRNA synthetase (aaRS) able to recognize and load this non-natural amino acid to the corresponding tRNA. It is important that the non-natural amino acid is metabolically stable and able to be imported by the host cell. Also this non-natural amino acid should not be a substrate for endogenous synthetases. Another condition is that the aaRS/tRNA pair does not cross react with the endogenous translational

system of the host organism. For this an aaRS/tRNA pair from another organism can be introduced, e.g. the *E. coli* EcTyrRS/tRNA^{Tyr} pair can be used in *S. cerevisiae*. Finally, a unique codon must be used, that cannot be recognized by any of the endogenous tRNA. A common method is the so-called nonsense suppression that uses a stop (nonsense) codon and a suppressor tRNA that recognizes this codon. Figure 1.5 gives an overview on the methodology for nonsense suppression. This method is discussed in detail in section 2.2.
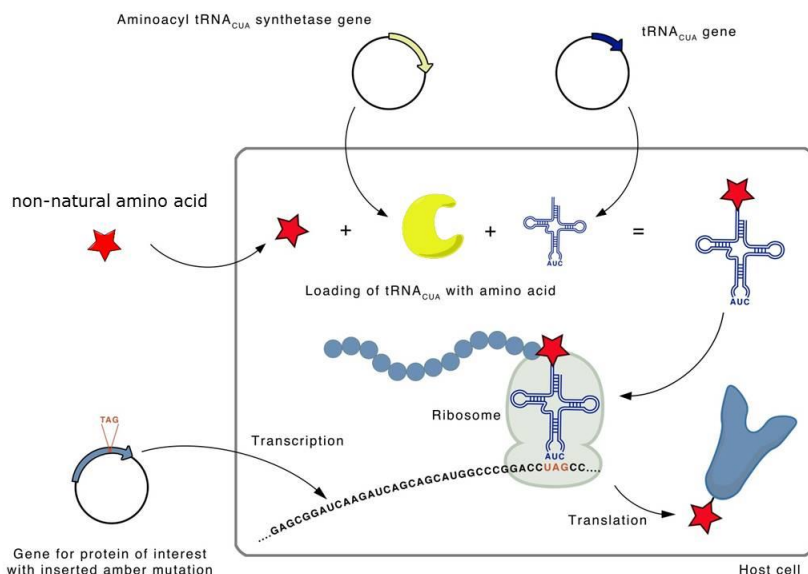
Figure 1.5. Schematic representation of the site-specific incorporation of unnatural amino acids into proteins by amber suppression(adapted from Steen Redeker *et al.*[3]). In this method, the gene for the protein of interest is mutated with an amber codon (UAG) at the desired site of modification. A mutant aminoacyl-tRNA$_{CUA}$ synthetase recognizes the unnatural amino acid bearing the desired functionality and loads it to the corresponding tRNA$_{CUA}$. This leads to the expression of the protein of interest with the desired functionality at the genetically encoded site.

1.3 Aims and structure of the thesis

1.3.1 Aims

Protein-based biotechnological applications share a strong demand for more controlled methodologies to modify and/or immobilize proteins on solid surfaces. All these applications demand a high sensitivity and selectivity combined with a high reproducibility. The strategies available nowadays only allow the formation of either an orientated or a covalent coupling, but not both simultaneously, and therefore do not meet the requirements that current technologies demand.

The aim of this work consists in the evaluation and application of a strategy towards spatially oriented and covalently coupled proteins to solid surfaces (e.g. for miniaturized biosensors).

For this reason a generic *in vivo* method for the site-specific modification of proteins, based on nonsense suppression, will be used in the yeast *Saccharomyces cerevisiae*. A genetically encoded mutant *E. coli* tyrosyl tRNA-synthetase(EcTyRS)/$tRNA_{CUA}^{Tyr}$ pair that incorporates *p*-azidophenylalanine (*p*-azidoPhe) (Figure 1.6) in response to an amber codon (UAG) is used. Therefore, the amber codon will encode a non-natural amino acid. The $EcTyrRS/tRNA_{CUA}^{Tyr}$ pair is orthogonal in *S. cerevisiae,* meaning that the tRNA is not aminoacylated by any of the endogenous synthetases, nor that EcTyrRS charges any of the endogenous tRNA's in yeast.
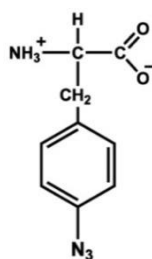
Figure 1.6. The non-natural amino acid *p*-azidophenylalanine.

In order to obtain such a mutant EcTyrRS a library needs to be constructed and screened against the non-natural amino acid. The EcTyrRS was the first aaRS to be modified and introduced in *S. cerevisiae* for the amber suppression approach and it has been extensively described[116,117]. All protein-engineering studies accomplished on EcTyrRS in the past were based on the crystal structure of the TyrRS

of *Bacillus stearothermophilus* and not on the structure of EcTyrRS, resulting in mutant EcTyrRS capable of incorporating both *p*-azidophenylalanine as well as *p*-propargyloxyphenylalanine. It was opted to construct and screen an own library of mutant EcTyrRS, based on the crystal structure of EcTyrRS, not only to acquire more selective EcTyrRS mutants, but also to apply it later on other aaRSs.

Although at the moment the focus is on the site-specific modification of proteins for improved biosensor applications, this method is also beneficial for improved therapeutics and numerous other applications as mentioned in section 1.1. The protein model system employed is the nanobody BCII10, which is a stable nanobody (Nb), easy to express in yeast.

The goal of this work is to use an *in vivo* method for the site-specific modification of proteins with a "click" functionality, allowing an oriented and covalent coupling to a complementary functionalized surface for the design of innovative bio-active materials.

In order to realize this, a number of major steps needs to be reached, as illustrated in Figure 1.7:

- Development of the DNA constructs (plasmids) needed for the production of azide-modified nanobodies as well as for the screening of a library of mutant EcTyrRSs, in order to obtain more specific mutants.
- Construction and screening of the library of mutant EcTyrRSs, for a mutant that specifically aminoacylates the tRNA$_{CUA}^{Tyr}$ with *p*-azidoPhe.
- The strategic introduction of the amber codon in the NbBCII10 gene and the synthesis of the modified NbBCII10 in *S. cerevisiae* (MaV203).
- Covalent coupling of the modified nanobody to a complementary functionalized surface by means of 'click' chemistry.

For the sake of clarity this overview figure (Figure 1.7) will be used throughout the thesis to highlight the specific steps discussed in the relevant chapters. Besides this, it was noticed during the experiments that a small fraction of MaV203 cells are capable of growing on single drop-out media SD-URA (medium lacking uracil), even though the URA3 gene is supposed to be actively repressed, in the absence of GAL4. We were the first to notice this spontaneous URA3 derepression, which can

have a negative impact on techniques that require a strict URA⁻ phenotype (e.g. the screening of a library of mutant EcTyrRSs, reverse yeast-two-hybrid). Therefore, a series of experiments needed to be conducted in order to explain this phenomenon.
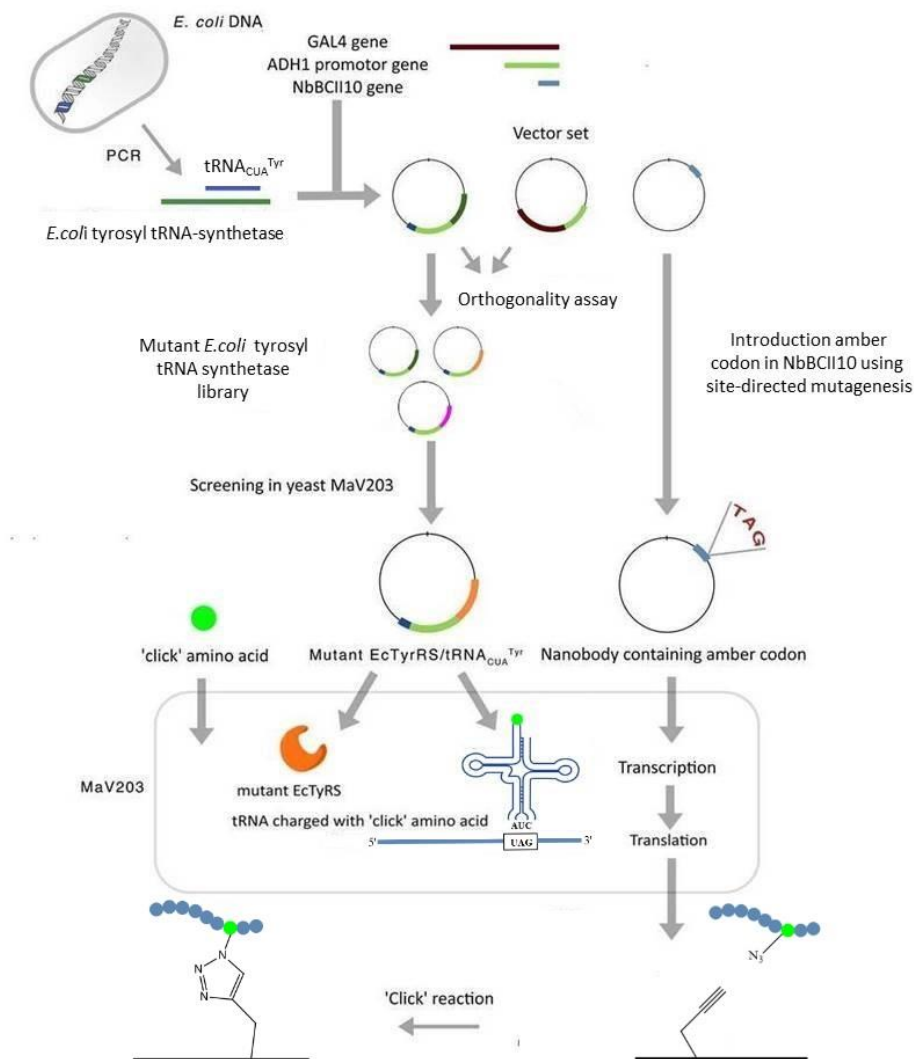
Figure 1.7. Schematic overview of the aims of this thesis.

1.3.2 Structure of the thesis

After a short introduction in chapter one, chapter two discusses both the protein model system (nanobodies) and the host organism (*S. cerevisiae*), as well as their advantages and possible applications. Also, the nonsense suppression approach to site-specifically introduce a functionality into nanobodies is described in detail.

Chapter three describes the components and the construction of the DNA constructs needed for the production of azide-modified nanobodies as well as the DNA constructs required for the proof of principle experiment for the use of diploid cells for the screening of libraries of mutant EcTyrRS.

Chapter four contains a series of experiments performed in order to explain the spontaneous and unwanted expression of the URA3 gene in MaV203 cells. It proposes a solution for this unwanted URA3 expression that could have an impact on techniques that require MaV203 cells with a strict URA⁻ phenotype (e.g. the screening of a library of mutant aaRS and yeast-two-hybrid).

In chapter five both the benefits as well as the difficulties of constructing a library of mutant EcTyrRS are discussed. Different strategies for library construction are given.

In chapter six the synthesis of the site-specifically functionalized nanobody BCII10 in *S. cerevisiae* will be discussed. The expression of full-length functionalized green fluorescent protein (GFP) is used as a proof of principle, to investigate whether or not the $EcTyrRS/tRNA_{CUA}^{Tyr}$ can be used to produce modified proteins.

Finally, chapter 7 contains a general discussion and conclusion regarding the site-specific functionalization of proteins in *S. cerevisiae* by means of nonsense suppression.

1.4 References

1. Campàs, M. *et al.*, A review of the use of genetically engineered enzymes in electrochemical biosensors, *Semin. Cell Dev. Biol.* **20**, 3–9 (2009).

2. Chambers, J. P. *et al.*, Biosensor recognition elements, *Curr. Issues Mol. Biol.,* **10**, 1–12 (2008).

3. Steen Redeker, E. *et al.,* Protein engineering for directed immobilization, *Bioconjug. Chem.* **24**, 1761–1777 (2013).

4. Xie, J. *et al.*, A genetically encoded bidentate, metal-binding amino acid, *Angew. Chemie.* **119**, 9399–9402 (2007).

5. Hyun, S. L. *et al.*, Biosynthesis of a site-specific DNA cleaving protein, *J. Am. Chem. Soc.* **130**, 13194–13195 (2008).

6. Xie, J. *et al.*, A genetically encoded metabolically stable analogue of phosphotyrosine in *Escherichia coli*, *ACS Chem. Biol.* **2**, 474–478 (2007).

7. Grünewald, J. *et al.,* Immunochemical termination of self-tolerance, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11276–11280 (2008).

8. Grünewald, J. *et al.,* Mechanistic studies of the immunochemical termination of self-tolerance with unnatural amino acids, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4337–4342 (2009).

9. Liu, C. C. *et al.*, Adding new chemistries to the genetic code, *Annu. Rev. Biochem.* **79**, 413–44 (2010).

10. Roberts, M. J. *et al.*, J. M., Chemistry for peptide and protein PEGylation, *Adv. Drug Deliv. Rev.* **64**, 116–127 (2012).

11. Harmsen, M. M. *et al.,* Passive immunization of guinea pigs with llama single-domain antibody fragments against foot-and-mouth disease. *Vet. Microbiol.* **120,** 193–206 (2007).

12. Vaneycken, I. *et al.,* Preclinical screening of anti-HER2 nanobodies for molecular imaging of breast cancer, *FASEB J.* **25,** 2433–2446 (2011).

13. Huang, L. *et al.,* SPECT imaging with ^{99m}Tc-labeled EGFR-specific nanobody for *in vivo* monitoring of EGFR expression, *Mol. Imaging Biol.* **10,** 167–175 (2008).

14. De Groeve, K. *et al.,* Nanobodies as tools for *in vivo* imaging of specific immune cell types, *J. Nucl. Med.* **51,** 782–9 (2010).

15. Nayak, T. R. *et al.*, *Cancer Theranostics*. *Cancer Theranostics,* 347-361, (2014).

16. Cellitti, S. E. *et al., In vivo* incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy, *J. Am. Chem. Soc.* **130,** 9268–9281 (2008).

17. Schultz, K. C. *et al.,* A genetically encoded infrared probe, *J. Am. Chem. Soc.* **128,** 13984–13985 (2006).

18. Chen, S. *et al.*, A., An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in *Saccharomyces cerevisiae*, *J. Mol. Biol.* **371,** 112–22 (2007).

19. Xie, J. *et al.,* The site-specific incorporation of *p*-iodo-L-phenylalanine into proteins for structure determination, *Nat. Biotechnol.* **22,** 1297–1301 (2004).

20. Sakamoto, K. *et al.,* Genetic encoding of 3-iodo-I-tyrosine in *Escherichia coli* for single-wavelength anomalous dispersion phasing in protein crystallography, *Structure.* **17,** 335–344 (2009).

21. Lee, H. S. *et al.*, Genetic incorporation of a metal-ion chelating amino acid into proteins as a biophysical probe, *J. Am. Chem. Soc.* **131,** 2481–2483 (2009).

22. Wang, Q. *et al.*, Expanding the genetic code for biological studies, *Chem. Biol.* **16,** 323–36 (2009).

23. Wang, J. *et al.*, A genetically encoded fluorescent amino acid, *J. Am. Chem. Soc.* **128,** 8738–9 (2006).

24. Hyun, S. L. *et al.*, Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in *Saccharomyces cerevisiae*, *J. Am. Chem. Soc.* **131,** 12921–12923 (2009).

25. Mills, J. H. *et al.*, A genetically encoded direct sensor of antibody-antigen interactions, *ChemBioChem* **10,** 2162–2164 (2009).

26. Peters, F. B. *et al.*, Photocleavage of the polypeptide backbone by 2-nitrophenylalanine, *Chem. Biol.* **16,** 148–152 (2009).

27. Bose, M. *et al.*, The incorporation of a photoisomerizable amino acid into proteins in *E. coli* the incorporation of a photoisomerizable amino acid into proteins in *E . coli*. **128,** 388–389 (2006).

28. Lee, H. S. *et al.*, Protein-DNA photo-crosslinking with a genetically encoded benzophenone-containing amino acid, *Bioorganic Med. Chem. Lett.* **19,** 5222–5224 (2009).

29. Lemke, E. A. *et al.*, Control of protein phosphorylation with a genetically encoded photocaged amino acid, *Nat. Chem. Biol.* **3,** 769–772 (2007).

30. Hino, N. *et al.,* Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid, *Nat. Methods* **2,** 201–206 (2005).

31. Camarero, J. A. *et al.*, Recent developments in the site-specific immobilization of proteins onto solid supports, *Biopolym. - Pept. Sci. Sect.* **90,** 450–458 (2008).

32. Stephanopoulos, N. *et al.*, Choosing an effective protein bioconjugation strategy, *Nat. Chem. Biol.* **7,** 876–884 (2011).

33. Camarero, J. A. *et al.*, New developments for the site-specific attachment of protein to surfaces, *Biophys. Rev. Lett.* (2005).

34. Rusmini, F. *et al.*, Protein immobilization strategies for protein biochips, *Biomacromolecules* **8,** 1775–89 (2007).

35. Spahn, C. *et al.*, Enzyme immobilization in Biotechnology, *Recent Patents Eng.* **2,** 195–200 (2008).

36. Nakanishi, K. *et al.*, Recent advances in controlled immobilization of proteins onto the surface of the solid substrate and its possible application to proteomics, *Curr. Proteomics* **5,** 161–175 (2008).

37. Brady, D. *et al.*, Advances in enzyme immobilisation, *Biotechnol. Lett.* **31,** 1639–1650 (2009).

38. Gauthier, M. A *et al.*, Peptide/protein-polymer conjugates: synthetic strategies and design concepts, *Chem. Commun. (Camb).* 2591–2611 (2008).

39. Jonkheijm, P. *et al.*, Chemical strategies for generating protein biochips, *Angew. Chemie - Int. Ed.* **47,** 9618–9647 (2008).

40. Mateo, C. *et al.,* Some special features of glyoxyl supports to

immobilize proteins, *Enzyme Microb. Technol.* **37,** 456–462 (2005).

41. Rao, S. V. *et al.*, Oriented immobilization of proteins, *Mikrochim. Acta* **128,** 127–143 (1998).

42. Wang, B. *et al.*, High-resolution single-molecule recognition imaging of the molecular details of ricin-aptamer interaction, *J. Phys. Chem. B* **116,** 5316–5322 (2012).

43. Pei, Z. *et al.,* Optimizing immobilization on two-dimensional carboxyl surface: pH dependence of antibody orientation and antigen binding capacity, *Anal. Biochem.* **398,** 161–168 (2010).

44. Chen, R. J. *et al.*, Noncovalent sidewall functionalization of carbon nanotubes for protein immobilization, *J. Am. Chem. Soc.* **123,** 3838–3839 (2001).

45. Baslé, E. *et al.*, Protein chemical modification on endogenous amino acids, *Chem. Biol.* **17,** 213–227 (2010).

46. Kim, Y. *et al.,* Efficient site-specific labeling of proteins via cysteines, *Bioconjug. Chem.* **19,** 786–791 (2008).

47. Ménard, A. *et al.*, Site-specific fluorescent labeling and oriented immobilization of a triple mutant of CYP3a4 via C64, *Bioconjug. Chem.* **23,** 826–836 (2012).

48. Masri, M. S. *et al.*, Protein reactions with methyl and ethyl vinyl sulfones, *J. Protein Chem.* **7,** 49–54 (1988).

49. Porath, J. *et al.*, Metal chelate affinity chromatography, a new approach to protein fractionation, *Nature* **258,** 598–599 (1975).

50. Ganesana, M. *et al.*, Site-specific immobilization of a $(His)_6$-tagged acetylcholinesterase on nickel nanoparticles for highly sensitive toxicity biosensors, *Biosens. Bioelectron.* **30,** 43–48 (2011).

51. Oshige, M. *et al.*, Immobilization of His-tagged proteins on various solid surfaces using NTA-modified chitosan, **2013,** 6–10 (2013).

52. Hopp, T. P. *et al.,* A short polypeptide marker sequence useful for recombinant protein identification and purification, *Bio/Technology* **6,** 1204–1210 (1988).

53. Knappik, A. *et al.*, An improved affinity tag based on the FLAG peptide for the detection and purification of recombinant antibody

Chapter 1: Introduction

fragments, *Biotechniques* **17,** 754–761 (1994).

54. Evan, G. I. *et al.*, Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product, *Mol. Cell. Biol.* **5,** 3610–3616 (1985).

55. Sudheer, P. D. *et al.*, Cyclization tag for the detection and facile purification of backbone-cyclized proteins, *Anal. Biochem.* **436,** 137–141 (2013).

56. Wingren, C. *et al.,* Microarrays based on affinity-tagged single-chain Fv antibodies: Sensitive detection of analyte in complex proteomes, *Proteomics* **5,** 1281–1291 (2005).

57. Holland-Nell, K. *et al.*, Specifically immobilised Aldo/Keto reductase AKR1A1 shows a dramatic increase in activity relative to the randomly immobilised enzyme, *ChemBioChem* **8,** 1071–1076 (2007).

58. Yu, C.-C. *et al.,* Site-specific immobilization of enzymes on magnetic nanoparticles and their use in organic synthesis, *Bioconjug. Chem.* **23,** 714–24 (2012).

59. Fruk, L. *et al.,* DNA-directed immobilization of horseradish peroxidase-DNA conjugates on microelectrode arrays: Towards electrochemical screening of enzyme libraries, *Chem. - A Eur. J.* **13,** 5223–5231 (2007).

60. Niemeyer, C. M. *et al.* Semisynthetic DNA-protein conjugates for biosensing and nanofabrication., *Angew. Chemie - Int. Ed.* **49,** 1200–1216 (2010).

61. Chong, S. *et al.,* Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element, *Gene* **192,** 271–281 (1997).

62. Oh, B.-K. *et al.,* Surface plasmon resonance immunosensor for the detection of *Salmonella typhimurium*, *Biosens. Bioelectron.* **19,** 1497–1504 (2004).

63. Berrade, L. *et al.*, Protein microarrays: novel developments and applications, *Pharm. Res.* **28,** 1480–99 (2011).

64. Benešová, E. *et al.*, Affinity Interactions as a tool for protein immobilization, *Affin. Chromatogr.* 29–46 (2012).

65. Mateo, C. *et al.*, Multifunctional epoxy supports: a new tool to improve the covalent immobilization of proteins. The promotion of

physical adsorptions of proteins on the supports before their covalent linkage, *Biomacromolecules* **1,** 739–745 (2000).

66.   Kolb, H. C. *et al.*, Click Chemistry : diverse chemical function from a few good reactions, *Angew. Chem. Int. Ed.* **40,** 2004-2021(2001).

67.   Best, M. D. *et al.* Click chemistry and bioorthogonal reactions: unprecedented selectivity in the labeling of biological molecules, *Biochemistry.* **48,** 6571–84 (2009).

68.   Lutz, J.-F. *et al.*, Efficient construction of therapeutics, bioconjugates, biomaterials and bioactive surfaces using azide-alkyne 'click' chemistry, *Adv. Drug Deliv. Rev.* **60,** 958–970 (2008).

69.   Huisgen, R. *et al.*, 1,3-Dipolar Cycloadditions. Past and Future, *Angew. Chemie Int. Ed. English* **2,** 565–598 (1963).

70.   Huisgen, R. *et al.*, Kinetics and mechanism of 1,3-dipolar cycloadditions, *Angew. Chemie Int. Ed. English* **2,** 633–645 (1963).

71.   Kalia, J. *et al.*, Advances in bioconjugation, *Curr. Org. Chem.* **14,** 138–147 (2010).

72.   Van Hest, J. C. M. *et al.*, Protein modification by strain-promoted alkyne-azide cycloaddition, *ChemBioChem* **12,** 1309–1312 (2011).

73.   Baskin, J. M. *et al.,* Copper-free click chemistry for dynamic *in vivo* imaging, *Proc. Natl. Acad. Sci. U.S.A.* **104,** 16793–16797 (2007).

74.   Agard, N. J. *et al.*, A strain-promoted [3+2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. 15046–15047 (2004).

75.   Devaraj, N. K. *et al.*, Copper catalyzed azide-alkyne cycloadditions on solid surfaces: Applications and future directions, *QSAR Comb. Sci.* **26,** 1253–1260 (2007).

76.   Deiters, A. *et al.*, *In vivo* incorporation of an alkyne into proteins in *Escherichia coli*, *Bioorg. Med. Chem. Lett.* **15,** 1521–1524 (2005).

77.   Chin, J. W. *et al.,* Addition of *p*-azido-L-phenylalanine to the genetic code of *Escherichia coli*, *J. Am. Chem. Soc.* **124,** 9026–7

(2002).

78.   Young, T. S. *et al.,* Expanding the genetic repertoire of the
       methylotrophic yeast *Pichia pastoris*. *Biochemistry* **48,** 2643–2653
       (2009).

79.   Liu, W. *et al.*, Genetic incorporation of unnatural amino acids into
       proteins in mammalian cells, *Nat. Methods* **4,** 239–244 (2007).

80.   Saxon, E. *et al.,* Cell surface engineering by a modified Staudinger
       reaction. *Science* **287,** 2007–2010 (200).

81.   De Araújo, A. D. *et al.,* Diels-Alder ligation and surface
       immobilization of proteins, *Angew. Chemie - Int. Ed.* **45,** 296–301
       (2005).

82.   Rusmini, F. *et al.*, Protein immobilization strategies for protein
       biochips, *Biomacromolecules* **8,** 1775–1789 (2007).

83.   Jonkheijm, P. *et al.,* Photochemical surface patterning by the
       thiol-ene reaction, *Angew. Chemie - Int. Ed.* **47,** 4421–4424
       (2008).

84.   Köhn, M. *et al.,* Immobilization strategies for small molecule,
       peptide and protein microarrays, *J. Pept. Sci.* **15,** 393–7 (2009).

85.   Chen, Y. X. *et al.*, Bioorthogonal chemistry for site-specific
       labeling and surface immobilization of proteins, *Acc. Chem. Res.*
       44, 762–773 (2011).

86.   Milles, S. *et al.,* Click strategies for single molecule protein
       fluorescence, *Synthesis (Stuttg).* 1–13 (2012).

87.   Dirksen, A. *et al.*, Rapid oxime and hydrazone ligations with
       aromatic aldehydes for biomolecular labeling, *Bioconjug. Chem.*
       19, 2543–2548 (2008).

88.   Wang Aijun, G. K. *et al.*, Protein engineering with non-natural
       amino acids, *Methods Mol. Biol.* **77,** 325–354 (1998).

89.   Lempens, E. H. M. *et al.*, Efficient and chemoselective surface
       immobilization of proteins by using aniline-catalyzed oxime
       chemistry, *ChemBioChem* **10,** 658–662 (2009).

90.   De Graaf, A. J. *et al.*, Nonnatural amino acids for site-specific
       protein conjugation, *Bioconjug. Chem.* **20,** 1281–1295 (2009).

91.   Hutchins, B. M. *et al.,* Site-specific coupling and sterically
       controlled formation of multimeric antibody Fab fragments with

unnatural amino acids, *J. Mol. Biol.* **406,** 595–603 (2011).

92. Chen, I. *et al.*, Site-specific labeling of proteins with small molecules in live cells, *Curr. Opin. Biotechnol.* **16,** 35–40 (2005).

93. Prescher, J. A *et al.*, Chemistry in living systems, *Nat. Chem. Biol.* **1,** 13–21 (2005).

94. Chen, I. *et al.*, Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase, *Nat. Methods* **2,** 99–104 (2005).

95. Slavoff, S. A. *et al.*, Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species, *J. Am. Chem. Soc.* **130,** 1160–1162 (2008).

96. Proft, T. *et al.,* Sortase-mediated protein ligation: An emerging biotechnology tool for protein modification and immobilisation, *Biotechnol. Lett.* **32,** 1–10 (2009).

97. Chan, L. *et al.,* Covalent attachment of proteins to solid supports and surfaces via sortase-mediated ligation, *PLoS One* **2,** 1–5 (2007).

98. Jiang, R. *et al.*, End-point immobilization of recombinant thrombomodulin via sortase-mediated ligation, *Bioconjug. Chem.* **23,** 643–649 (2012).

99. Rabuka, D. *et al.*, Chemoenzymatic methods for site-specific protein modification, *Curr. Opin. Chem. Biol.* **14,** 790–6 (2010).

100. Sletten, E. M. *et al.*, Bioorthogonal chemistry: fishing for selectivity in a sea of functionality, *Angew. Chem. Int. Ed. Engl.* **48,** 6974–98 (2009).

101. Hinner, M. J. *et al.*, How to obtain labeled proteins and what to do with them, *Curr. Opin. Biotechnol.* **21,** 766–776 (2010).

102. Milles, S. *et al.*, What precision-protein-tuning and nano-resolved single molecule sciences can do for each other, *BioEssays* **35,** 65–74 (2013).

103. Keppler, A. *et al.,* A general method for the covalent labeling of fusion proteins with small molecules *in vivo*, *Nat. Biotechnol.* **21,** 86–89 (2003).

104. Engin, S. *et al.*, SNAP-tag as a tool for surface immobilization, *Curr. Pharm. Des.* **19,** 5443–8 (2013).

105. Juillerat, A. *et al.,* Engineering substrate specificity of O6-alkylguanine-DNA alkyltransferase for specific protein labeling in living cells, *ChemBioChem* **6,** 1263–1269 (2005).

106. Engin, S. *et al.,* Benzylguanine thiol self-assembled monolayers for the immobilization of snap-tag proteins on microcontact-printed surface structures, *Langmuir* **26,** 6097–6101 (2010).

107. Kwok, C. W. *et al.* Selective immobilization of Sonic hedgehog on benzylguanine terminated patterned self-assembled monolayers, *Biomaterials* **32,** 6719–6728 (2011).

108. Recker, T. *et al,* Directed covalent immobilization of fluorescently labeled cytokines, *Bioconjug. Chem.* **22,** 1210–1220 (2011).

109. Johnsson, K. *et al.*, Switchable fluorophores for protein labeling in living cells, 768–774 (2011).

110. Gautier, A. *et al.,* An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.* **15,** 128–136 (2008).

111. Los, G. V. *et al.,* HaloTag: A novel protein labeling technology for cell imaging and protein analysis, *ACS Chem. Biol.* **3,** 373–382 (2008).

112. Muralidharan, V. *et al.*, Protein ligation: an enabling technology for the biophysical analysis of proteins, *Nat. Methods* **3,** 429–438 (2006).

113. Lin, P. C. *et al.,* Site-specific protein modification through CuI-catalyzed 1,2,3-triazole formation and its implementation in protein microarray fabrication, *Angew. Chemie - Int. Ed.* **45,** 4286–4290 (2006).

114. Van Hest, J. C. M. *et al.*, Efficient incorporation of unsaturated methionine analogues into proteins *in vivo*, *J. Am. Chem. Soc.* **122,** 1282–1288 (2000).

115. Kiick, K. L. *et al.*, Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation, *Proc. Natl. Acad. Sci. U. S. A.* **99,** 19–24 (2002).

116. Chin, J. W. *et al.,* Progress Toward an Expanded Eukaryotic Genetic Code, *Chem Biol.* **10,** 511–519 (2003).

117. Cropp, T. A. *et al.*, Reprogramming the amino-acid substrate specificity of orthogonal aminoacyl-tRNA synthetases to expand the genetic code of eukaryotic cells, *Nat.Protoc.* 2590–600 (2007).

Chapter 2

Nanobodies and amber suppression in *S. cerevisiae*

In this chapter both the protein model system (nanobodies) and the host organism (*S. cerevisiae*)*,* as well as their advantages and possible applications will be discussed. Also, the nonsense suppression approach used to site-specifically introduce a functionality into nanobodies will be explained in detail.

2.1 Nanobodies

Nanobodies were discovered in the eighties by professor Raymond Hamers of the Vrije Universiteit Brussel (VUB)[1]. The antibodies of mammals consist of two identical "heavy chains" (H-chains) and two identical "light chains" (L-chains). The Immunoglobulin G (IgG) antibodies of *Camelidae* are an exception. Besides the conventional antibodies, the serum of *Camelidae* also contains a large fraction of so-called 'heavy chain antibodies' (HCAbs)[1–3]. These HCAbs are structurally different from conventional antibodies (Figure 2.1) *i.e.* they contain a single variable domain (VHH), a hinge region and two constant domains (CH2 and CH3).

The two constant domains are homologous to the Fc domains (fragment crystallizable domain) of classical antibodies[4]. Although the first domain of the constant region (CH1) is transcribed from the genome, it is spliced out during mRNA processing. Therefore, HCAbs lack the CH1 domain, explaining the absence of the light chain, since CH1 is the anchoring point for the constant domain of the light chain in classical antibodies[5,6]. The antigen-binding domain in HCAbs is reduced to one single variable domain, the VHH. In addition, the isolated and cloned VHH domain, also called nanobody (Nb), is a very stable polypeptide and it is one of the smallest intact antigen-binding fragments [7–9].
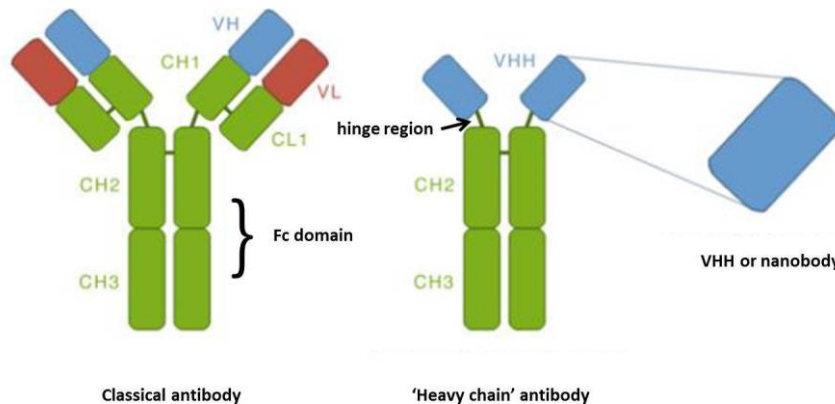
Figure 2.1. Comparison of a classical antibody, a 'heavy chain' antibody (HCAb) and a VHH or nanobody.

Nanobodies contain the full antigen-binding capacity of the original HCAbs structure and have evolved to be fully functional without the presence of light chains[6,10]. In general, nanobodies contain four framework regions that form the core structure of the immunoglobulin domain and three complementarity-determining regions (CDRs), involved in the antigen binding (Figure 2.2).

Furthermore, there are some structural differences between the variable domain of a classical antibody (VH) and the variable domain of an HCAb (VHH or nanobody). For example, the antigen-binding loops of nanobodies have a larger structural repertoire than observed in conventional VHs[11]. Especially the CDR3 region of nanobodies is on average longer than that of VHs. Besides, nanobody sequences also contain amino acid substitutions in the framework regions. These substitutions are not present in VH domains that pair with the variable domain of the light chain (VL). The substitutions observed in nanobodies might compensate for the absence of the VL domain in the antigen-binding site and might accommodate more flexible loops than VH. The framework-2 region also contains some hydrophobic to hydrophilic amino acid substitutions, improving the solubility of nanobodies[5,6,12,13]. For comparison, the structures of the VH and VHH are shown in Figure 2.2.
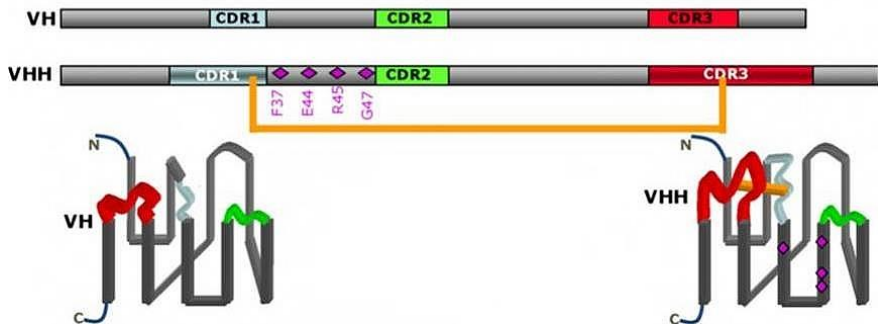
Figure 2.2. Representation of the variable domain structure of a classical antibody (VH) compared with a nanobody (VHH) [2]. In this figure, some structural differences between the VH and the VHH are observed. In particular, the CDR3 regions of nanobodies are on average longer than those of VH. Nanobody sequences also contain amino acid substitutions in the framework regions. The internal disulfide bridge is depicted in orange.

Nanobodies share features with classical antibodies, such as: i) high affinity and selectivity for their targets and, ii) low inherent toxicity. In addition, nanobodies possess extra beneficial characteristics that normal antibodies lack:

- A small molecular weight (~15 kDa[13,14]) compared to classical antibodies (~150 kDa).

- The ability to recognize small and hidden epitopes or antigenic sites (e.g. active sites[15,16] and conserved cryptic epitopes[17]). This is possible due to the small size and extended CDR loop of nanobodies, which allows them to penetrate into the epitope sites[16,18].

- Nanobodies do not undergo spontaneous dimerization, as can occur for normal antibodies (e.g. single-chain variable fragments scFv turn into scFv$_2$).

- Soluble in aqueous solutions, provided by the substitution of the original hydrophobic amino acids for hydrophilic ones in the framework-2 region.

- High stability under harsh conditions, which originates from the nanobodies' efficient refolding after thermal or chemical denaturation[19–21]. Some examples are:

    - Stable to heat exposure. Some nanobodies have shown to remain functional at 90 °C or after incubation at high temperatures[19,22].

- ▪ Long shelf-life, useful for biosensors and microarrays. Literature reports that nanobodies retain > 80% of their binding activity after 1 week of incubation at 37 °C[14].

- ▪ Stable in contact with chaotropic agents, proteases and extreme pH conditions, allowing their use for oral administration treatments[6,13].

In addition, nanobodies are encoded by a single gene, requiring no splicing nor post-translational modifications, which makes them very easy to manipulate[23]. It has been shown that nanobodies can efficiently be expressed in micro-organisms. When expressed in shake-flask cultures, levels up to 20 mg/l are obtained in *E. coli*[24]. In the yeast *S. cerevisiae* levels > 100 mg/l were obtained from shake-flask cultures[25,26] and > 1 g/l from a 10 l fed-batch fermentation. When scaled up, yields of > 1 kg of nanobodies have been obtained in a 15 m^3 fermentation tank[5,13].

2.1.1 Applications of nanobodies

As stated above, the numerous advantages of nanobodies portrays them as suitable candidates for applications ranging from capturing reagents in immunoaffinity purification[27], biosensors[28], directed treatment of cancer[10] and even for the prevention of dandruff in shampoos[29]. In this context, some of their most relevant applications are presented in detail.

Therapeutics and drug delivery. Nanobodies have been reported to present a high homology with human VH frameworks (approximately 90%). However, for use as therapeutics, nanobodies can be further humanized with 95-99% homology by means of a few amino acid substitutions in the framework region[13].

Due to the high stability of nanobodies at extreme pH conditions as well as at high chaotropic agent concentrations, they are perfect candidates for oral immunotherapy[20,30]. Besides, the small size of nanobodies (~15 kDa) enables a fast tissue penetration and allows them to pass the renal filter (cutoff ~60 kDa), resulting in a fast blood clearance. This is an advantage for therapeutic applications where the nanobodies are coupled to toxic substances or when used for *in vivo* imaging.

Furthermore, nanobodies can be used either to target toxic enzymes or to block certain molecular interactions. For example, Baral *et al.*[31] successfully treated sleeping sickness by using a nanobody fused to the apolipoprotein L-1 enzyme. This dimeric nanobody selectively binds to a trypanosome coat protein, which leads to trypanosome lysis. Moreover, a nanobody has been genetically fused to β-lactamase and directed against a carcinoembryonic antigen typically used to target tumor cells. After injecting a prodrug, β-lactamase converts it into a toxic drug in the vicinity of the tumor cells, inducing cellular death[10].

Nanobody therapies, based on the blocking of molecular interactions, have also been developed for the treatment of cancers and inflammatory diseases, such as rheumatoid arthritis[32–34]. For these applications, the fact that nanobodies can be excreted in about two hours after administration (short serum half-life), might be a disadvantage[10,35]. The serum half-life of nanobodies can be prolonged in two ways. One method is to target the nanobody to long-lived serum proteins (e.g. albumin or immunoglobulin), using bispecific nanobodies that recognize these serum proteins together with the therapeutic target. This can result in serum half-lives that are equal to immunoglobulin (nine days) or albumin (two days)[32,34]. Another common approach is the chemical functionalization of the nanobody with polyethylene glycol (PEG), thus increasing the serum half-life[36].

Potential immunogenicity of nanobodies could compromise their therapeutic use, especially in treatments that require periodic inoculations. Currently, no immunogenicity was shown in mice after repeated injections, which was verified by analysis of: i) T cells proliferation, ii) the presence of specific antibodies and, iii) cytokine levels[34,37]. However, studies have shown that some immunogenicity issues might occur un humans[38,39]. In case the immunogenicity of a nanobody presents a risk, technologies exist to reduce the immunogenicity[40].

Protein purification. Because of their monomeric nature, stability and potential immobilization on solid supports, nanobodies are useful as ligands for protein purification. Due to their intrinsic stability a high column regeneration capacity is possible and their small size allows a higher number of paratopes per gram of support material. For example, nanobodies with an anti-human IgG specificity were made for the depletion of IgG's from blood[41]. Besides, VHHs against human serum

albumin, human growth hormone and mouse IgG have been commercialized by BAC BV/Life Technologies (Naarden, the Netherlands). This company also brought a new protein affinity tag of four amino acids (EPEA) to the market (developed by the VUB), the so-called CaptureSelect C-tag and its corresponding anti-CaptureSelect C-tag VHH.

Nanobodies as diagnostic tools. The characteristics that detection probes should meet are: i) stability and high selectivity towards the antigen, ii) high probe accessibility and, iii) cost-effective for large-scale production. Nanobodies fulfill all these characteristics, plus their small size allows a higher probe density. In this context, diagnostic nanobodies have been developed for the *in vitro* detection of *Trypanosoma* parasites[42] or as part of a VHH-based agglutination reagent for HIV diagnosis[43]. Additionally, nanobodies can be used as tracers for non-invasive *in vivo* tumour imaging. A good tracer has a: i) high target specificity, ii) high tumor-to-background signal, iii) high stability, iv) fast clearance of unbound tracer, v) good solubility and, vi) low immunogenicity. For this reason, nanobodies have been tested as tracers, confirming a positive tumor penetration and fast renal clearance of excess tracer, which allowed sensitive imaging of the target within a few hours after injection[44,45].

Furthermore, radiolabeled nanobodies are already in use, for example: i) to discriminate between moderate and high expression of the epidermal growth factor receptor, used for the evaluation of cancer therapy[44], ii) to evaluate the diagnosis of breast cancer in a preclinical setting[45] or, iii) to follow up the status of inflammatory responses[46]. Besides nanobodies labelled with radionuclides, nonradioactive VHH-based probes have been developed for tumor imaging as well. These tracers labelled with near-infrared fluorophores allow faster imaging compared to approved classical antibodies targeting the same antigens[47,48].

2.1.2 The nanobody BCII10

In this study, the nanobody BCII10 (NbBCII10) will be used as a model protein. The NbBCII10 gene, containing a C-terminal His_6-tag, was amplified from the pHEN6-BCII10 vector (kindly provided by Prof. Dr. Muyldermans and Prof. Dr. Devoogdt (Vrije Universiteit Brussel (VUB)/ Vlaams Instituut voor Biotechnologie (VIB)) using PCR. The His_6-tag

allows for the detection of the nanobodies by Western blot using anti-his antibodies, as well as their purification with a Ni-column. Since this nanobody is directed against β-lactamase, it inhibits the resistance that nosocomial strains of bacteria have towards β-lactam antibiotics. The nanobody BCII10 was chosen for two reasons: 1) the plasticity of the BCII10 framework, which allows successful transfer of antigen specificity from donor VHH onto the scaffold. This exchange of antigen specificities is done by means of complementarity determining region (CDR) grafting and, 2) the fact that NbBCII10 is very stable, has good expression levels and stays functional, even in the absence of the conserved disulfide bond in the CDR domain[49].

2.2 Nonsense suppression in prokaryotic and eukaryotic cells

Origin of nonsense suppression. Regulation of gene expression occurs at different levels, one of which happens at the level of termination in protein synthesis. For prokaryotic and eukaryotic cells, the termination of protein synthesis is signaled by the presence of a nonsense (stop) codon at the ribosomal A site. In both prokaryotic and eukaryotic cells, three stop codons exist which are responsible for the termination of protein translation. On the mRNA level, these nonsense or stop codons are UAG (amber), UAA (ochre) and UGA (opal). Protein release factors bind to these codons and mediate the release of the nascent polypeptide chain from the ribosome[50]. Mutations can occur, altering an amino acid encoding codon into a nonsense codon. This results in an early termination and therefore most likely in an inactive protein. For this reason, both prokaryotes and eukaryotes have developed some mechanisms to circumvent this unwanted early termination.

In this work, the focus lies on the so-called nonsense suppressors. Nonsense suppressors are alleles of tRNA genes of which the anticodon is altered to recognize a nonsense codon instead of a natural encoding codon[51] during protein synthesis. This leads to the insertion of an amino acid in response to the termination codon and as a result the early termination is suppressed, allowing the expression of the full-length protein. Because there are three termination codons on the DNA level (TAG, TAA and TGA), nonsense mutations are divided into amber, ochre, or opal mutations (resulting in UAG, UAA or UGA respectively) at the mRNA level. For each of these nonsense codons a suppressor tRNA can be found or made[52,53]. In *E. coli* and other organisms there are often

multiple copies of a gene for a tRNA present in the genome. Therefore a mutation in the anticodon of one copy of a tRNA gene that results in the recognition of a nonsense codon, is not lethal because there are still multiple copies of the tRNA gene able to recognize the conventional codon[51].

Potential use of aminoacyl-tRNA synthetase (aaRS)/tRNA pairs in other organisms. Some of the suppressor tRNA's can be copied, along with their cognate aaRS into other organisms without interfering with the endogenous translational machinery. In this case the aaRS/tRNA pair is called orthogonal. This means that the suppressor tRNA is not aminoacylated by any of the endogenous synthetases, nor that the aaRS charges any of the endogenous tRNA's. Nevertheless, they are capable of working with the endogenous ribosomes.

For example, the first orthogonal aminoacyl-tRNA synthetase (aaRS)/ tRNA pair in *E. coli* was generated from the archaeal bacteria *Methanococcus jannaschii*[54]. Also a number of *E. coli* aaRS/tRNA pairs have been proven to be orthogonal in eukaryotic cells[55,56]. Moreover, it has been demonstrated that the *E. coli* tyrosyl-tRNA synthetase (EcTyrRS)/$tRNA_{CUA}^{Tyr}$ pair is able to suppress amber codons in *S. cerevisiae* by means of tyrosine incorporation. The *E. coli* $tRNA_{CUA}$ is not charged by any of the yeast endogenous aminoacyl tRNA synthetases[57,58] and the EcTyrRS does not charge any yeast tRNA[59,60]. For this reason, this pair is a suitable candidate to be used as an orthogonal pair in *S. cerevisiae.* Yeast combines the ease of microbial growth with a eukaryotic environment, enabling post-translational modifications. Using yeast cells a generic platform can be developed in which more complex proteins, requiring post-translational modifications, can also be functionalized with a non-natural amino acid. An extra benefit is that the translational machinery is highly conserved in eukaryotes, meaning that the orthogonal pair developed in yeast can also be used in higher eukaryotes[61]. The benefits of using yeast for the production of functionalized proteins will be discussed in detail in section 2.3.

Nonsense suppression and the site-specific functionalization of nanobodies. In this work, the focus lies on the amber suppressor *E. coli* tyrosine $tRNA_{CUA}$ ($tRNA_{CUA}^{Tyr}$) containing a mutation in its anticodon (GUA $\rightarrow$ CUA). This tRNA recognizes the amber codon (UAG) and incorporates a tyrosine in the growing peptide chain. Following the idea

that each tRNA is aminoacylated by its corresponding aminoacyl-tRNA synthetase, the *E. coli* tRNA$_{CUA}^{Tyr}$ is aminoacylated by the *E. coli* tyrosyl-tRNA synthetase (EcTyrRS). A mutant EcTyrRS is obtained by the mutation of key amino acids in the binding pocket of EcTyrRS, so that it has affinity for a non-natural amino acid instead of tyrosine. In this work the nanobodies are site-specifically modified with a so-called 'click' functionality. The 'click' functionality is introduced in the nanobody by means of amber suppression, using a mutant EcTyrRS with an affinity for the non-natural amino acid *p*-azidophenylalanine (*p*-azidoPhe).

Consequently, the tRNA$_{CUA}^{Tyr}$ will be charged with a *p*-azidoPhe and incorporate this non-natural amino acid in response to an amber codon[62,63]. The fact that the EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair is orthogonal in eukaryotic cells, allows it to be added to the genetic repertoire of *S. cerevisiae*, thereby encoding for a 21st non-natural amino acid. This can be done since the amber codon is the least used stop codon in yeast and it is not used for the termination of essential genes[64,65]. By means of the site-specific introduction of an amber codon into the target protein (NbBCII10_His$_6$) the modification site can be controlled. The advantage of this method, in comparison with other functionalization techniques, is that the modification can be introduced at any location. Figure 2.3 gives an overview of the amber suppression method.
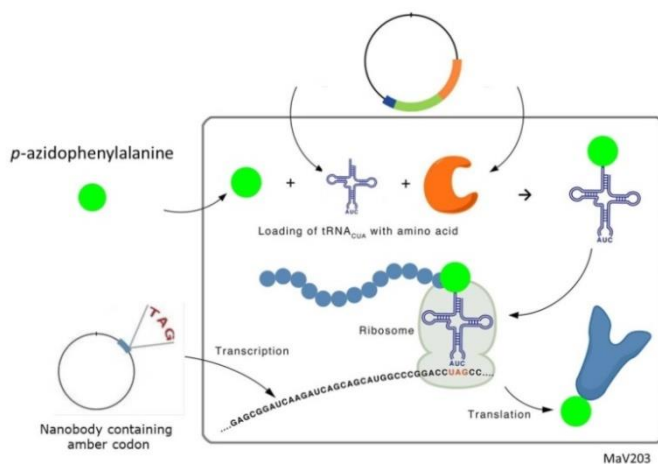
Figure 2.3. Overview of the site-specific incorporation of *p*-azidophenylalanine (*p*-azidoPhe) into NbBCII10 by amber suppression. In this method, the gene for the protein of interest is mutated with an amber codon at the desired site of modification. A mutant tyrosyl-tRNA$_{CUA}^{Tyr}$ synthetase (EcTyrRS) recognizes *p*-azidoPhe and loads it to the corresponding tRNA$_{CUA}^{Tyr}$. This leads to the expression of NbBCII10 with the desired functionality at the genetically encoded site.

To acquire such a mutant EcTyrRS, a library of mutant EcTyrRSs needs to be constructed and screened against the non-natural amino acid of interest. There are many advantages in having such a library, even though the construction and screening process of it is complicated and time-consuming work. The benefits and details of the library construction will be discussed in Chapter 5.

As mentioned above, the construction and screening process of a mutant library is time-consuming and there is no guarantee of success. For this reason, protein expression experiments for the production of modified NbBCII10 were performed in parallel with the library construction. For these expression experiments a specific EcTyrRS mutant, that has been described in literature by Deiters *et al.*[66], was evaluated. Deiters *et al.* managed to construct and screen a library of mutant EcTyrRS for mutants that selectively aminoacylate the *E. coli* $tRNA_{CUA}^{Tyr}$ with *p*-azidoPhe or *p*-propargyloxyPhe, depending on which non-natural amino acid was supplied to the growth medium of yeast.

2.3 Yeast strains for the production of functionalized nanobodies

Since the early eighties, yeast cells are used for the production of intracellular and extracellular proteins of human, animal and plant origin[67]. A wide variety of recombinant products are made in *S. cerevisiae* such as insulin, urate oxidase, hepatitis B surface antigen, glucagons and many more[68]. Nowadays, most of the recombinant proteins approved by the European Medicines Agency (EMEA) and the Food and Drug Agency (FDA) are produced in *S. cerevisiae*[69]. An extra advantage is the fact that *S. cerevisiae* has been recognized as a "generally regarded as safe (GRAS)" strain[70].

Yeast combines the ease of microbial growth and simple manipulation with an eukaryotic environment, enabling the possibility to introduce eukaryotic specific (post-translational) modifications, such as protein folding, disulfide bond formation, proteolytic processing and glycosylation[71,72]. When compared to other eukaryotic systems used for protein expression, e.g. Chinese Hamster Ovary Cells (CHOs), yeast shows more economical benefits.

Because of the eukaryotic environment, yeast cells are capable of performing post-translational modifications, meaning that more complex proteins requiring these modifications can also be functionalized.

Besides, the translational machinery of eukaryotes is very well conserved. Therefore genes involved in the site-specific incorporation of non-natural amino acids in yeast can used in higher eukaryotes. The tRNA aminoacyl synthetase/$tRNA_{CUA}$ pairs that are orthogonal in *S. cerevisiae*, are also orthogonal in higher eukaryotes such as mammalian cells[73].

A major challenge for the application of aaRS/tRNA pairs in mammalian cells is the evolution of mutant orthogonal aaRSs specific for an unnatural amino acid. Because of the low transformation efficiencies of mammalian cells in comparison to *E. coli* or yeast, it is very difficult to generate large mutant aaRS libraries. Besides, the survival-death selection in mammalian cells is not as effective as in *E. coli* or yeast, making it very difficult to select an aaRS out of a large mutant library directly in mammalian cells[73].

To circumvent this issue, a transfer strategy can be used, where the mutant aaRS library is constructed and screened in yeast after which the optimal aaRS is transferred for use in mammalian cells. For example, it has been shown that the *E. coli* tyrosyl tRNA-synthetase can be used with the *Bacillus staerothermophilus* $tRNA_{CUA}^{Tyr}$ to efficiently suppress an amber codon in mammalian cells[74,75]. Wang *et al.*[76] successfully used a mutant EcTyrS, selected in yeast, together with the cognate *E. coli* $tRNA_{CUA}^{Tyr}$ to suppress an amber codon in mammalian cells.

AaRS/$tRNA_{CUA}^{Tyr}$ pairs evolved in *S. cerevisiae* can also be used in other yeast species, such as *Pichia pastoris, Kluyveromyces lactis, Schizoschaccharomyces pombe* and more[70]. This is a great advantage since all the different species of yeast have their own strong points depending on the type of protein and type of expression that is required, although small modifications or extra promotor sequences are sometimes needed to express the tRNA properly[61,74,77].

Yeast strains. Different strains of yeast were used in this work. The first type of yeast strains to be used are the *S. cerevisiae* strains MaV203 (*MATα*) (Life Technologies, Naarden, the Netherlands) and MaV103 (*MATa)* cells (kindly provided by Ir. Dr. Twizere, from the Université de Liège) (Appendix page 185)*.* Since the MaV103 cells are identical to the MaV203 cells, but of the opposite mating type, only MaV203 will be discussed. The MaV203 strain is originally developed for the so-called "reverse two-hybrid system" and "reverse one-hybrid

system"[78]. These applications facilitate the investigation of protein-protein and DNA-protein interactions, as well as the elucidation of structure-function relationships. It can be used to analyze mutations in interaction partners as well as to identify additional trans-acting factors that dissociate protein interactions that have potential to be therapeutic agents[78]. The genotype of MaV203 is *MATα; leu2-3,112; trp1-901; his3Δ200; ade2-101; cyh2R; can1R; gal4Δ; gal80Δ; GAL1::lacZ; HIS3UASGAL1::HIS3@LYS2; SPAL10::URA3*.

The main characteristic of the MaV203 strain is that it contains deletions of the endogenous GAL4 and GAL80 genes and that three GAL4-inducible reporter genes (*GAL1::LacZ*, *HIS3$_{UASGAL1}$*::*HIS3* and *SPAL10::URA3*[78,78]) are present. Therefore, in the absence of GAL4, MaV203 is auxotrophic for both histidine and uracil, meaning that histidine and uracil are not produced and need to be added to the growth medium. Besides, the MaV203 cells are also auxotrophic for leucine and tryptophan, allowing for the selection of successful transformation with plasmids containing the *LEU2* or *TRP1* genetic marker. The *SPAL10::URA3* is counter selective when GAL4 is present, *i.e.* it can be used for both positive as well as negative selection (Figure 2.4). The URA3 gene expresses orotidine-5'-phosphate decarboxylase, catalyzing the sixth step in the *de novo* biosynthesis of uracil by means of converting oritidine-5'-monophosphate (OMP) into uridine-5'-monophosphate (UMP). This permits growth on single drop-out medium SD-URA (medium lacking uracil). However, this enzyme also converts 5-fluoroorotic acid (5-FOA) into the toxic compound, 5-fluorouracil, causing cell death.

Most yeast promotors fused to the URA3 gene confer a URA3[+] phenotype (growth on SD-URA) when no specific transcription factor is present[80,81]. Basal levels of expression of the URA3 gene are sufficient to promote growth on a medium lacking uracil and induce sensitivity for 5-FOA. In order to prevent the basal expression of the URA3 gene, a fusion between the SPO13 promoter and the URA3 open reading frame was made by Vidal *et al.*[78]. The SPO13 promoter is only active under sporulation conditions[82] and is strictly repressed under normal conditions[83-87]. Thus, this fusion confers a very strict URA[-] phenotype (no growth in absence of uracil), 5-FOA resistant phenotype under normal growth conditions. Derivatives of the SPO13 promoter containing GAL4 binding sites, e.g. SPAL10, were constructed by Vidal *et al.*[78].
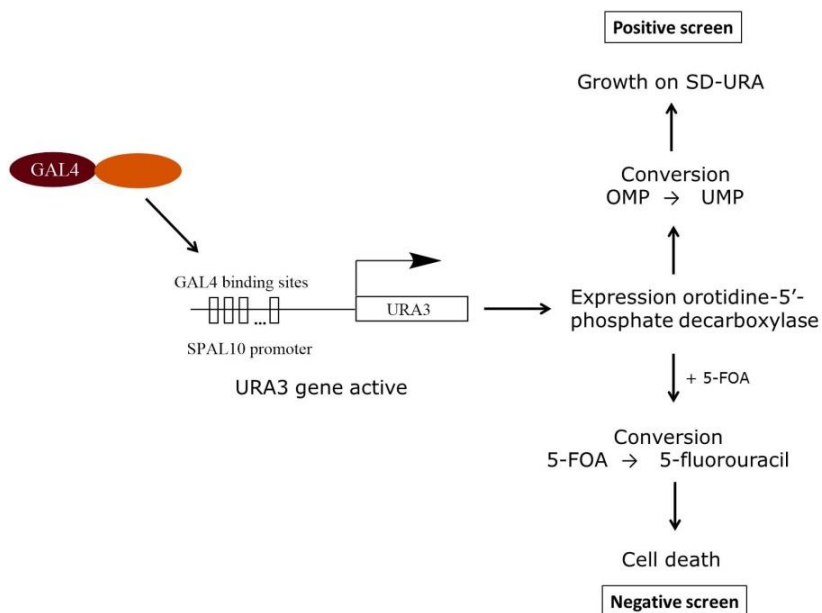
Figure 2.4. Positive and negative screening via the counter selective *SPAL10::URA3* gene fusion. In the presence of GAL4 the URA3 gene is active and expresses orotidine-5'-phosphate decarboxylase. This enzyme catalyzes the conversion of oritidine-5'-monophosphate (OMP) into uridine-5'-monophosphate (UMP), allowing growth on single drop-out medium SD-URA (positive screen). However, this enzyme also converts 5-fluoroorotic acid (5-FOA), when added to the growth medium, into the toxic compound 5-fluorouracil (negative screen).

The SPAL10 promoter in MaV203 cells is a SPO13 derivative containing ten GAL4 binding sites. This means that the URA3 gene is only expressed if GAL4 is present. Due to the *SPAL10::URA3* fusion, these cells can be used for the screening of large libraries of mutant EcTyrRSs. The details of library screening will be discussed in section 4.1.1. Moreover, these cells were also used for the expression of modified NbBCII10 and green fluorescent protein (GFP) (Chapter 6).

INVSc1. Besides the MaV203 cells, another strain of *S. cerevisiae* was used for the expression of NbBCII10_His$_6$ and GFP, namely the strain INVSc1 (Life Technologies, Naarden, the Netherlands) (Appendix page 185). The INVSc1 strain is a fast-growing diploid strain, optimized for protein expression. The genotype is *MATa his3D1 leu2 trp1-289 ura3-52 MAT his3D1 leu2 trp1-289 ura3-52.*

2.4 References

1.　Hamers-Casterman, C. *et al.,* Naturally occurring antibodies devoid of light chains, *Nature.* **363,** 446–448 (1993).

2.　Muyldermans, S. *et al.,* Camelid immunoglobulins and nanobody technology, *Vet. Immunol. Immunopathol.* **128,** 178–83 (2009).

3.　Woolven, B. P. *et al.*, The structure of the Ilama heavy chain constant genes reveals a mechanism for heavy-chain antibody formation, *Immunogenetics.* **50**, 98–101 (1999).

4.　Muyldermans, S. *et al.,* Camelid immunoglobulins and nanobody technology, *Vet. Immunol. Immunopathol.* **128,** 178–183 (2009).

5.　Muyldermans, S. *et al.*, Single domain camel antibodies: Current status, *Rev. Mol. Biotechnol.* **74,** 277–302 (2001).

6.　Revets, H. *et al.*, Nanobodies as novel agents for cancer therapy, *Expert Opin. Biol. Ther.* **5,** 111–124 (2005).

7.　Khong Nguyen, V. *et al.*, Loss of splice consensus signal is responsible for the removal of the entire C(H)1 domain of the functional camel IGG2A heavy-chain antibodies, *Mol. Immunol.* **36,** 515–524 (1999).

8.　Maass, D. R. *et al.*, Alpaca (Lama pacos) as a convenient source of recombinant camelid heavy chain antibodies (VHHs), *J. Immunol. Methods* **324,** 13–25 (2007).

9.　Muyldermans, S *et al.*, Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains, *Protein Eng.* **7,** 1129–1135 (1994).

10.　Cortez-Retamozo, V. *et al.,* Efficient cancer therapy with a nanobody-based conjugate, *Cancer Res.* **64,** 2853–2857 (2004).

11.　Deffar, K. *et al.*, Nanobodies - the new concept in antibody engineering, *J. Biotechnol.* **8,** 2645–2652 (2009).

12.　Riechmann, L. *et al.*, Single domain antibodies: comparison of camel VH and camelised human VH domains, *J. Immunol. Methods* **231,** 25–38 (1999).

13.　Harmsen, M. M. *et al.*, Properties, production, and applications of camelid single-domain antibody fragments, *Appl. Microbiol. Biotechnol.* **77,** 13–22 (2007).

14. De Genst, E. *et al.,*Antibody repertoire development in camelids, *Dev. Comp. Immunol.* **30,** 187–98 (2006).

15. Lauwereys, M. *et al.,* Potent enzyme inhibitors derived from dromedary heavy-chain antibodies, *EMBO J.* **17,** 3512–3520 (1998).

16. De Genst, E. *et al.,* Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies, *Proc. Natl. Acad. Sci. U.S.A.* **103,** 4586–4591 (2006).

17. Stijlemans, B. *et al.,* Efficient targeting of conserved cryptic epitopes of infectious agents by single domain antibodies: African trypanosomes as paradigm, *J. Biol. Chem.* **279,** 1256–1261 (2004).

18. Desmyter, A *et al.,* Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme, *Nat. Struct. Biol.* **3,** 803–811 (1996).

19. Pérez, J. M. J. *et al.,* Thermal unfolding of a llama antibody fragment: A two-state reversible process, *Biochemistry* **40,** 74–83 (2001).

20. Dumoulin, M. *et al.,* Single-domain antibody fragments with high conformational stability, *Protein Science.* **11,** 500–515 (2002).

21. Ewert, S. *et al.*, Biophysical properties of camelid VHH domains compared to those of human VH3 domains, *Biochemistry* **41,** 3628–3636 (2002).

22. Van der Linden, R. H. *et al.,* Comparison of physical chemical properties of llama VHH antibody fragments and mouse monoclonal antibodies, *Biochim. Biophys. Acta* **1431,** 37–46 (1999).

23. De Meyer, T. *et al.*, Nanobody-based products as research and diagnostic tools, *Trends Biotechnol.* **32,** 263–270 (2014).

24. Ta, D. T. *et al.,* An efficient protocol towards site-specifically clickable nanobodies in high yield: cytoplasmic expression in *Escherichia coli* combined with intein-mediated protein ligation, *Protein Eng. Des. Sel.* **28,** 351–363 (2015).

25. Frenken, L. G. *et al.,* Isolation of antigen specific llama VHH antibody fragments and their high level secretion by Saccharomyces cerevisiae, *J. Biotechnol.* **78,** 11–21 (2000).

26. Thomassen, Y. E. *et al.*, Large-scale production of V HH antibody fragments by *Saccharomyces cerevisiae, Production* **30,** 273–278 (2002).

27. Verheesen, P. *et al.*, Beneficial properties of single-domain antibody fragments for application in immunoaffinity purification and immuno-perfusion chromatography, *Biochim. Biophys. Acta - Gen. Subj.* **1624,** 21–28 (2003).

28. Pleschberger, M. *et al.*, An S-layer heavy chain camel antibody fusion protein for generation of a nanopatterned sensing layer to detect the prostate-specific antigen by surface plasmon resonance technology, *Bioconjug. Chem.* **15,** 664–671 (2004).

29. Dolk, E. *et al.,* Isolation of llama antibody fragments for prevention of dandruff by phage display in shampoo, *Appl. Environ. Microbiol.* **71,** 442–450 (2005).

30. Dumoulin, M. *et al.*, A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme, *Nature* **424,** 783–788 (2003).

31. Baral, T. N. *et al.,* Experimental therapy of African trypanosomiasis with a nanobody-conjugated human trypanolytic factor, *Nat. Med.* **12,** 580–584 (2006).

32. Roovers, R. C. *et al.,* Efficient inhibition of EGFR signalling and of tumour growth by antagonistic anti-EGFR Nanobodies, *Cancer Immunol. Immunother.* **56,** 303–317 (2007).

33. Reichert, J. M. *et al.*, Development trends for monoclonal antibody cancer therapeutics, *Nat. Rev. Drug Discov.* **6,** 349–56 (2007).

34. Coppieters, K. *et al.,* Formatted anti-tumor necrosis factor VHH proteins derived from camelids show superior potency and targeting to inflamed joints in a murine model of collagen-induced arthritis, *Arthritis Rheum.* **54,** 1856–1866 (2006).

35. Harmsen, M. M. *et al.*, Prolonged *in vivo* residence times of llama single-domain antibody fragments in pigs by binding to porcine immunoglobulins, *Vaccine* **23,** 4926–4934 (2005).

36. Harmsen, M. M. *et al.,* Passive immunization of guinea pigs with llama single-domain antibody fragments against foot-and-mouth disease, *Vet. Microbiol.* **120,** 193–206 (2007).

37. Cortez-Retamozo, V. *et al.,* Efficient tumor targeting by single-domain antibody fragments of camels, *Int. J. Cancer.* **98,** 456–

462 (2002).

38. Holland, M. *et al.*, Autoantibodies to variable heavy (V_H) chain Ig sequences in humans impact the safety and clinical pharmacology of a V_H domain antibody against TNF-α receptor 1, *J. Clin. Immunol.* **33,** 1192–1203 (2013).

39. Papadopoulos, K. *et al.,* Unexpected hepatotoxicity in a phase I study of TAS266, a novel tetravalent agonistic Nanobody targeting the DR5 receptor, *Cancer Chemother Pharmacol.* **75,** 887–895 (2015).

40. Presta, L. G. *et al.*, Engineering of therapeutic antibodies to minimize immunogenicity and optimize function, *Adv. Drug Deliv. Rev.* **58,** 640–656 (2006).

41. Klooster, R. *et al.,* Improved anti-IgG and HSA affinity ligands: Clinical application of VHH antibody technology, *J. Immunol. Methods* **324,** 1–12 (2007).

42. Saerens, D. *et al.,* Parallel selection of multiple anti-infectome Nanobodies without access to purified antigens, *J. Immunol. Methods* **329,** 138–150 (2008).

43. Habib, I. *et al.,* VHH (nanobody) directed against human glycophorin A: A tool for autologous red cell agglutination assays, *Anal. Biochem.* **438,** 82–89 (2013).

44. Vaneycken, I. *et al.,* Preclinical screening of anti-HER2 nanobodies for molecular imaging of breast cancer, *FASEB J.* **25,** 2433–2446 (2011).

45. Huang, L. *et al.,* SPECT imaging with ^{99m}Tc-labeled EGFR-specific nanobody for *in vivo* monitoring of EGFR expression, *Mol. Imaging Biol.* **10,** 167–175 (2008).

46. De Groeve, K. *et al.,* Nanobodies as tools for *in vivo* imaging of specific immune cell types, *J. Nucl. Med.* **51,** 782–9 (2010).

47. Kijanka, M. *et al.,* Rapid optical imaging of human breast tumour xenografts using anti-HER2 VHHs site-directly conjugated to IRDye 800CW for image-guided surgery, *Eur. J. Nucl. Med. Mol. Imaging* **40,** 1718–1729 (2013).

48. Oliveira, S. *et al.,* Rapid visualization of human tumor xenografts through optical imaging with a near-infrared fluorescent anti-epidermal growth factor receptor nanobody, *Mol. Imaging* **11,** 33–46 (2012).

49. Saerens, D. *et al.,* Identification of a universal VHH framework to graft non-canonical antigen-binding loops of camel single-domain antibodies, *J. Mol. Biol.* **352,** 597–607 (2005).

50. Beier, H. *et al.*, Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs, *Nucleic Acids Res.* **29,** 4767–4782 (2001).

51. Herrington, M. B. *et al.*, Nonsense Mutations and Suppression Suppression In: eLS. John Wiley & Sons, Ltd: Chichester (2013).

52. Kohli, J. *et al.*, Characterization of a UGA-suppressing serine tRNA from *Schizosaccharomyces pombe* with the help of a new *in vitro* assay system for eukaryotic suppressor tRNAs, *J. Biol. Chem.* **254,** 1546–1551 (1979).

53. Scotia, N. *et al.,* Defined Set of Cloned Termination Suppressors, *J. Bacteriol.* **158,** 849–859 (1984).

54. Wang, L. *et al.*, A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the *in vivo* incorporation of unnatural amino acids into proteins, *J. Am. Chem. Soc.* **122,** 5010–5011 (2000).

55. Drabkin, H. J. *et al.*, Amber suppression in mammalian cells dependent upon expression of an *Escherichia coli* aminoacyl-tRNA synthetase gene, *Microbiology,* 907-913 (1996).

56. Kowal, A K. *et al.*, Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria, *Proc. Natl. Acad. Sci. U. S. A.* **98,** 2268–2273 (2001).

57. Edwards, H. *et al.*, bacterial amber suppressor in *Saccharomyces cerevisiae* is selectively recognized by a bacterial aminoacyl-tRNA synthetase, *Mol. Cell. Biol.* **10,** 1633–41 (1990).

58. Edwards, H. *et al.*, An *Escherichia coli* tyrosine transfer RNA is a leucine-specific transfer RNA in the yeast *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 1153–1156 (1991).

59. Kwok, Y. *et al.*, Evolutionary relationship between *Halobacterium cutirubrum* and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes, *Can. J. Biochem.* **58,** 213–218 (1980).

60. Wakasugi, K. *et al.*, Genetic code in evolution: switching species-

specific aminoacylation with a peptide transplant, *EMBO J.* **17,** 297–305 (1998).

61. Köhrer, C. *et al.*, Import of amber and ochre suppressor tRNAs into mammalian cells: a general approach to site-specific insertion of amino acid analogues into proteins, *Proc. Natl. Acad. Sci. U.S.A.* **98,** 14310–5 (2001).

62. Chin, J. W. *et al.,* Progress Toward an Expanded Eukaryotic Genetic Code, *Chem Biol.* **10,** 511–519 (2003).

63. Young, T. S. *et al.*, Beyond the canonical 20 amino acids: expanding the genetic lexicon, *J. Biol. Chem.* **285,** 11039–44 (2010).

64. Hatfield, D. L. *et al.*, *Transfer RNA in Protein Synthesis*. (1992).

65. Yasukazu, N. Codon Usage Database. *Genet. codes* (2000). at <http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=4932>

66. Deiters, A. *et al.,* Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae, *J. Am. Chem. Soc.* **125,** 11782–3 (2003).

67. Cereghino, G. P. *et al.*, Applications of yeast in biotechnology: protein production and genetic analysis, *Curr. Opin. Biotechnol.* **10,** 422–7 (1999).

68. Demain, A. L. *et al.*, Production of recombinant proteins by microbes and higher organisms, *Biotechnol. Adv.* **27,** 297–306 (2009).

69. Huang, C. *et al.*, Recombinant immunotherapeutics: Current state and perspectives regarding the feasibility and market, *Appl. Microbiol. Biotechnol.* **87,** 401–410 (2010).

70. Çelik, E. *et al.*, Production of recombinant proteins by yeast cells, *Biotechnol. Adv.* **30,** 1108–1118 (2012).

71. Eckart, M. R. *et al.*, Quality and authenticity of heterologous proteins synthesized in yeast, *Curr. Opin. Biotechnol.* **7,** 525–30 (1996).

72. Yin, J. *et al.*, Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes, *J. Biotechnol.* **127,** 335–347 (2007).

73. Wang, Q. *et al.*, Expanding the genetic code for biological studies. *Chem. Biol.* **16,** 323–36 (2009).

74. Sakamoto, K. *et al.,* Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells, *Nucleic Acids Res.* **30,** 4692–9 (2002).

75. Liu, W. *et al.*, Genetic incorporation of unnatural amino acids into proteins in mammalian cells, *Nat. Methods* **4,** 239–244 (2007).

76. Wang, W. *et al.,* Genetically encoding unnatural amino acids for cellular and neuronal studies, *Nat. Neurosci.* **10,** 1063–72 (2007).

77. Liu, C. C. *et al.*, Adding new chemistries to the genetic code, *Annu. Rev. Biochem.* **79,** 413–44 (2010).

78. Vidal, M. *et al.*, Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions, *Proc. Natl. Acad. Sci. U.S.A.* **93,** 10315–20 (1996).

79. Vidal, M. *et al.*, Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system, *Proc. Natl. Acad. Sci. U.S.A.* **93,** 10321–10326 (1996).

80. Losson, R. *et al.*, *In vivo* transcription of a eukaryotic regulatory gene, *EMBO Rep.* **2,** 2179–2184 (1983).

81. Aparicio, O. M. *et al.*, Overcoming telomeric silencing: a trans-activator competes to establish gene expression in a cell cycle-dependent way, *Genes Dev.* **8,** 1133–1146 (1994).

82. Wang, H. T. *et al.,* Developmental regulation of SPO13, a gene required for separation of homologous chromosomes at meiosis I, *Mol. Cell. Biol.* **7,** 1425–35 (1987).

83. Luche, R. M. *et al.*, A cis-acting element present in multiple genes serves as a repressor protein binding site for the yeast CAR1 gene, *Mol. Cell. Biol.* **10,** 3884–95 (1990).

84. Park, H. O. *et al.*, Transcriptional regulation of a yeast HSP70 gene by heat shock factor and an upstream repression site-binding factor, *Genes Dev.* **5,** 1299–1308 (1991).

85. Bowdish, K. S. *et al.*, Bipartite structure of an early meiotic upstream activation sequence from *Saccharomyces cerevisiae*, *Mol. Cell. Biol.* **13,** 2172-2181 (1993).

86. John, M. *et al.*, The INO1 promoter of *saccharomyces cerevisiae*

includes upstream repressor sequence (URS1) common to a diverse set of yeast genes. **175,** 4235–4238 (1993).

87. Vidal, M. *et al.*, Identification of essential nucleotides in an upstream repressing sequence of *Saccharomyces cerevisiae* by selection for increased expression of TRK2., *Proc. Natl. Acad. Sci. U.S.A.* **92,** 2370–4 (1995).

Chapter 3

DNA constructs for the synthesis of modified nanobodies

This chapter discusses the DNA constructs that are necessary for the production of azide-modified nanobodies as well as to perform the orthogonality assay/proof of principle experiment for diploid cells. The first part starts with an outline of the amber suppression technique, in particular for the site-specific modification of nanobodies. This is done by adding the genes of the *E. coli* tyrosyl-tRNA synthetase (EcTyrRS)/tRNA$_{CUA}^{Tyr}$ pair to the genetic repertoire of *S. cerevisiae*. In addition, a detailed discussion regarding the construction of the vectors is given. In the second part of this chapter, the orthogonality assay/proof of principle experiment for diploid cells and the construction of the corresponding vectors will be discussed. This proof of principle/ orthogonality assay is performed to i) verify that diploid cells can be used for the screening of a library of mutant EcTyrRSs, and ii) that the EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair does not interact with the endogenous translational machinery of the yeast, *i.e.* that EcTyrRS does not aminoacylate any yeast tRNA's, nor that tRNA$_{CUA}^{Tyr}$ is aminoacylated by any yeast aminoacyl-tRNA synthetase (aaRS). Figure 3.1 gives an overview of the methodology used for the synthesis of site-specifically modified nanobodies with the construction of the vectors highlighted.
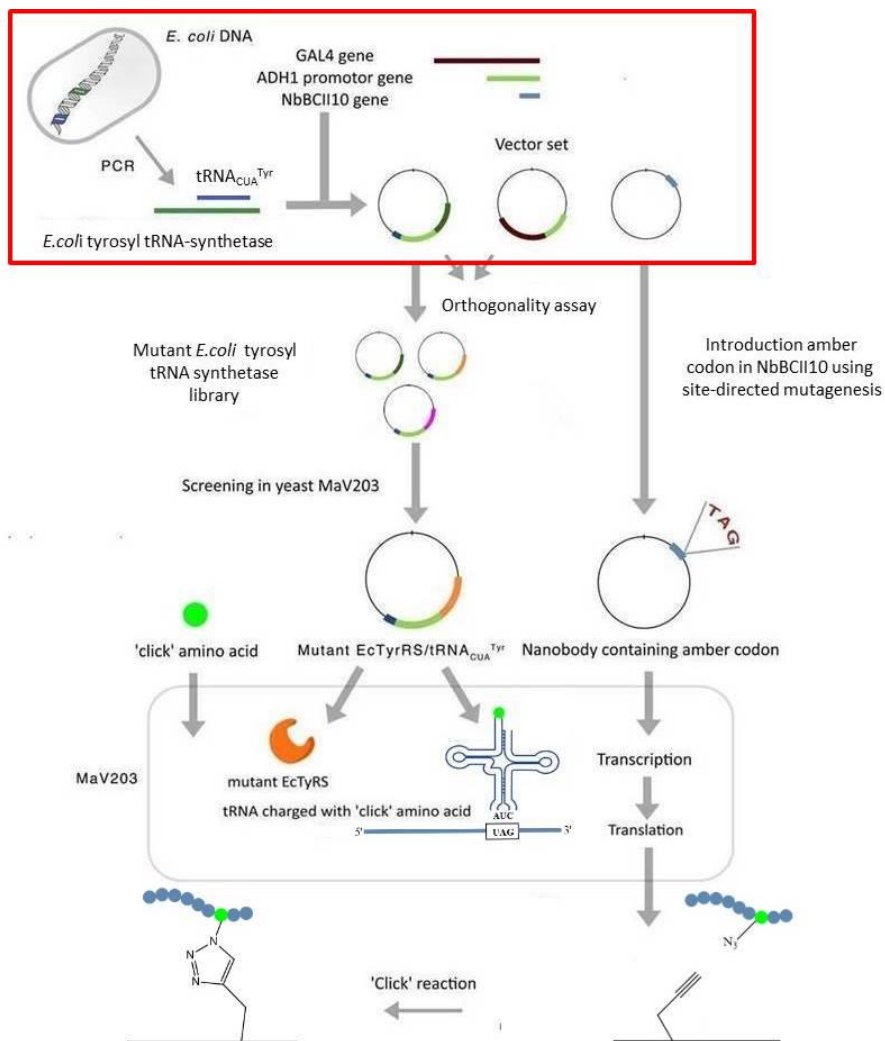
Figure 3.1. Methodology used for the synthesis of site-specifically modified nanobodies. The construction of the necessary DNA constructs (indicated by the red frame) is the first step in the synthesis of site-specifically functionalized nanobodies.

3.1 Overview of site-specific protein modification using the amber suppression approach

The site-specific modification of the nanobody BCII10 (NbBCII10_His$_6$) is based on the incorporation of p-azidophenylalanine (p-azidoPhe) at a well-chosen position by means of amber suppression. The principle of amber suppression is described in detail in section 2.2. In this context, the amber suppressor *E. coli* tyrosine tRNA$_{CUA}$ (tRNA$_{CUA}$^{Tyr}) is used in

combination with a mutant *E. coli* tyrosyl-tRNA synthetase (EcTyrRS)[1–4]. Site-directed mutagenesis is performed to alter key amino acids in the binding pocket of EcTyrRS so that it has affinity for *p*-azidoPhe instead of tyrosine. As a result, the *E. coli* tRNA$_{CUA}^{Tyr}$ will be charged with *p*-azidoPhe and incorporates this amino acid in response to an amber codon. Figure 3.2 represents the methodology.
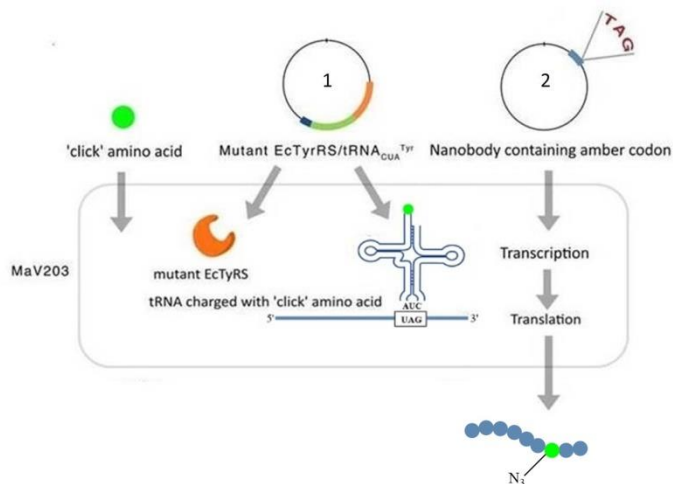
Figure 3.2. Basic representation of the amber suppression approach for the site-specific modification of NbBCII10_His$_6$. Vectors pEcTyrRS(mutant)/SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t (1) and pTEF-MF/NbBCII10_TAG (2) are co-transformed in the yeast MaV203. The non-natural amino acid *p*-azidoPhe ('click' amino acid) is added to the growth medium. Vector 1 will express both the amber suppressor tRNA$_{CUA}^{Tyr}$ and the mutant EcTyrRS with affinity for *p*-azidoPhe. This pair will incorporate the 'click' amino acid in response to an amber codon in NbBCII10_His$_6$, expressed by vector 2. Since the NbBCII10 gene is fused to the pre pro leader sequence of mating factor α (MF), the azido-modified nanobody will be secreted to the growth medium.

The following section describes in detail the development of the DNA constructs necessary to introduce *p*-azidoPhe in the NbBCII10_His$_6$.

As shown in Figure 3.2, it concerns the following vectors:

1. **pEcTyrRS(mutant)/SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t**,   containing the *mutant E. coli* tyrosyl-tRNA synthetase (EcTyrRS) and the amber suppressor tRNA$_{CUA}^{Tyr}$

2. **pTEF-MF/NbBCII10_TAG** containing the NbBCII10 gene with a C-terminal His$_6$-tag, located after the mating factor α leader sequence (MF). The NbBCII10 gene contains an amber codon at

position 132 (the codon for glycine 132 has been converted to an amber codon).

As shown in Figure 3.2 these two vectors will be co-transformed[5] in the yeast MaV203[6]. Here, the vector pEcTyrRS(mutant)/ SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t will express i) the amber suppressor tRNA$_{CUA}$^{Tyr} and ii) the mutant EcTyrRS with an affinity for *p*-azidoPhe. This pair will be able to incorporate *p*-azidoPhe (when present in the growth medium) in respons to an amber codon in the mRNA of NbBCII10_His$_6$. Since the NbBCII10 gene is fused to the pre pro leader sequence of Mating Factor α (MF), the modified NbBCII10_His$_6$ will be secreted to the growth medium.

3.2 Development of DNA constructs for the synthesis of modified NbBCII10 (general method)

The details on the composition and construction of the vectors pEcTyrRS(mutant)/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t and pTEF-MF/ NbBCII10_TAG, as well as the other vectors used throughout the protein expression experiments will be discussed here and in section 3.3.

An overview of all the vectors is listed below:

- pEcTyrRS/tRNA$_{CUA}$^{Tyr} (contains wildtype EcTyrRS)
- pEcTyrRS(mutant)/tRNA$_{CUA}$^{Tyr}
- pEcTyrRS/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t (contains wildtype EcTyrRS)
- pEcTyrRS(mutant)/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t
- pTEF-MF/NbBCII10_His$_6$
- pTEF-MF/ NbBCII10_TAG
- pTEF/GFP
- pTEF/GFP_Y39TAG
- pADH1/GFP
- pADH1/GFP_Y39TAG

In general, the construction of the vectors described in section 3.2, 3.3 and 3.4 can be summarized in three steps. A description of these steps is shown in Figure 3.3.

Figure 3.3. Overview of vector construction.

1. *Amplification of necessary genes.* By means of PCR[7–10] the genes were amplified from a single colony of *E. coli* BL21 (DE3) pLysS (Appendix page 185) or from plasmids. The BL21 strain is a standard *E. coli* strain commonly used in research laboratories. The primers were designed to introduce a unique restriction site up- and downstream of the gene, which was later used for cloning the gene in the correct vector. Table A1 (Appendix) gives an overview of the primers used.

2. *Ligation in pCR2.1 and transformation.* The amplified DNA was visualized by agarose gel electrophoresis and ligated in the pCR2.1 vector using the commercial TA Cloning Kit (Invitrogen, Carlsbad, USA). Subsequently, it was transformed in *E. coli* Top10F' competent cells (Appendix page 185), according to the manufacturer's protocol. During the PCR amplification (Figure 3.3 step 1) Taq polymerase adds a single deoxyadenosine monophosphate (dAMP) to the 3' ends of the PCR

products. The linearized pCR2.1 vector supplied in the kit has single 3' deoxythymidine monophosphate (dTMP) residues.
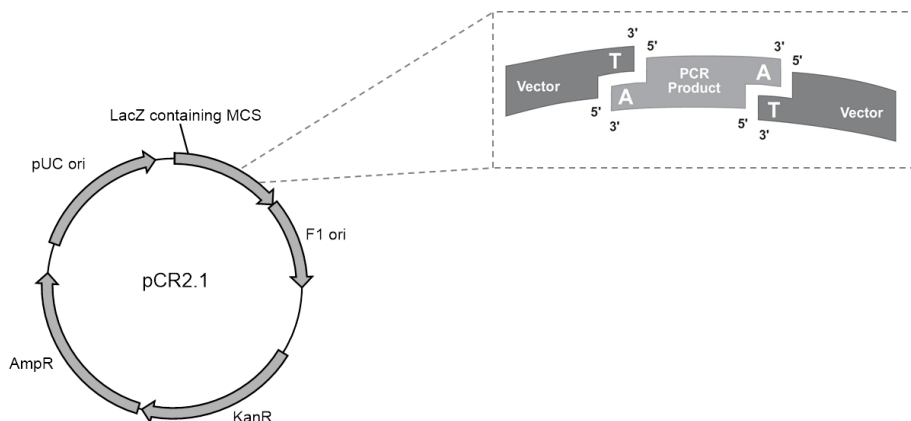
Figure 3.4. Schematic overview of the pCR2.1 vector. The pCR2.1 vector is supplied in a linearized form with single 3' deoxythymidine monophosphate (dTMP) residues. These residues are located in the *LacZ*α gene. This allows the insertion of the PCR product, containing dAMP overhangs introduced by the Taq polymerase, resulting in an inactive *LacZ*α gene.

These dTMP overhangs allow PCR products (containing dAMP overhangs) to ligate efficiently with the vector (Figure 3.4). An EcoRI restriction site is located close to the dTMP residue and can be used to verify the successful introduction of the PCR insert by means of restriction analysis. Since Taq polymerase has no proof-reading mechanism, there is a risk that mutations appeared during the PCR amplification. For that reason all the PCR reactions were performed with the Pfu polymerase, which has a proof-reading mechanism. However, this polymerase does not add single dAMP residues to the 3' ends of the PCR products. Therefore, after the PCR reaction, 0.5 μl of Taq polymerase was added to the PCR mix followed by incubation at 72 °C for 10 minutes.

After the ligation reaction the pCR2.1 vector was transformed into Top10F' *E. coli* cells. Successful ligations of the insert are detected by a blue-white screening. Therefore, both X-gal (5-bromo-4-chloro-3-indolyl β-D-galactopyranoside) (40 μl; 40 mg/ml) and IPTG (isopropyl-beta-D-thiogalactopyranoside) (40 μl; 100 mM) were added to the LB plates (Luria-Bertani) of transformed Top10F' cells. The pCR2.1 vector contains the fragment of the *LacZ*α gene that is part of the *lac* operon and is responsible for the expression of β-galactosidase. The F' episome of the Top10F' cells carries the *lacI*q repressor which inhibits the production of β-galactosidase in the absence of lactose. The addition of IPTG, a lactose metabolite, induces the activation of the *LacZ*α gene and

subsequent expression of β-galactosidase. β-galactosidase cleaves the colorless X-gal, yielding galactose and 5-bromo-4-chloro-3-hydroxyindole which is oxidized into 5,5'-dibromo-4,4'-dichloro-indigo, an insoluble blue product resulting in blue colonies. However, the multiple cloning site (MCS) of the pCR2.1 vector is located in the *LacZα* gene (Figure 3.4).

When an insert is successfully introduced in the multiple cloning site of the pCR2.1 vector, the *LacZα* gene becomes inactive. The expression of β-galactosidase is thus inhibited and as a result, X-gal is not cleaved. Colonies containing an insert will appear white, whereas colonies without the insert will appear blue. Liquid cultures of white colonies are made in Luria Bertani (LB) medium containing ampicillin (100 µg/ml final concentration) and tetracycline (12,5 µg/ml final concentration), after which the plasmid DNA is isolated using the QIAprep Spin Miniprep Kit (Qiagen, Venlo, the Netherlands) and the DNA sequences analyzed.

3. *Isolation and ligation into the correct vector.* Both the pCR2.1 vector, containing the amplified gene insert, as well as the target vector are cut with two restriction endonucleases. The recognition sites of these endonucleases are present in the target vector and were introduced up- and downstream of the amplified gene by the primers. After agarose gel analysis the insert and the cut target vector are purified using the Qiagen gel extraction kit (Qiagen, Venlo, the Netherlands). Afterwards the genes are ligated into the target vector and transformed in Top10F' cells. The ligations are performed using T4 DNA ligase with a molar ratio of vector to insert of 1:3. The plasmid DNA is isolated and the presence of the correct insert is verified by means of restriction analysis[12] and DNA sequencing (outsourced to LGC Genomics, Augsburg, Germany).

3.2.1 Construction of the pEcTyrRS(mutant)/tRNA$_{CUA}$^{Tyr} vector

For the construction of the pEcTyrRS(mutant)/tRNA$_{CUA}$^{Tyr} vector, the pESC-TRP vector (Agilent Technologies, Brussels, Belgium) was used as the backbone vector. The pECS-TRP vector is a yeast episomal plasmid (YEp)[13] of 6.5 kbp, designed for the expression of proteins in *S. cerevisiae.* The pESC-TRP vector contains following features:

- the pUC plasmid origin of replication and the ampicillin-resistance gene, that allows replication and selection in *E. coli*

- the yeast 2μ origin, that enables autonomous replication of the plasmid in *S. cerevisiae*
- an auxotrophic marker that allows to select and maintain the expression vector in yeast cells. In this specific vector it is the TRP1 auxotrophic marker that enables the growth on single drop-out medium SD-TRP (growth medium lacking tryptophan).

The EcTyrRS gene and tRNA^{Tyr} gene were amplified from a single colony of *E. coli* BL21 (DE3) pLys. The full-length ADH1 promoter was amplified from the vector pMA424 (kindly provided by Prof. Dr. Fields of the Howard Hughes Medical Institute, U.S.A.). The ADH1 promoter is a constitutive promoter of the alcohol dehydrogenase 1 gene. When this promoter is placed upstream it will cause constitutive expression of EcTyrRS.

Amplification and cloning of the E. coli tyrosyl-tRNA synthetase (EcTyrRS) gene

Optimization of the PCR reaction. In order to determine the optimal PCR conditions to amplify the EcTyrRS gene, a gradient PCR was performed on a single colony of *E. coli* BL21 (DE3) pLys, and this for both the annealing temperature (35-55 °C) and the MgCl$_2$ concentration (2, 4, 6 and 8 mM). The PCR reaction mixtures were analyzed with agarose gel electrophoresis from which it could be concluded that 2 mM MgCl$_2$ was optimal. A second gradient PCR reaction was done to optimize the annealing temperature, ranging from 45-70 °C. The reaction mixtures were analyzed by agarose gel electrophoresis. The EcTyrRS gene has a length of 1275 bp. Taking the extra nucleotides added by the primers and the dAMP added by the Taq polymerase into account, a fragment of 1308 bp is expected. Primer annealing appeared to be most efficient at 54.3 °C and 49.9 °C as indicated by the clear bands at ~1300 bp on the agarose gel (Figure 3.5, lanes 5 and 6 respectively). The reaction mixture from lane 6 (primer annealing temperature of 49.9 °C) was used for the ligation in the pCR2.1 vector. The resulting pCR2.1/EcTyrRS vector was transformed in Top10 F' cells. The composition of the reaction mixture and the settings for the PCR program can be found in Table A2 and A3 (appendix), respectively.
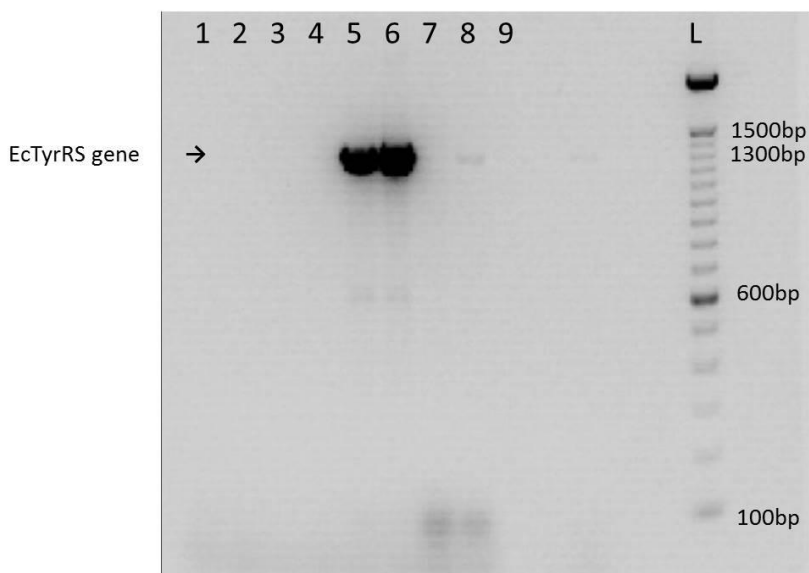
Figure 3.5. Amplification of the EcTyrRS gene from *E. coli* BL21 (DE3) pLys (1% agarose). A temperature gradient was applied to optimize the PCR result: 1=70.0 °C; 2=68.1 °C; 3=65.0 °C; 4=60.3 °C; 5=54.3 °C; 6=49.9 °C; 7=46.8 °C; 8=45.0 °C; 9= negative control (mastermix without DNA); L=100bp ladder (Invitrogen). A temperature of 49.9 °C (lane 6) appeared to be the optimal annealing temperature for the amplification of the EcTyrRS gene.

Restriction analysis and sequencing. The Top10F' cells, transformed with the pCR2.1/EcTyrRS vector, were allowed to grow overnight after which the plasmid DNA was isolated from white colonies and the insert was analyzed by restriction digestion using the two EcoRI restriction sites flanking the insert. After addition of the restriction enzyme EcoRI, two fragments were expected for clones containing EcTyrRS: i) a fragment of approximately 3.9 kbp (the open pCR2.1 vector) and ii) a fragment of 1323 bp (EcTyrRS + the extra nucleotides between the EcoRI site and the EcTyrRS gene). In total 10 clones were tested.

Figure 3.6 shows that the clones 3, 4, 5, 7, 8 and 9 contain fragments of the expected length (1323 bp). Clone 2 and 10 also appear to contain the EcTyrRS gene but also show an additional band at > 2 kbp (clone 2) and ~650 bp (clone 10). In order to verify these results, the inserts of clones 3, 4, 5, 7, 8 and 9 were sequenced and aligned against the sequence of *E. coli* BL21 TyrRS. Clones 4, 5 and 7 showed identical sequences as compared to the reference sequence of *E. coli* BL21 TyrRS. Clones 3, 8 and 9 on the other hand contained one or more point mutations. The sequence alignment of clones 3 and 4 are given as an example in S1 of the appendix.
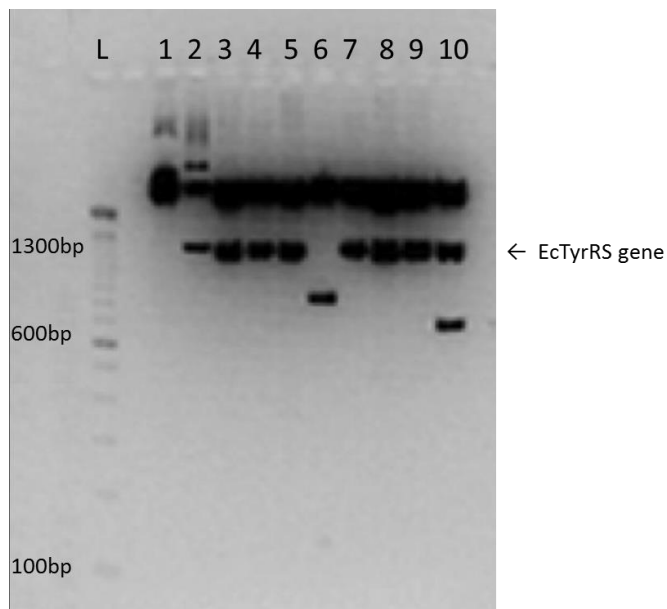
Figure 3.6. Restriction digestion of pCR2.1/EcTyrRS with restriction enzyme EcoRI (1% agarose). Lanes 1-10=clones 1-10; L=100bp ladder (Invitrogen). Fragments of the expected length (1323bp) are present in lanes 3, 4, 5, 7, 8 and 9. Clone 2 and 10 also appear to contain the EcTyrRS gene but also show a band at >2000 bp (clone 2) and ~650 bp (clone 10).

Ligation of the EcTyrRS gene in the pESC-TRP vector. The EcTyrRS gene of clone 4 was cut from the pCR2.1 vector using the restriction enzymes EcoRI and NotI. Note that these restriction sites have been introduced by the primers used to amplify the EcTyrRS gene (see Table A1 of the appendix). This digestion should result in i) a fragment of 1323 bp (the EcTyrRS gene + the "sticky ends" from the restriction sites) and, ii) a fragment of approximately 3.9kbp (the open pCR2.1 vector). The pESC-TRP vector was cut with the same enzymes, which should generate two fragments: i) a fragment of approximately 6.5 kbp (the open pESC-TRP vector) and, ii) a fragment of approximately 22 bp (not visible on a 1% agarose gel). After restriction digestion, the reaction mixtures were analyzed using agarose gel electrophoresis. Bands are visible at the expected height on the agarose gel (Figure 3.7), indicating a successful restriction digestion.
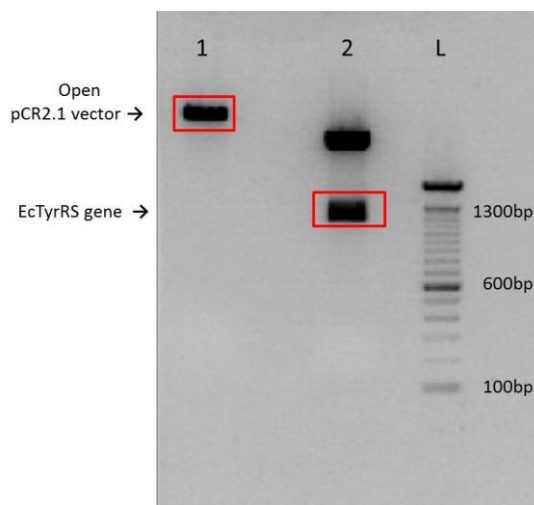
Figure 3.7. Restriction digestion of the pESC-TRP vector and pCR2.1/EcTyrRS with restriction enzymes EcoRI and NotI (1% agarose). Lane 1=open pESC-TRP vector (6.5 kbp) indicated by the red frame; lane 2=the highest band represents the opened pCR2.1 vector, the lowest band, indicated by the red frame, represents the EcTyrRS gene (1323 bp); L= 100bp ladder (Invitrogen).

The EcTyrRS gene and the open pESC-TRP vector were excised from the agarose gel and after purification of the DNA, the EcTyrRS gene was ligated into the pESC-TRP vector and subsequently transformed in competent Top10 F' *E. coli* cells, resulting in the pESC-TRP/EcTyrRS vector (Figure 3.8). Plasmid DNA from four colonies was isolated and sequenced. At this point, all clones contained the EcTyrRS insert without mutations.
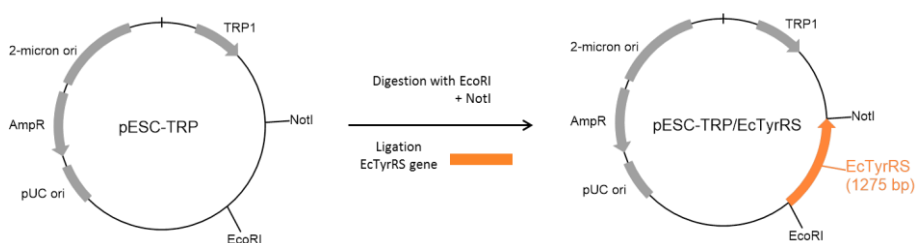
Figure 3.8. Scheme of the ligation of the EcTyrRS gene in the pESC-TRP vector to form the pESC-TRP/EcTyrRS vector.

Amplification and cloning of the *E. coli* tRNA^{Tyr}gene

Optimization of the PCR reaction. A PCR was performed on a single colony of *E. coli* BL21 (DE3) pLys, in order to amplify the conventional *E. coli* tRNA^{Tyr} gene. The amber suppressor tRNA$_{CUA}^{Tyr}$ can be easily acquired by mutating the anticodon of the conventional tRNA^{Tyr}. This will be discussed later in this section.

The first set of experiments focused on determining the optimal PCR conditions. A gradient was applied for both the annealing temperature (45-70 °C) and the MgCl$_2$ concentration (2, 4, 6 and 8 mM). The PCR reaction mixtures were analyzed by agarose gel electrophoresis from which it was clear that 4 mM MgCl$_2$ was optimal. A second gradient PCR, ranging from 55 °C to 75 °C, was performed to optimize the annealing temperature. The PCR products were analyzed by agarose gel electrophoresis (Figure 3.9), from which it was observed that 62.5 °C was the optimal annealing temperature. Taking into account the length of *E. coli* tRNA^{Tyr} (156 bp), plus the nucleotides added by the primers and the dAMP added at the 3' end by Taq polymerase, a fragment of 181 bp was expected.
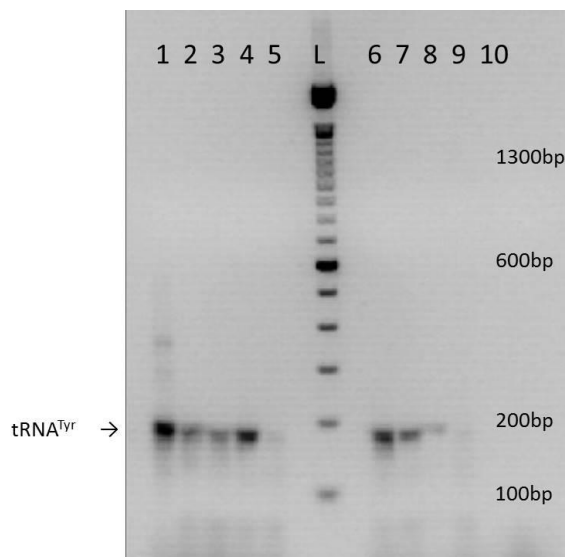
Figure 3.9. Amplification of the tRNA^{Tyr} gene from a *E. coli* BL21 (DE3) pLysS (2% agarose). A temperature gradient was applied to optimize the PCR. L=100bp ladder (Invitrogen); 1=62.5 °C; 2=73.5 °C; 3=71.0 °C; 4=67.2 °C; 5=75.0 °C; 6=58.9 °C; 7=56.5 °C; 8= 55.0 °C; 9=no sample loaded (the visible bands are the result of leakage from the adjacent lane); 10=negative control (mastermix without DNA).

Figure 3.9 shows that starting from 75.0 °C, a decrease in temperature results in an increase of density of the tRNA^{Tyr} band until a maximum is reached at 62.5 °C. Upon decreasing the annealing temperature further, the density lowered again. The tRNA$_{CUA}^{Tyr}$ amplified under the most optimal conditions (primer annealing at 62.5 °C and 4 mM MgCl$_2$) was cloned into the pCR2.1 vector, followed by the pCR2.1/tRNA^{Tyr} transformation in Top10 F' cells. The composition of the reaction mixture and the settings for the PCR program can be found in respectively Table A2 and A4 (appendix).

Restriction analysis and sequencing The pCR2.1/tRNA^{Tyr} vector was isolated and analyzed by restriction digestion with the restriction enzyme EcoRI (Figure 3.10). Here, two fragments were expected: i) a fragment of approximately 3.9 kbp (the open pCR2.1 vector) and, ii) a fragment of 196 bp (tRNA^{Tyr} + the nucleotides introduced by the primers). In total 10 clones were tested, but only the clones in lanes 5, 9 and 10 contain fragments of the expected lengths (marked by the red frames).
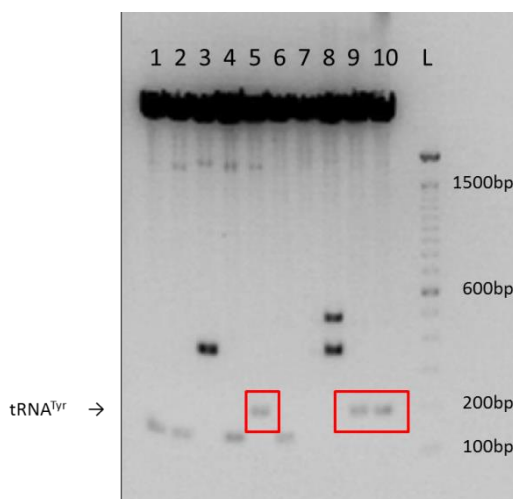
Figure 3.10. Restriction digestion of pCR2.1/tRNA^{Tyr} with EcoRI (2% agarose). Lanes 1-10=clones 1-10; L=100bp ladder (Invitrogen). Fragments of the expected length (196 bp, in red squares) are present in the clones of lane 5, 9 and 10.

Moreover, these results were verified by sequencing (S2 in the appendix). All clones showed only one nucleotide difference in comparison to the *E. coli* tRNA$_{CUA}^{Tyr}$ sequence. The secondary structure of *E. coli* tRNA$_{CUA}$ as well as of the amplified tRNA was modelled using

the online tool 'tRNAscan-SE Search Server'[14,15] (Figure 3.11). The model revealed that the different nucleotide was located in the anticodon (Figure 3.11, marked by red square). The anticodon GUA of the amplified *E. coli* tyrosyl tRNA was modified into CUA using site-directed mutagenesis.
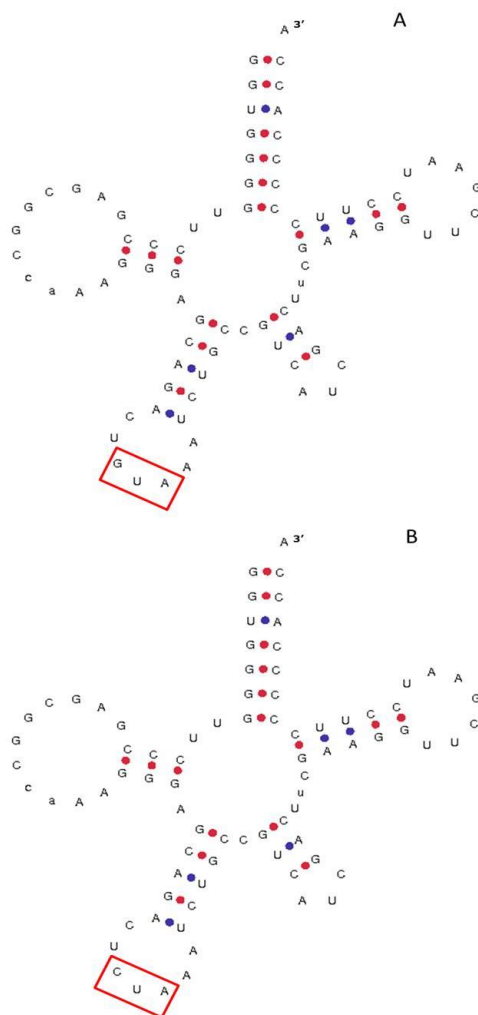
Figure 3.11. (A) Secondary structure of the tRNA^{Tyr} amplified from *E. coli* BL21 (DE3) pLysS and, (B) The *E. coli* amber suppressor tRNA$_{CUA}^{Tyr}$ modelled using the online tool 'tRNAscan-SE Search Server'. The different nucleotide corresponds to the anticodon of the tRNA^{Tyr}. As expected, the amplified tRNA is not the amber suppressor tRNA$_{CUA}^{Tyr}$ but the conventional tRNA$_{GUA}^{Tyr}$.

Site-Directed Mutagenesis of the tRNA^{Tyr} anticodon. Site-directed mutagenesis was performed using the QuickChange Lightning Site-Directed Mutagenesis kit (Stratagene, Agilent Technologies, Brussels, Belgium), according to the manufacturer's protocol. The pCR2.1 vector containing the *E. coli* tyrosyl-tRNA insert (pCR2.1/ tRNA^{Tyr}) was used as template DNA. In addition, the forward and reverse primers are displayed in Table A5 (appendix). After the mutagenesis, the sequence of the tRNA^{Tyr} was verified and it was confirmed that the anticodon had been changed from GUA to CUA and so recognizes the amber stop codon (S3 in the appendix).

Ligation of the tRNA$_{CUA}^{Tyr}$ gene in the pESC-TRP/EcTyrRS vector. In order to perform this ligation, the *E. coli* tRNA$_{CUA}^{Tyr}$ gene was cut from the pCR2.1 vector using the restriction enzymes AgeI and NheI. This should result in a fragment of 166 bp (the tRNA$_{CUA}^{Tyr}$ gene + the "sticky ends" from the restriction sites) and a fragment of approximately 3.9 kbp (the open pCR2.1 vector). The plasmid pESC-TRP/EcTyrRS was cut with the same enzymes, which results in a fragment of approximately 500 bp and a fragment of 6 kbp (the open pESC-TRP/EcTyrRS vector). After verifying that the restriction digestion was successful (Figure 3.12), the corresponding bands of 166 bp (tRNA$_{CUA}^{Tyr}$ gene) and 6 kbp (the open pESC-TRP/EcTyrRS vector) were excised from the gel and purified using the Qiagen Gel Extraction kit.
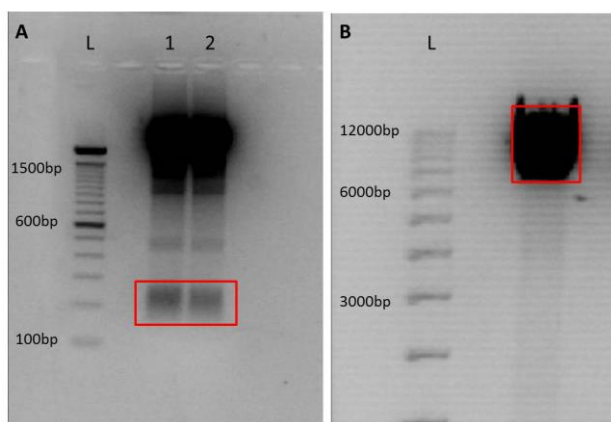
Figure 3.12. Restriction digestion of pCR2.1/tRNA$_{CUA}^{Tyr}$ (A) and pECS-TRP/EcTyrRS (B) with restriction enzymes NheI and AgeI. A) Agarose gel (2%). Restriction digestion of pCR2.1/ tRNA$_{CUA}^{Tyr}$ (performed in duplo): L=100bp ladder (Invitrogen); Lanes 1+2=the highest bands represent the open pCR2.1 vector (3.9 kbp). The lowest band, indicated with a red frame, represents the tRNA$_{CUA}^{Tyr}$ gene (166bp). (B) Agarose gel (1%). Restriction digestion of pECS-TRP/EcTyrRS. L=1 kbp plus ladder (Invitrogen). The band, indicated with a red frame, represents the open pESC-TRP/EcTyrRS.

Finally the *E. coli* tRNA$_{CUA}$^{Tyr} gene was ligated in the pESC-TRP/EcTyrRS vector, resulting in the pESC-TRP/EcTyrRS/tRNA$_{CUA}$^{Tyr} vector (Figure 3.13). After transformation in Top10F' cells, plasmid DNA from four colonies was isolated and sequenced, confirming the ligation and correct sequence of the tRNA$_{CUA}$^{Tyr} gene in the vector.
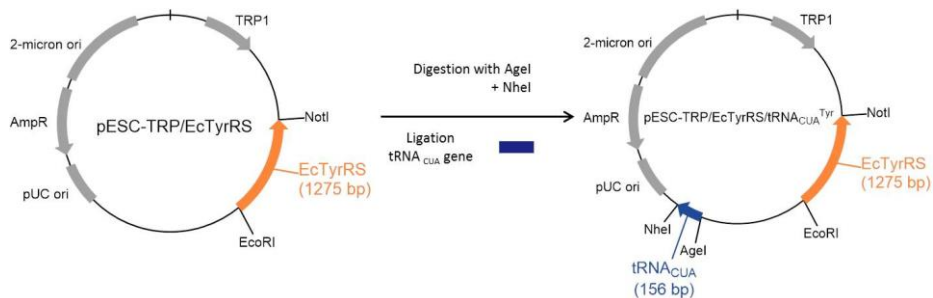
Figure 3.13. Schematic view of the ligation of the tRNA$_{CUA}$^{Tyr} gene in the pESC-TRP/EcTyRS vector to form the vector pESC-TRP/EcTyrRS/tRNA$_{CUA}$^{Tyr}.

Amplification and cloning of the S. cerevisiae ADH1 promoter

Optimization of the PCR reaction. A PCR was performed on the plasmid pMA424 in order to amplify the full-length yeast ADH1 promoter. Four different concentrations of pMA424 DNA were tested, ranging from 2 ng to 40 ng. After the amplification reaction (Table A6 of the appendix gives an overview of the PCR program settings), the reaction mixtures were analyzed by agarose gel electrophoresis. From the gel, it could be concluded that no difference was present for the different plasmid DNA concentrations used. The reaction mixture with the lowest DNA concentration was used for the ligation in the pCR2.1 vector. The resulting pCR2.1/ADH1 construct was transformed into Top10 F' competent *E. coli* cells and, plasmid DNA was isolated from 10 white colonies.

Restriction analysis and sequencing The presence of the correct insert was verified with restriction digestion and sequencing (S4 of the appendix). The results of the restriction analysis are shown in Figure 3.14. Also here, two fragments were expected: i) a fragment of approximately 3.9 kbp (the open pCR2.1 vector) and ii) a fragment of 1483 bp (the ADH1 promoter + the nucleotides introduced by the primers).
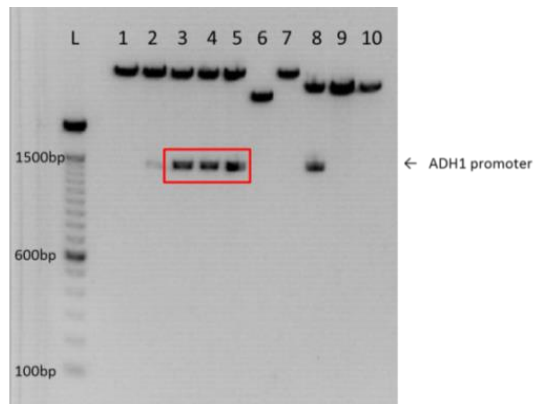
Figure 3.14. Restriction digestion of pCR2.1/ADH1 with EcoRI (1.5% agarose). L=100bp ladder (Invitrogen); Lanes 1-10=clones 1-10 from pCR2.1/ADH1. Fragments of the expected length were clearly visible in clones 3, 4, 5 (red frame). Clone 8 also seemed to contain the correct insert although the position of the open pCR2.1 vector (3.9 kbp)is lower than expected.

Fragments of the expected length were clearly observed in lanes 3, 4 and 5. Also in lane 8 a clear band of 1483 bp is present but the band corresponding to the open pCR2.1 vector appears at lower position than expected. Therefore, the presence of the correct ADH1 promoter in the pCR2.1 vector of clones in lanes 3, 4 and 5 was verified by sequencing. Whereas clone 4 contained two mutations, the clones 3 and 5 showed complete sequence similarity with the reference ADH1 gene (pMA424 vector) (S4 in the appendix).

Ligation of the ADH1 promoter in pESC-TRP/EcTyrRS/tRNA$_{CUA}$^{Tyr}. In order to complete the pEcTyrRS/tRNA$_{CUA}$^{Tyr} construct, the insert corresponding to the amplified ADH1 promoter of clone 3 was ligated into the pESC-TRP/EcTyrRS/tRNA$_{CUA}$^{Tyr} vector.

Hereto, the ADH1 promoter was excised from the pCR2.1 vector using the restriction enzymes EcoRI and AgeI. Digestion of pCR2.1/ADH1 should result in two fragments: i) a fragment of 1453 bp (the ADH1 promoter) and, ii) a fragment of 3.9 kbp (the open pCR2.1 vector). Next, the pESC-TRP/EcTyrRS/tRNA$_{CUA}$^{Tyr} vector was cut using the same enzymes and should generate a fragment of approximately 6.2 kbp (the open vector) and a fragment of approximately 300 bp. The fragments were than analyzed by agarose gel electrophoresis by which it was confirmed that the restriction reaction was successful (Figure 3.15). The ADH1 promoter and the open pESC-TRP/EcTyrRS/tRNA$_{CUA}$^{Tyr} vector were excised and purified from the agarose gel. Then, the ADH1 promoter

was ligated into the pESC-TRP/EcTyrRS/tRNA$_{CUA}^{Tyr}$ vector, resulting in the complete pEcTyrRS/tRNA$_{CUA}$ vector (Figure 3.16).
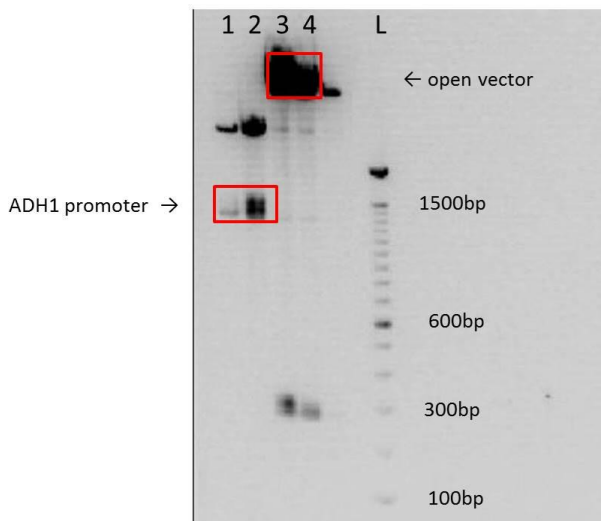
Figure 3.15. Restriction digestion of pCR2.1/ADH1 and pESC-TRP/EcTyrRS/tRNA$_{CUA}^{Tyr}$ with EcoRI and AgeI (1% agarose). Lanes 1 & 2: the red frame indicates the ADH1 promoter, successfully excised from the vector, while the red frame in lanes 3 & 4 indicates the open pESC-TRP/EcTyrRS/tRNA$_{CUA}^{Tyr}$ vector.
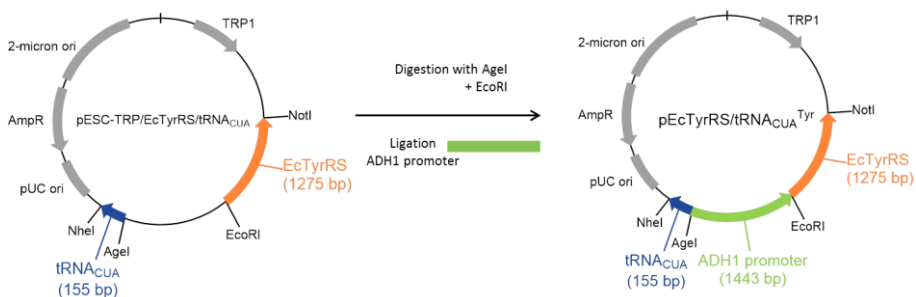
Figure 3.16: Schematic representation of the ligation of the ADH1 promoter in the pESC-TRP/EcTyrRS /tRNA$_{CUA}^{Tyr}$ construct to form the pEcTyrRS /tRNA$_{CUA}^{Tyr}$ vector.

Site-directed mutagenesis of EcTyrRS. It was decided to develop a mutant EcTyrRS for which the high affinity for *p*-azidoPhe was already described in literature[4]. To obtain this mutant, three amino acids in the binding pocket were mutated, namely Y37T, D182S and F183A[4,16]. The site-directed mutagenesis was performed using the Quickchange Lightning Site-directed Mutagenesis kit and the vector

pEcTyrRS/tRNA$_{CUA}$^{Tyr} was used as a template. The primers used are shown in Table A7 (appendix). In S5 of the appendix the protein sequence of the wildtype EcTyrRS is compared with that of the obtained mutant EcTyrRS and confirms the successful mutagenesis of the EcTyrRS gene, which resulted in the pEcTyrRS(mutant)/tRNA$_{CUA}$^{Tyr} vector.

Optimization of the tRNA$_{CUA}$^{Tyr} expression

The transcription of tRNA is different in prokaryotes compared to eukaryotes. Therefore, prokaryotic tRNA is not properly expressed in eukaryotic cells, resulting in lower incorporation efficiencies of the modified amino acid[17]. In prokaryotes, tRNA transcription occurs under the influence of promoters located upstream of the tRNA gene, while in eukaryotes this promoter sequence is present in the tRNA gene itself (Figure 3.17A). These sequences are known as the so-called "A-box" and "B-box" (Figure 3.17). Without an external promoter the tRNA$_{CUA}$^{Tyr} gene in pEcTyrRS/tRNA$_{CUA}$^{Tyr} is expressed at a basal level. This basal expression is enough to perform the orthogonality assays and to screen libraries of mutant EcTyrRSs, but not to incorporate modified amino acids at high rate in heterologous proteins[18].

Adding the A- and B-box in the *E. coli* tRNA$_{CUA}$^{Tyr} gene by means of site-directed mutagenesis is not an option as it has been described that the tRNA$_{CUA}$^{Tyr} will no longer be functional[19]. However, it is possible to efficiently express *E. coli* tRNA$_{CUA}$^{Tyr} in yeast by the addition of an external promoter, containing an A- and B-box, in front of the tRNA$_{CUA}$^{Tyr} gene[20] (Figure 3.17B). As mentioned in literature[17], the expression of tRNA$_{CUA}$^{Tyr} can be elevated in eukaryotes by adding the promoter sequence of SNR52 (small nuclear RNA) (upstream of the tRNA$_{CUA}$^{Tyr} gene) in combination with the SUP4 terminator sequence (SUP4=yeast tRNA^{Tyr} gene) (downstream of the tRNA$_{CUA}$^{Tyr} gene). The SNR52 promoter sequence will be post-transcriptionally removed from the mature *E. coli* tRNA$_{CUA}$^{Tyr}. It is noteworthy to mention that the elevated tRNA-activity is not a consequence of higher transcription levels, but of better tRNA processing[17].
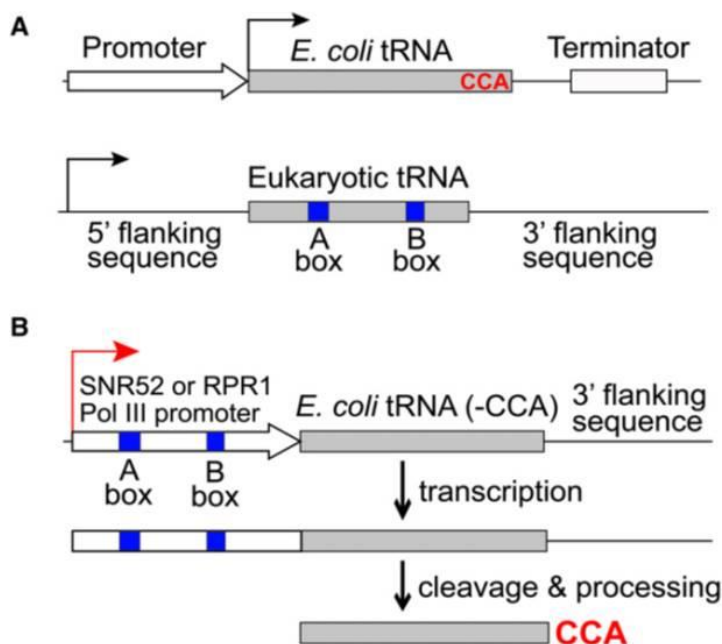
Figure 3.17. (A) Comparison of the genetic elements for tRNA transcription in prokaryotic and eukaryotic cells[17]. (B) Expression of prokaryotic tRNA in eukaryotic cells can be done by using an external PolIII promoter, such as the SNR52 promoter, containing both the A- and B-box sequences. This promoter will be post-transcriptionally cleaved from the primary transcript.

Introducing the SNR52 promoter sequence and the SUP4 terminator sequence in pECTyrRS/tRNA$_{CUA}^{Tyr}$ and in pECTyrRS(mutant)/tRNA$_{CUA}^{Tyr}$. In order to introduce the SNR52 promoter and the SUP4 terminator sequence, the following construct was made (IDT, Leuven, Belgium) (S6 in the appendix). This SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t construct contained the following components:

"AgeI site---SNR52promoter---tRNA$_{CUA}^{Tyr}$---SUP4t---NdeI site"

The tRNA gene will be removed from the vector pEcTyrRS/tRNA$_{CUA}^{Tyr}$ as well as from vector pEcTyrRS(mutant)/tRNA$_{CUA}^{Tyr}$ and will be replaced by the SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t construct. The tRNA$_{CUA}^{Tyr}$ gene is located between the AgeI and the NheI restriction site in the vector pEcTyrRS/tRNA$_{CUA}^{Tyr}$. Since the NheI restriction site is also present in the SNR52 promoter, it cannot be used for the introduction of the construct. This issue was solved by introducing an NdeI restriction site between the AgeI and NheI sites as follows.

First the tRNA$_{CUA}^{Tyr}$ gene was cut from the vectors pECTyrRS(mutant)/tRNA$_{CUA}^{Tyr}$ and pECTyrRS/tRNA$_{CUA}^{Tyr}$ using AgeI and NheI, resulting in a linear vector with an AgeI and a NheI overhang (shown in Figure 3.18A). Two oligonucleotides were designed to be partially each other's reverse complement (IDT, Leuven, Belgium) (Figure 3.18B). When hybridized, these oligonucleotides will form a double stranded "linker" with the NdeI restriction site in the center (marked in green) and, at the ends the overhangs which are complementary to the overhangs generated when a vector is cut with AgeI and NheI (Figure 3.18C). The double stranded oligonucleotide is then ligated between the AgeI and NheI overhangs of the vector resulting in the addition of an NdeI restriction site between the AgeI and NheI restriction sites (Figure 3.18D).
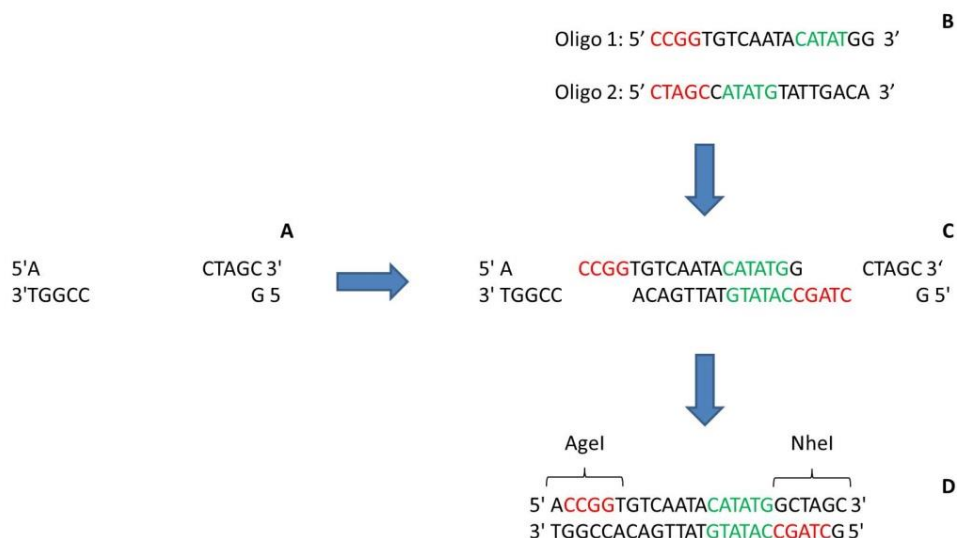
Figure 3.18. Introduction of a NdeI restriction site (marked in green) between the AgeI and NheI restriction sites. (A) The AgeI and NheI overhangs present in the linearized vector, (B) the two oligonucleotides that were designed to be partially each other's complement, (C) the hybridization of the oligonucleotides to form a linker with the NdeI restriction site in its centre, (D) the introduction of the NdeI site, between the AgeI and NheI restriction site.

After cutting the resulting vectors and the SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t construct with the AgeI and NdeI restriction enzymes, the construct can be ligated in the correct position. This results in the vector pEcTyrRS/SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t (and pEcTyrRS(mutant)/SNR52_ tRNA$_{CUA}^{Tyr}$_SUP4t) (Figure 3.19). Finally, the correct ligation of the construct was verified with restriction analyses and DNA sequencing.
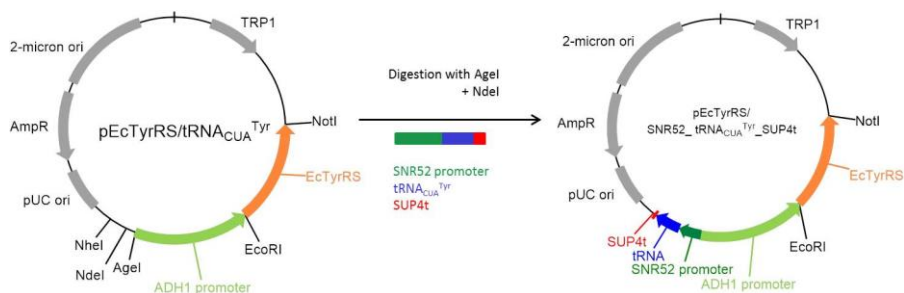
Figure 3.19: Schematic representation of the ligation of the SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t construct in the pEcTyrRS/tRNA$_{CUA}^{Tyr}$ vector and pEcTyrRS(mutant)/tRNA$_{CUA}^{Tyr}$ vector to form the pEcTyrRS/SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t and pEcTyrRS(mutant) /SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t vectors, respectively.

3.2.2 Construction of the pTEF-MF/NbBCII10_His$_6$ vector

PCR reaction for the amplification of the NbBCII10_His$_6$ gene For the construction of the pTEF-MF/NbBCII10_His$_6$ vector, the NbBCII10_His$_6$ gene is ligated into the pTEF-MF vector (Dualsystems Biotech, Schlieren, Switzerland). This pTEF-MF vector is a yeast vector for the expression and secretion of heterologous proteins. It contains a strong TEF1 promoter driving the constitutive expression of a cDNA (NbBCII10_His$_6$ gene) fused to the pre-pro leader sequence of Mating Factor alpha (MFα), ensuring the secretion of the protein into the growth medium. The NbBCII10 gene, containing a C-terminal His$_6$-tag, was amplified from the pHEN6-BCII10 vector using PCR. The EcoRI and XhoI restriction sites were introduced up- and down-stream of the NbBCII10_His$_6$ gene using the primers listed in Table A1 (appendix). After confirmation by 1.5% agarose gel electrophoresis that the PCR was successful, the NbBCII10 gene was ligated in the pCR2.1 vector, and its sequence was verified by sequencing (S7 in the appendix). The PCR reaction mixture and the PCR program can be found in the Table A2 and A8 (appendix), respectively.

Ligation of the NbBCII10_His$_6$ gene in the pTEF-MF vector With the purpose of cloning NbBCII10_His$_6$ in the pTEF-MF yeast expression vector, both pCR2.1/NbBCII10_His$_6$ and pTEF-MF were digested with the restriction enzymes XhoI and EcoRI. Afterwards, the NbBCII10_His$_6$ gene was successfully ligated into the pTEF-MF vector, downstream of the MF α leader sequence, resulting in the vector pTEF-MF/NbBCII10_His$_6$ (Figure 3.20).
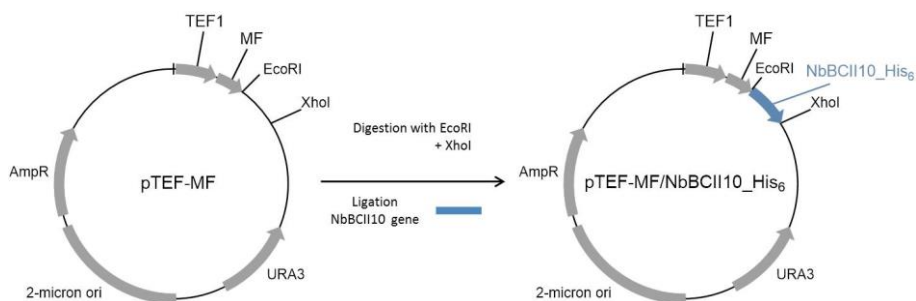
Figure 3.20. Scheme of the ligation of the NbBCII10_His$_6$ gene in the pTEF-MF vector to construct the pTEF-MF/NbBCII10_His$_6$ vector.

Site-directed mutagenesis of NbBCII10_His$_6$ The vector pTEF-MF/NbBCII10_His$_6$ contains the wildtype NbBCII10 gene which is used as a positive control in the protein expression experiments. However to produce site-specifically modified nanobodies, an amber codon needs to be introduced at a location opposite to the active site in order to allow maximal availability for the target binding. As can be seen on Figure 2.2, the C-terminus is opposite to the active site and it is the optimal location for functionalization. Between the C-terminus of NbBCII10 and the His$_6$-tag are three amino acids present (RGR), wich are not part of the nanobody but are a result of the restriction site (NotI) used to construct the NbBCII10_His$_6$ fusion. It was chosen to alter the codon encoding this glycine (G) to an amber codon, resulting in an amber codon between the C-terminus of NbBCII10 and the His$_6$-tag (Figure 3.21). In this way, if the amber suppression is successful, the modified nanobodies contain the His$_6$-tag and can be purified using a nickel column or can be detected using an anti-his antibody. In case amber suppression is not succesfull, no His$_6$-tag will be present. In this context, a site-directed mutagenesis was carried out using the QuickChange Lightning Site-Directed Mutagenesis kit (Stratagene, Agilent Technologies, Brussels, Belgium), using the vector pTEF-MF/NbBCII10_His$_6$ as a template.

Figure 3.21. The NbBCII10_TAG gene with an amber codon located between the C-terminus and the His$_6$-tag. Therefore, only when the amber codon (UAG) is successfully suppressed, will the nanobody contain a C-terminal His$_6$-tag.

Therefore, in case of successful amber suppression the resulting nanobodies contain a C-terminal His$_6$-tag, allowing their detection using mouse IgG monoclonal anti-histidine antibody. If the amber codon is not suppressed, no His$_6$-tag will be present. The forward and reverse primers are displayed in Table A9 (appendix). After the site-directed mutagenesis experiment, sequencing (S8 in the appendix) confirmed that the amber codon was successfully introduced, resulting in the vector pTEF-MF/NbBCII10_TAG.

3.3 DNA constructs for the cytoplasmic expression of GFP

As a proof of principle for amber suppression in yeast, Green Fluorescent Protein (GFP), with an amber codon at location 39 (GFP_Y39TAG) was cytosolically expressed. If the amber codon is successfully suppressed, full-length GFP is formed in the cytosol and the cells will fluoresce when exposed to UV-light. When amber suppression is not successful, the premature termination will result in a non-fluorescent truncated GFP molecule. Therefore, by expressing GFP_Y39TAG in the cytosol of MaV203 cells, successful amber suppression in GFP can be verified by simply monitoring the fluorescence of the cells. For this purpose, two types of vectors were constructed:

- pTEF/GFP and pTEF/GFP_Y39TAG

- pADH1/GFP and pADH1/GFP_Y39TAG

3.3.1 Construction of the vectors pTEF/GFP and pTEF/GFP_Y39TAG

For the construction of these vectors, the GFP gene is ligated into the pTEF-MF vector (Dualsystems Biotech, Schlieren, Switzerland). However, when ligating the GFP_Y39TAG gene in this vector the MF α leader sequence was removed. As a consequence, the GFP will not be transported to the growth medium but will stay in the cytosol. To create this vector, the GFP gene containing the amber codon at location 39 (GFP_Y39TAG) was amplified from the vector pBAD_GFP-39TAG (kindly provided by Dr. Lemke (EMBL, Heidelberg)). The primers used introduced a SpeI as well as an XhoI restriction site (Table A1 in the appendix). After confirmation on a 1.5% agarose gel, both GFP_Y39TAG

and pTEF-MF were digested with the restriction enzymes SpeI and XhoI. Afterwards, GFP_Y39TAG was successfully ligated into the pTEF-MF vector, removing the MF α leader sequence in the process and resulting in the pTEF/GFP_Y39TAG vector (Figure 3.22) (S9 in the appendix). The PCR reaction mixture and PCR program for GFP_Y39TAG can be found in Table A2 and Table A10 (appendix), respectively.
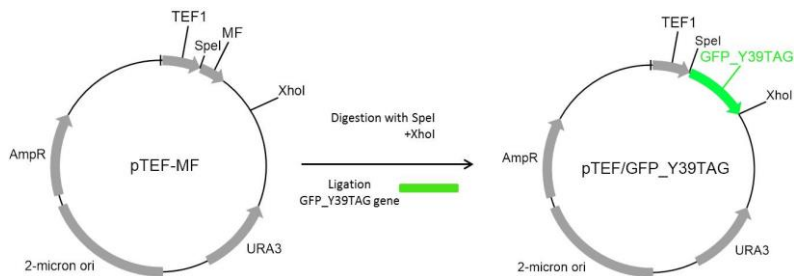
Figure 3.22. Scheme of the ligation of the GFP_Y39TAG gene in the pTEF-MF vector to construct the pTEF/GFP_Y39TAG vector.

Site-Directed Mutagenesis of GFP_Y39TAG to wildtype GFP. The vector pTEF/GFP_Y39TAG contains the GFP_Y39TAG gene. So, as a positive control for the expression experiments, it was necessary to generate the same vector containing the wild type GFP.

A site-directed mutagenesis was carried out using the QuickChange Lightning Site-Directed Mutagenesis kit (Stratagene, Agilent Technologies, Brussels; Belgium) according to the manufacturer's protocol. The pTEF/GFP_Y39TAG vector was used as template DNA. After the site-directed mutagenesis, it was verified by sequencing (S10 in the appendix) that the amber codon was successfully mutated to a tyrosine (TAG → TAT), resulting in the wild type GFP and thus the vector pTEF/GFP.

3.3.2 Construction of the vectors pADH1/GFP and pADH1/GFP_Y39TAG

For the construction of the pADH1/GFP and pADH1/GFP_Y39TAG vectors, the pESC-LEU vector (Agilent Technologies; Brussels; Belgium) was used as the backbone vector. The pECS-LEU vector is a yeast episomal plasmid (YEp)[13] of 7.8 kbp, designed for the expression of proteins in the yeast *S. cerevisiae.* The pESC-LEU vector contains following features:

- The pUC plasmid origin of replication and the ampicillin-resistance gene allowing replication and selection in *E. coli*
- The yeast 2μ origin, enabling autonomous replication of the plasmids in *S. cerevisiae*
- An auxotrophic selectable marker to select and maintain the expression vector in yeast cells. This vector contains the LEU2 auxotrophic marker allowing the growth on SD-LEU.

First, the ADH1 promoter was amplified using the vector pMA424 The required primers (Table A1 of the appendix) contain an NheI and BamHI restriction site. After confirmation that the PCR was successful on a 1.5% agarose gel, the ADH1 promoter was ligated in the pESC-LEU vector. Hereto, both the ADH1 promoter and the pESC-LEU vector were digested with the restriction enzymes NheI and BamHI. Successful ligation of the ADH1 promoter in the pESC-LEU vector, as confirmed by DNA sequencing, resulted in the vector pESC-LEU/ADH1 (Figure 3.23).
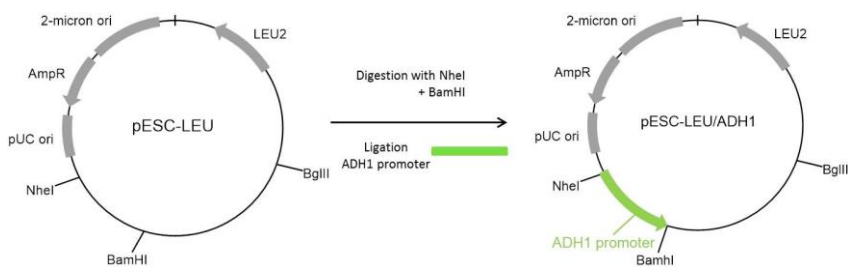
Figure 3.23. Schematic representation of the ligation of the ADH1 promoter in the pESC-LEU vector to form the vector pESC-LEU/ADH1.

Next, both the wildtype GFP as well as GFP_Y39TAG were amplified using the vectors pTEF/GFP and pTEF/GFP_Y39TAG respectively. The primers used introduced a BamHI and a BglII restriction site (Table A1). After confirming by 1% agarose gel electrophoresis that the PCR was successful, the GFP gene (both wildtype GFP as well as GFP_Y39TAG) was ligated in the pESC-LEU/ADH1 vector. Hereto both the GFP genes as well as the pESC-LEU/ADH1 vector were digested with the restriction enzymes BamHI and BglII. After this the GFP gene was successfully ligated into the pESC-LEU/ADH1 vector, resulting in the vectors pADH1/GFP and pADH1/GFP_Y39TAG (Figure 3.24).
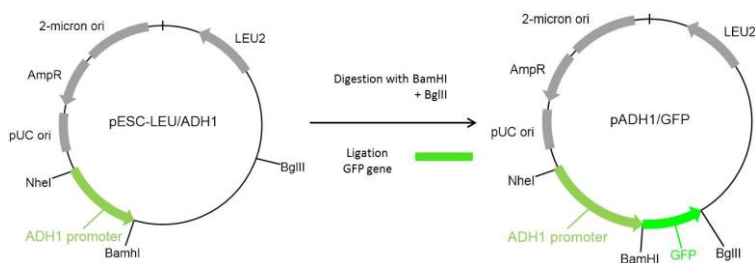
Figure 3.24. Schematic view of the ligation of the GFP gene in the pESC-LEU/ADH1 vector to form the vector pADH1/GFP.

3.4 DNA constructs used in the proof of principle experiment for the use of diploid cells and the validation of the orthogonality of the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair

The second part of this chapter discusses the proof of principle experiment to verify if diploid cells can be used for the screening of the mutant library of aaRSs and the development of the DNA constructs needed for it. This proof of principle experiment was performed because the results discussed in chapter 4 showed that a single recessive mutation occurs in the haploid MaV203 cells. This recessive mutation makes the MaV203 cells not suited for the library screening. The proof of principle experiment has the same conditions as the positive selection step of the library screening using the wildtype EcTyrRS/tRNA$_{CUA}$^{Tyr} pair (discussed in detail in section 4.2.6). It was verified whether the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair can suppress amber codons in GAL4 and thus activate the URA3 gene in diploid cells. In literature it has been stated that the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair is orthogonal in *S. cerevisiae* for the suppression of amber codons by the incorporation of tyrosine[2,3,21]. This means that the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair does not interact with the endogenous translational machinery of yeast. The EcTyrRS does not aminoacylate any tRNA's from yeast, nor is the *E. coli* tRNA$_{CUA}$^{Tyr} aminoacylated by any aminoacyl tRNA synthetases from yeast. Additionality, with this experiment the orthogonality of the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair is confirmed as well. For the proof of principle experiment, four vectors are needed:

1. **pEcTyrRS/tRNA$_{CUA}$^{Tyr}**, containing the **wildtype** *E. coli* tyrosyl-tRNA synthetase (EcTyrRS) gene and the amber suppressor tRNA$_{CUA}$^{Tyr} gene (the construction of this vector is discussed in section 3.2.1)

2. **ptRNA$_{CUA}^{Tyr}$**: pESC-TRP vector containing the *E. coli* tRNA$_{CUA}^{Tyr}$ gene

3. **pEcTyrRS**: the pESC-TRP vector containing only the wildtype EcTyrRS gene behind the constitutive ADH1 promoter

4. **pADH1/GAL4_T44_R110**: pESC-LEU vector with the GAL4 gene containing two amber mutations (T44TAG and R110TAG). The GAL4 gene is expressed by the constitutive ADH1 promoter.

In order to test if diploid cells can be used for the library screening, the vectors pEcTyrRS/tRNA$_{CUA}^{Tyr}$ (1), ptRNA (2) and pEcTyrRS (3) are each co-transformed with the vector pADH1/GAL4_T44_R110 (4). pADH1/GAL4_T44_R110 contains the GAL4 gene with an amber codon at location 44 and 110. When both amber codons in GAL4 are suppressed, full-length GAL4 will be produced, activating the URA3 gene causing the MaV203 cells to grow on SD-URA. Since the EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair is orthogonal in yeast, growth should only take place if the vector pEcTyrRS/tRNA$_{CUA}^{Tyr}$ is combined with pADH1/GAL4_T44_R110. The other vector combinations should not result in successful suppression of the amber codons in GAL4 and therefore should not promote growth on SD-URA. The construction of the vectors needed for the orthogonality assay will be discussed in detail in section 3.4. The proof of principle experiment/orthogonality assay and its outcome will be discussed in detail in section 4.2.6.

3.4.1 Construction of the vectors ptRNA$_{CUA}^{Tyr}$ and pEcTyrRS

ptRNA$_{CUA}^{Tyr}$. The ptRNA$_{CUA}^{Tyr}$ vector was constructed by ligating the amber suppressor tRNA$_{CUA}^{Tyr}$ in the pESC-TRP vector. The *E. coli* tRNA$_{CUA}^{Tyr}$ gene was excised from the pCR2.1/ptRNA$_{CUA}^{Tyr}$ vector using the restriction enzymes AgeI and NheI (the construction of the pCR2.1/ ptRNA$_{CUA}^{Tyr}$ vector was described in section 3.2.1.). The plasmid pESC-TRP was cut with the same enzymes. After verifying that the restriction digestion was successful, the tRNA$_{CUA}^{Tyr}$ insert was successfully ligated into the pESC-TRP vector, resulting in the vector ptRNA$_{CUA}^{Tyr}$ (Figure 3.25).
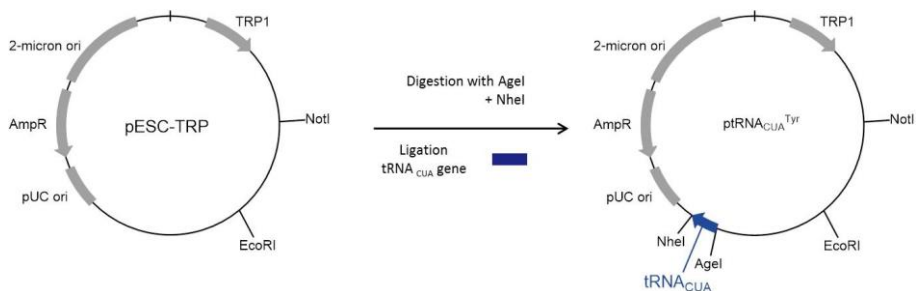
Figure 3.25. Schematic view of the ligation of the tRNA$_{CUA}$^{Tyr} gene in the pESC-TRP vector to form the ptRNA$_{CUA}$^{Tyr} vector.

pEcTyrRS. This vector contains the wildtype EcTyrRS gene behind the constitutive ADH1 promoter. The pEcTyrRS vector was constructed by ligating the ADH1 promoter in the vector pESC-TRP/EcTyrRS. Both the construction of the pESC-TRP/EcTyrRS vector as well as the amplification of the ADH1 promoter are discussed in section 3.2.1. First, the ADH1 promoter was excised from the pCR2.1/ADH1 vector using the restriction enzymes EcoRI and AgeI. The pESC-TRP/EcTyrRS vector was cut using the same enzymes. After verifying that the restriction digestion was successful, the ADH1 promoter was successfully ligated into the vector pESC-TRP/EcTyrRS, resulting in the pEcTyrRS vector (Figure 3.26).
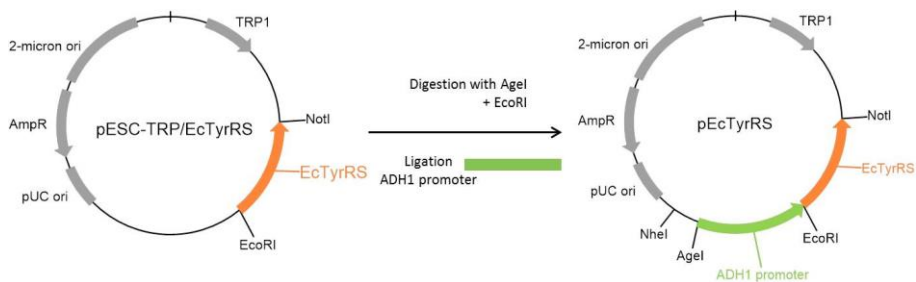
Figure 3.26. Schematic view of the ligation of the ADH1 promoter in the pESC-TRP/EcTyrRS vector to form the pEcTyrRS vector.

3.4.2 Construction of the pADH1/GAL4_T44_R110 vector

For the construction of the pADH1/GAL4_T44_R110 vector, the full-length GAL4 gene was amplified using the pCMV-GAL4 vector[22]. The primers used (Table A1 of the appendix) introduced a BamHI and BglII restriction site. After confirming on a 1% agarose gel that the PCR was successful, the GAL4 gene and the pESC-LEU/ADH1 vector (section 3.3.2) were digested with the restriction enzymes BamHI and BglII.

Thereafter the GAL4 gene was successfully ligated into the pESC-LEU/ADH1 vector, resulting in the vector pADH1/GAL4 (Figure 3.27). The PCR reaction mixture and program for ADH1 can be found in Table A2 and Table A6 and for GAL4 in Table A2 and Table A11 (appendix).
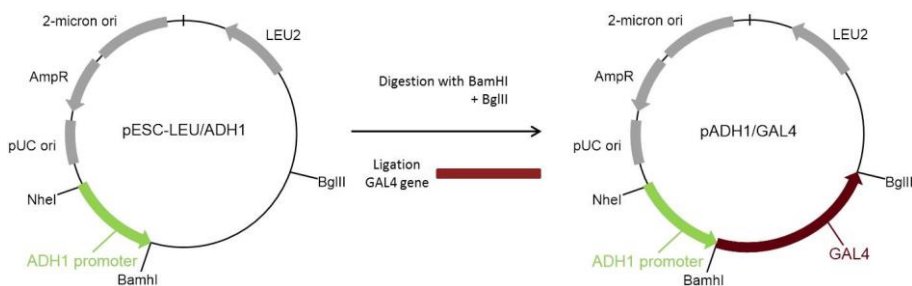
Figure 3.27. Schematic view of the ligation of the GAL4 gene in the pESC-LEU/ADH1 vector to form the pADH1/GAL4 vector.

Site-Directed Mutagenesis of GAL4. The pADH1/GAL4 vector contains the wildtype GAL4 gene. However, for the proof of principle assay and to test the orthogonality of the EcTyrS/tRNA$_{CUA}$^{Tyr} pair in yeast, two amber codons were introduced corresponding to threonine 44 and arginine 110[2]. A site-directed mutagenesis was carried out in two steps using the QuickChange Lightning Site-Directed Mutagenesis kit. In the first step the codon encoding threonine 44 was mutated to an amber codon (T44TAG). After confirming that this reaction was successful, a second mutagenesis reaction was done to change the codon encoding arginine 110 to an amber codon (R110TAG), resulting in the pADH1/GAL4_T44_R110 vector. The primers used for the mutagenesis are shown in Table A12. After the site-directed mutagenesis, sequencing (S11) confirmed that the amber codons were successfully introduced at the desired locations.

3.5 Conclusions

In this chapter the construction and composition of all the DNA constructs (vectors) used throughout this work was described.

For the synthesis of functionalized nanobodies (chapter 6) following vectors were constructed:

- **pEcTyrRS/tRNA$_{CUA}$^{Tyr}**: which contains the wildtype EcTyrRS/tRNA$_{CUA}$^{Tyr} pair.

- **pEcTyrRS(mutant)/tRNA$_{CUA}$^{Tyr}**: that contains the mutant EcTyrRS/tRNA$_{CUA}$^{Tyr} pair, with affinity for *p*-azidoPhe.
- **pTEF-MF/NbBCII10_His$_6$**: which contains the NbBCII10_His$_6$.
- **pTEF-MF/NbBCII10_TAG**: that contains the NbBCII10_His$_6$, with an amber codon located between the nanobody and the C-terminal His$_6$-tag.

For the protein expression experiments using GFP (chapter 6) following vectors were constructed:

- **pEcTyrRS/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t:** which contains the wildtype EcTyrRS/tRNA$_{CUA}$^{Tyr} pair. The activity of the tRNA$_{CUA}$^{Tyr} was elevated by the addition of the SNR52 promoter sequence and the SUP4 terminator sequence.
- **pEcTyrRS(mutant)/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t:** that contains the mutant EcTyrRS/tRNA$_{CUA}$^{Tyr} pair, with an affinity for *p*-azidoPhe. The activity of the tRNA$_{CUA}$^{Tyr} was elevated by the addition of the SNR52 promoter sequence and the SUP4 terminator sequence.
- **pADH1/GFP:** pESC-LEU vector with the wild type GFP gene.
- **pADH1/GFP_Y39TAG:** pESC-LEU vector with the GFP_Y39TAG gene, in which the tyrosine location 39 was substituted for an amber codon.
- **pTEF/GFP:** The pTEF-MF vector in which the MF leader sequence had been removed by the ligation of the wild type GFP gene, resulting in the cytosolic expression of GFP.
- **pTEF/GFP_Y39TAG:** The pTEF-MF vector in which the MF leader sequence had been removed by the ligation of the GFP_Y39TAG gene, resulting in the cytosolic expression of truncated GFP_Y39TAG.

For the experiments conducted in chapter 4, following vectors were constructed:

- **ptRNA$_{CUA}$^{Tyr}**: pESC-TRP vector containing the *E. coli* tRNA$_{CUA}$^{Tyr} gene.

- **pEcTyrRS**: pESC-TRP vector containing the wildtype EcTyrRS gene behind the constitutive ADH1 promoter.

- **pADH1/GAL4**: pESC-LEU vector with the wild type GAL4 gene, expressed by the constitutive ADH1 promoter.

- **pADH1/GAL4_T44_R110**: pESC-LEU vector with the GAL4 gene containing two amber mutations (T44TAG and R110TAG).

All the vectors needed in this project have been successfully constructed and have been verified by restriction analysis as well as DNA sequencing (section S of the appendix).

3.6 References

1.   Kowal, A K. *et al.*, Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria, *Proc. Natl. Acad. Sci. U. S. A.* **98,** 2268–2273 (2001).

2.   Chin, J. W. *et al.,* Progress Toward an Expanded Eukaryotic Genetic Code, *Chem Biol.* **10,** 511–519 (2003).

3.   Young, T. S. *et al.*, Beyond the canonical 20 amino acids: expanding the genetic lexicon, *J. Biol. Chem.* **285,** 11039–44 (2010).

4.   Deiters, A. *et al.,* Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*, *J. Am. Chem. Soc.* **125,** 11782–3 (2003).

5.   Gietz, R. D. *et al.*, Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure, *Yeast* **11,** 355–60 (1995).

6.   Vidal, M. *et al.*, Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions, *Proc. Natl. Acad. Sci. U.S.A.* **93,** 10315–20 (1996).

7.   Mullis, K. B. *et al.*, Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction, *Methods Enzymol.* **155,** 335–350 (1987).

8.   Bartlett, J. M. S. *et al.*, A short history of the polymerase chain reaction, *Methods Mol. Biol.* **226,** 3–6 (2003).

9.   Garret, R. H. & Grisham, C. M., *Biochemistry*. (Harris, David, 2005).

10. Berg, J. M., Tymoczko, J. L. & Stryer, L., *Biochemistry*. (Sara Tenney, 2007).

11. Pingoud, A. *et al.*, Type II restriction endonucleases: Structure and mechanism, *Cell. Mol. Life Sci.* **62,** 685–707 (2005).

12. Voytas, D. *et al.*, in *Curr. Protoc. Immunol.* (John Wiley & Sons, Inc., 2001). doi:10.1002/0471142735.im1004s02

13. Gellissen, G. *et al.,* Heterologous protein production in yeast, *Antonie Van Leeuwenhoek* **62,** 79–93 (1992).

14.   Lowe, T. M. *et al.*, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res.* **25,** 955–64 (1997).

15.   Schattner, P. *et al.*, The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs, *Nucleic Acids Res.* **33,** W686–W689 (2005).

16.   Deiters, A. *et al.,* Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*, *J. Am. Chem. Soc.* **125,** Supporting Information. 1–8 (2003).

17.   Wang, Q. *et al.*, Expanding the genetic code for biological studies, *Chem. Biol.* **16,** 323–36 (2009).

18.   Chen, S. *et al.*, An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in *Saccharomyces cerevisiae*, *J. Mol. Biol.* **371,** 112–22 (2007).

19.   Sakamoto, K. *et al.,* Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells, *Nucleic Acids Res.* **30,** 4692–9 (2002).

20.   Wang, Q. *et al.*, New methods enabling efficient incorporation of unnatural amino acids in yeast, *J. Am. Chem. Soc.* **130,** 6066–7 (2008).

21.   Edwards, H. *et al.*, A bacterial amber suppressor in *Saccharomyces cerevisiae* is selectively recognized by a bacterial aminoacyl-tRNA synthetase, *Mol. Cell. Biol.* **10,** 1633–41 (1990).

22.   Bowdish, K. S. *et al.*, Bipartite structure of an early meiotic upstream activation sequence from *Saccharomyces cerevisiae*, *Mol. Cell. Biol.* **13,** 2172-2181 (1993).

Chapter 4

Spontaneous derepression of the URA3 gene in haploid MaV203 yeast cells can be solved by the use of diploid cells

4.1 Introduction

The site-specific modification of NbBCII10 is based on the incorporation of *p*-azidophenylalanine (*p*-azidoPhe) by means of amber suppression. The principle of the amber suppression methodology is described in detail in section 2.2. For the incorporation of *p*-azidoPhe, the genetic code of *S. cerevisiae* is expanded with genes encoding for a mutant *E. coli* tyrosyl-tRNA synthetase (EcTyrRS)/tRNA$_{CUA}$^{Tyr} pair. The mutant EcTyrRS has key amino acids altered in its binding pocket, so that it has affinity for *p*-azidoPhe instead of tyrosine. In order to obtain such a mutant EcTyrRS, a library needs to be constructed and screened for the appropriate mutant EcTyrRS.

It was opted to construct and screen an own library of mutant EcTyrRS, not only to acquire more selective EcTyrRS mutants, but also to apply it later on other aaRSs. This library is generated by randomly mutating key amino acids present in the binding pocket of the EcTyrRS (Tyr37, Asp81, Tyr175, Gln179 and Asp182). The screening of such a library is performed in the haploid yeast strain MaV203[1] by means of a double-sieve method, based on the activation/repression of the GAL4 inducible URA3 gene. The MaV203 yeast strain is described in detail in section 2.3. In the first part of the current chapter, the double-sieve screening method[2,3] is explained in detail with a focus on the importance of the strict repression of the URA3 gene in the absence of GAL4.

However, it was observed that a fraction of the MaV203 cells are capable of growing on SD-URA, in the absence of GAL4. A series of experiments is performed and discussed in order to explain the observed unwanted activation of the URA3 gene. In the second part of this chapter, the results of a proof of principle/orthogonality assay using diploid yeast cells are evaluated.

4.1.1 Screening of a library of mutant EcTyrRSs in MaV203

In literature the procedure to construct and screen a library of mutant EcTyrRS is described[2-7]. According to literature one starts with the preparation of a library of EcTyrRS mutants in *E. coli*. The library of mutant EcTyrRSs is then transformed into the haploid yeast strain MaV203, together with the tRNA$_{CUA}^{Tyr}$ gene and the GAL4 gene, the latter having two amber codons. The main characteristic of the MaV203 strain is that it contains deletions of the endogenous GAL4 and GAL80 genes and that three GAL4-inducible reporter genes, *GAL1::LacZ*, *HIS3$_{UASGAL1}$::HIS3* and *SPAL10::URA3*[1,8] are present. Therefore, in the absence of GAL4, MaV203 is auxotroph for both histidine and uracil, meaning that histidine and uracil are not produced and need to be added to the growth medium in order for the cells to survive/grow. In addition, the MaV203 cells are also auxotroph for leucine and tryptophan, allowing for the selection of successful transformation with plasmids containing the *LEU2* or *TRP1* genetic marker. The *SPAL10::URA3* is counter selective when GAL4 is present, *i.e.* it can be used for positive as well as negative selection (described in section 2.3). The MaV203 cells are transformed with two vectors (illustrated in Figure 4.1).

1. The first vector contains the amber suppressor tRNA$_{CUA}^{Tyr}$ and the mutant EcTyrRS library.

2. The second vector contains a GAL4 gene, with two amber codons.
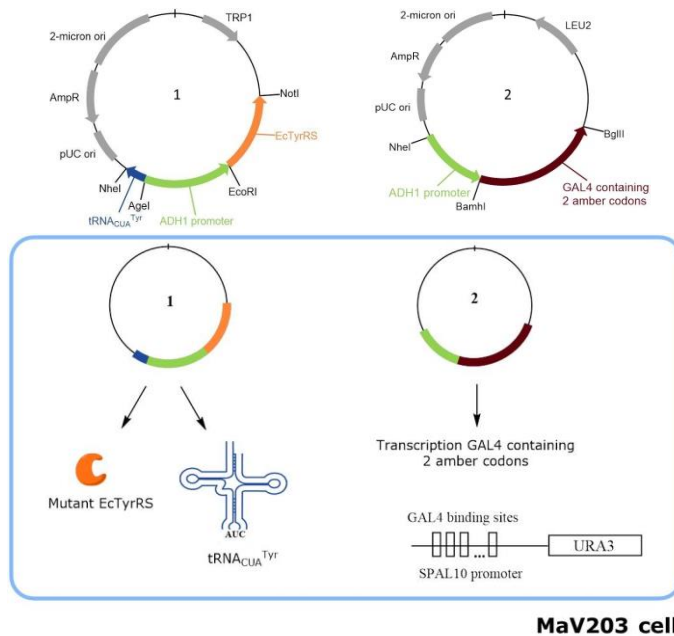
Figure 4.1. MaV203 cells transformed with two vectors. Vector 1 expresses the amber suppressor tRNA$_{CUA}$^{Tyr} and a mutant EcTyrRS library. Vector 2 expresses GAL4 that contains two amber codons.

Positive selection In the positive selection step, the MaV203 cells are grown in SD-URA containing the non-natural amino acid *p*-azidoPhe. Clones containing a mutant EcTyrRS capable of charging the tRNA$_{CUA}$^{Tyr} with either the non-natural amino acid or the natural tyrosine (Figure 4.2A), will be able to suppress the amber codons in GAL4. This will result in the production of full-length GAL4, which in turn will activate the URA3 gene, causing the cells to grow on SD-URA. Clones containing an inactive mutant EcTyrRS (Figure 4.2B) will not suppress the amber codons in GAL4 and will therefore not grow on SD-URA. As a result, after the first positive selection step, only clones expressing mutant EcTyrRS selective for either the non-natural amino acid or for the natural tyrosine remain.
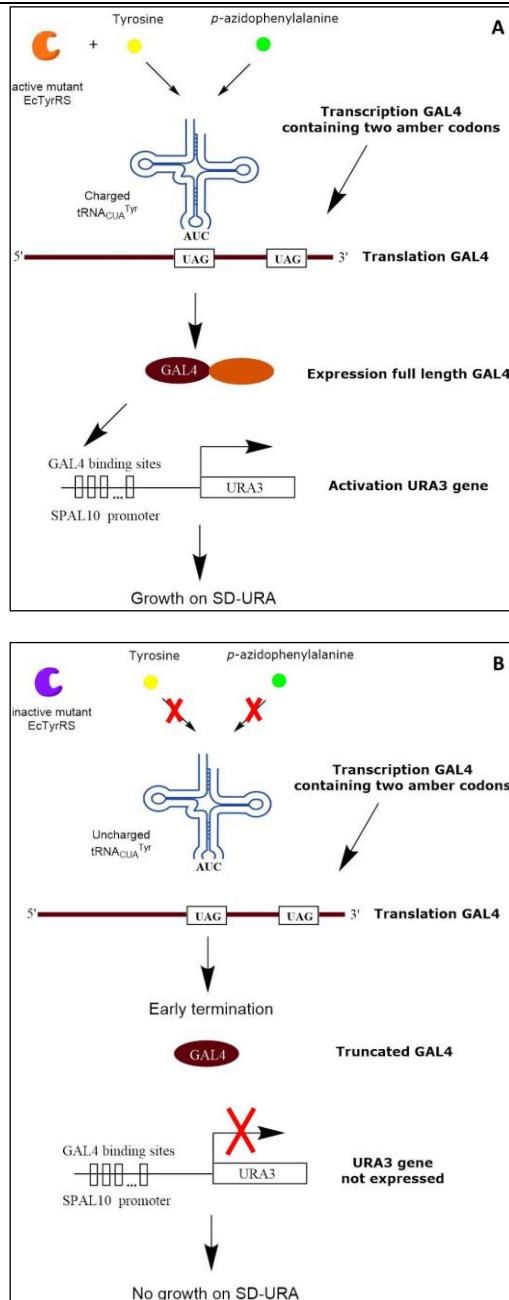
Figure 4.2. Positive selection for amber suppression: (A) MaV203 cells containing an active EcTyrRS mutant capable of charging the $tRNA_{CUA}^{Tyr}$ with either tyrosine or p-azidoPhe will express full-length GAL4, activate the URA3 gene and will be able to grow on SD-URA. (B) MaV203 cells containing an inactive EcTyrRS mutant will not express full-length GAL4 and will not express the URA3 gene, resulting in no growth on SD-URA.

Negative selection In the negative selection step, the cells that were obtained from the positive selection are grown in full medium containing uracil (YPD = yeast extract peptone dextrose, complete medium) and 5-fluoroorotic acid (5-FOA), but lack the non-natural amino acid. These cells contain a mutant EcTyrRS capable of coupling either tyrosine or *p*-azidoPhe. The gene product of the URA3 gene, orotidine-5'-phosphate decarboxylase, converts 5-FOA to the toxic compound 5-fluorouracil. Clones containing a mutant EcTyrRS capable of charging the $tRNA_{CUA}^{Tyr}$ with natural tyrosine (Figure 4.3A), will suppress the amber codons in GAL4 and thus express the URA3 gene. This will result in the conversion of 5-FOA to the toxic 5-fluorouracil, causing cell death. Clones containing a mutant EcTyrRS specific for *p*-azidoPhe (Figure 4.3B) will survive this negative selection step. Since no *p*-azidoPhe is present in the medium, the mutant EcTyRS cannot charge any $tRNA_{CUA}^{Tyr}$. Therefore no full-length GAL4 is expressed and the URA3 gene remains inactive.
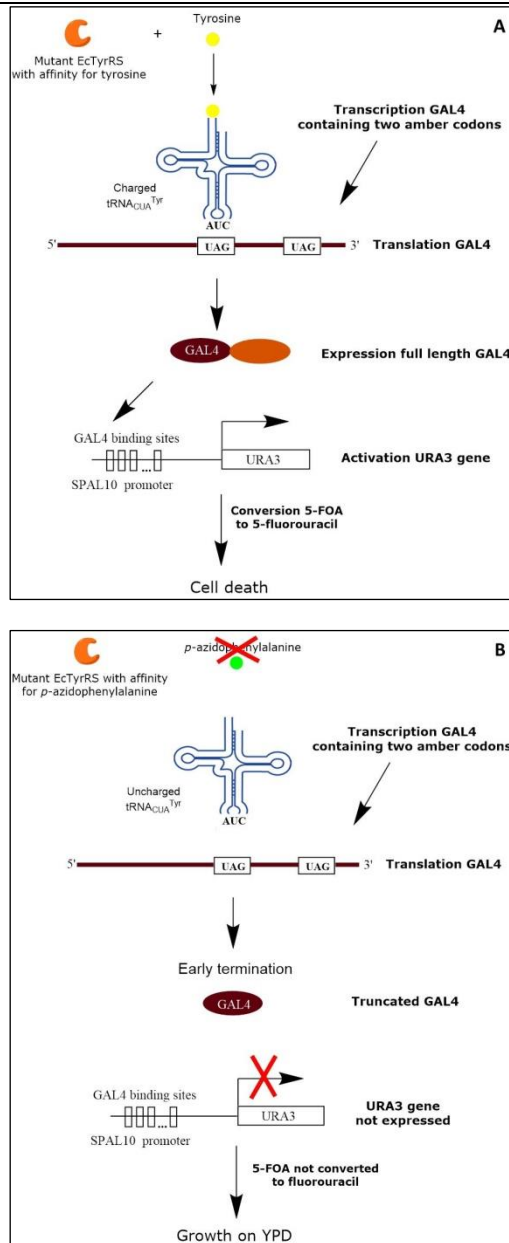
Figure 4.3. Negative selection for amber suppression: (A) Clones containing mutant EcTyrRS capable of charging the tRNA$_{CUA}^{Tyr}$ with natural tyrosine will suppress the amber codons in GAL4 and thus express the URA3 gene, resulting in the conversion of 5-FOA to the toxic 5-fluorouracil, causing cell death. (B) Clones containing a mutant EcTyrRS specific for the non-natural amino acid will survive this negative selection step. Since no *p*-azidoPhe is present in the medium, they cannot charge any tRNA$_{CUA}^{Tyr}$. Therefore no full-length GAL4 is expressed and the URA3 gene thus stays inactive.

Since the double-sieve method is based on the activation of the URA3 gene in case the full-length GAL4 is expressed, it is important that the URA3 gene stays actively repressed in the absence of GAL4. Most yeast promoters, fused to URA3, confer a Ura3$^+$ phenotype if no specific transcription factor is present[9,10]. This means that basal levels of URA3 expression are sufficient to promote growth on a medium lacking uracil and to induce sensitivity for 5-fluoroorotic acid (5-FOA).

For this reason, the URA3 gene is fused to the SPAL10 promoter in the MaV203 strain[1]. This SPAL10 promoter is a derivative of the SPO13 promoter, containing 10 GAL4 binding sites (Figure 4.1). The SPO13 promoter (and therefore the SPAL10 promoter) is only active under sporulation conditions[11] and is tightly repressed by the Rpd3 Histone DeAcetylase Complex (Rpd3 HDAC) under normal conditions[12–16]. Fusing the URA3 gene to the SPAL10 promoter thus confers a very strict Ura$^-$ and 5-FOA resistant phenotype under normal growth conditions. Consequently, the GAL4 binding sites in the SPAL10 promoter allow the activation of the promoter when GAL4 is present.

Interestingly, although designed to be very strict URA$^-$, our experiments showed that a small fraction of MaV203 cells was capable of growing on SD-URA medium, under normal conditions, even though GAL4 was not expressed. In order to address this issue of unwanted growth, a series of experiments was set up and performed.

4.2 Study of the origin of the unwanted URA3 activation

Several factors such as the presence of GAL4, the non-specific activation of the SPAL10 promoter and the nature of possible mutations (whether recessive or not) were investigated for MaV203 haploid cells. Furthermore, the stability of the SPAL10 promoter and the use of diploid cells were evaluated as a possible solution for the unwanted activation of the URA3 gene.

4.2.1 GAL4 is not present in the growth medium, nor produced in MaV203 cells

Since the URA3 gene is fused to the GAL4-inducible promoter SPAL10, the first hypothesis explaining the unwanted growth on SD-URA is the presence of GAL4 in the growth media or in the cells. To test for this, a β-galactosidase assay, more specific a Colony-Lift Filter Assay[17], was

performed. Using this assay the activity of the *LacZ* gene is evaluated. Since MaV203 cells have the *LacZ* gene behind a GAL4-inducible promoter, the *LacZ* gene will be expressed, resulting in the hydrolysis of 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal) when GAL4 is present, and the development of blue colonies. MaV203 cells transformed with pADH1/GAL4 were used as a positive control. The vector pADH1/GAL4 (described in section 3.4.2) contains full-length GAL4 behind a constitutive ADH1 promoter. These cells will produce GAL4, activate the *LacZ* gene and turn blue after the assay.
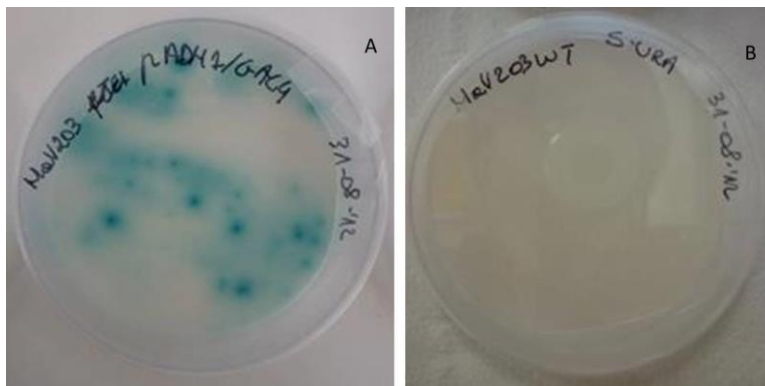
Figure 4.4. Results of the β-galactosidase assay performed on MaV203 cells containing (A) the vector pADH1/GAL4 (positive control), (B) untransformed MaV203 cells capable of growing on SD-URA. The fact that the untransformed cells (B) are not turning blue indicates that GAL4 is not present and can therefore not be responsible for the activation of the URA3 gene.

The β-galactosidase assay was performed in triplicate confirming the reproducibility of the results (Figure 4.4). From Figure 4.4A it is clear that the cells, transformed with the vector pADH1/GAL4, expressed GAL4 and turns blue. Figure 4.4B shows that the untransformed MaV203 cells, capable of growing on SD-URA, did not turn blue. This indicates that there is no GAL4 present in the medium, nor produced by the MaV203 cells themselves to activate the SPAL10 promoter.

4.2.2 The SPAL10 promoter is inefficiently repressed

Another explanation for the unwanted URA3 expression is that in a fraction of the cells, the SPAL10 promoter is not actively repressed and therefore is responsible for the growth on SD-URA plates. To test this hypothesis, an experiment was conducted in which a liquid overnight culture of untransformed MaV203 cells was plated on YPD (positive control); YPD+0.1% 5-Fluoroorotic Acid (5-FOA) and SD-URA plates.

Only cells with an active URA3 gene should be able to grow on SD-URA plates but the same URA3 gene product would also convert 5-FOA to the toxic 5-fluorouracil. Thus, if the URA3 gene is actively repressed in all the MaV203 cells, no growth should be visible on SD-URA. However, if a fraction of the cells has an active URA3 gene, growth should be visible on both the YPD+0.1% 5-FOA and SD-URA plates.

From Figure 4.5, it is clear that growth was visible on YPD (Figure 4.5A, positive control) and on both the YPD+0.1% 5-FOA (Figure 4.5B) and SD-URA plates (Figure 4.5C). This result suggests that in a fraction of the cells the URA3 gene is expressed, allowing growth on SD-URA, while other cells (visible on YPD+0.1% 5-FOA plates) have the URA3 gene successfully repressed. If this is the case, a simple solution to obtain MaV203 cells with a strict URA3⁻ phenotype would be to select a single colony which is capable of growing on YPD+0.1% 5-FOA and use it for the screening of libraries of mutant EcTyrRSs.
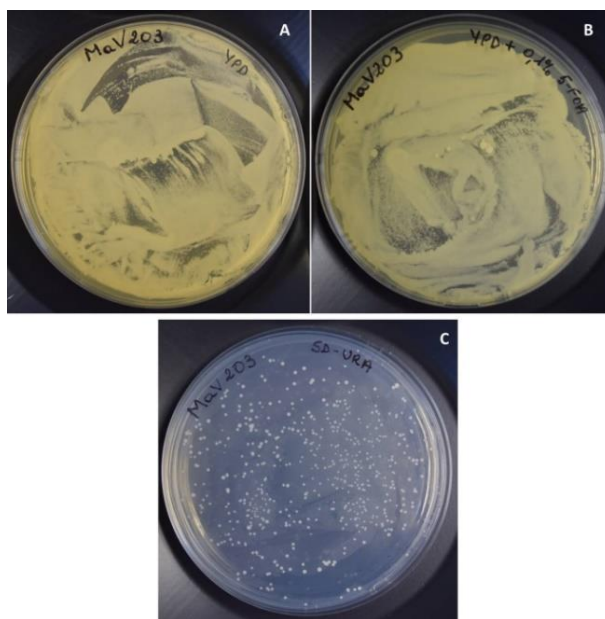
Figure 4.5. Verifying the repression of the SPAL10 promoter. MaV203 cells plated on (A) YPD medium (positive control), (B) YPD + 0.1% 5-FOA and (C) SD-URA. Growth is clearly seen on all media, indicating that a fraction of the cells have an active URA3 gene (C), while other cells have the URA3 gene successfully repressed (B).

It was investigated if this ability to grow on SD-URA is present in a
certain fraction of cells or if it is acquired after a number of cell
divisions. To test this, cells were selected from a YPD + 0.1% 5-FOA
plate (URA3 gene is actively repressed) and allowed to grow overnight
in YPD, followed by plating on YPD as well as on SD-URA.

Again growth was visible on both YPD and SD-URA plates, indicating
that after a number of cell divisions some cells lose their ability to
repress the URA3 gene. This suggests that although the SPAL10
promoter initially is actively repressed, spontaneous mutations occur,
most likely in the SPO13 (and thus also SPAL10) repressing genes. This
results in a derepression and thus basal expression of the URA3 gene. In
a personal communication with Prof. Dr. Vidal and Prof. Dr Boeke, who
developed the MaV203 strain in 1996, they hypothesized that this might
be the result of a mutation in the gene of a protein involved in Rpd3
HDAC.

The observation that these cells apparently undergo spontaneous
mutations leading to basal expression of the URA3 gene, resulted in a
number of questions. First, is the mutation dominant or recessive and
hence, in case of the latter, can diploid cells be used to address the
problem? Second, in case the mutation is recessive, is the growth on
SD-URA the result of a mutation in a single gene, or is it possible that a
mutation in different genes gives rise to the same phenotype.

4.2.3 Mutations in SPAL10 repressing genes are recessive

To determine whether a recessive or a dominant mutation is causing the
activation of the URA3 gene, diploid cells were constructed. Figure 4.6
shows a comparison between haploid and diploid cells and explains how
diploid cells can be used to determine whether a mutation is dominant
or recessive. The diploid cells are formed by the transformation of
MaV103(a) and MaV203(α) cells, with a vector containing a different
auxotrophic marker, *i.e.* pESC-LEU or pESC-TRP, respectively.
Subsequently the cells are plated on double drop-out media (SD-LEU-
TRP)[17]. Only the resulting diploid cells will be able to grow, since they
contain both the pESC-LEU and pESC-TRP vector. On the other hand,
the haploid cells present in the mixture either contain only pESC-LEU or
pESC-TRP, and are not able to grow on the SD-LEU-TRP plates.
Experiments were done in fourfold.
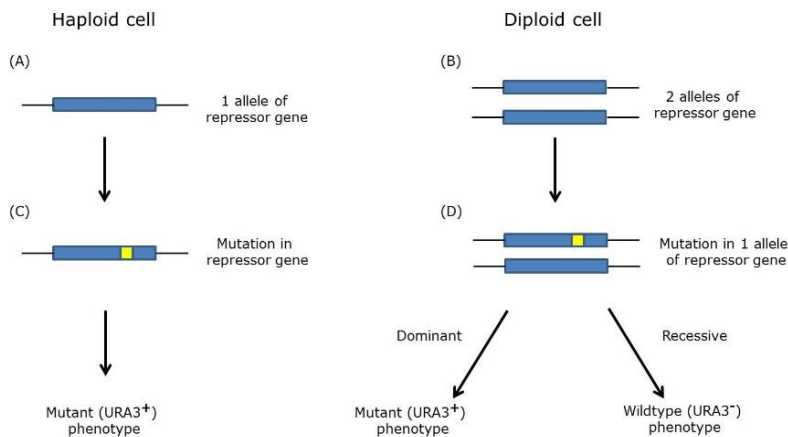
Figure 4.6. Comparison of a haploid and a diploid yeast cell. (A) Haploid cells contain one allele of the repressor gene while (B) diploid cells contain two alleles of the same repressor gene. (C) When a mutation occurs in the repressor gene of the haploid cell (C) it will result in a mutant phenotype (*i.e.* URA[+] phenotype). (D) However in diploid cells, when a mutation occurs in one allele of the repressor gene, a wildtype allele will still be present. Therefore, when the mutation is dominant it will result in a mutant phenotype, whereas if the mutation is recessive, the wildtype phenotype remains.

Table 4.1 represents the growth conditions for the haploid cells used to obtain the diploid cells. Based on their ability to grow on YPD+ 0.1% 5-FOA or SD-URA, it is possible to determine whether their URA3 gene is active or not. Growth on YPD+0.1% 5-FOA is only possible for cells that do not express the URA3 gene, while growth on SD-URA is only possible when the URA3 gene is expressed. Table 4.2 represents the combinations of haploid cells used to form diploid cells. Figure 4.7 visualizes the results of plating the diploid cells on both SD-LEU-TRP and SD-LEU-TRP-URA.

Table 4.1. Growth conditions for the haploid cells used for the formation of diploid cells. Since the identity of the repressor gene, *i.e.* which mutation is causing the derepression of the URA3 gene, is unknown it is referred to as "GENE" for the wildtype form and "gene" for the mutant form.

	Haploid strain	Growth medium	URA3 gene	Repressor gene
1	MaV203(α)	YPD+ 0.1% 5-FOA	not active	GENE
2	MaV103(a)	YPD+ 0.1% 5-FOA	not active	GENE
3	MaV203(α)	SD-URA	active	gene
4	MaV103(a)	SD-URA	active	gene

Table 4.2: Combination of haploid cells used to produce diploid cells.

	MaV103/MaV203	Growth on SD-LEU-TRP	Growth on SD-LEU-TRP-URA
1+2 (Figure 4.7 A and B)	GENE/ GENE	Plate covered with colonies (A)	No colonies present (B)
2+3 (Figure 4.7 C and D)	GENE/ gene	Plate covered with colonies (C)	Small amount of colonies (D)
1+4 (not shown)	gene/ GENE	Plate covered with colonies	Small amount of colonies
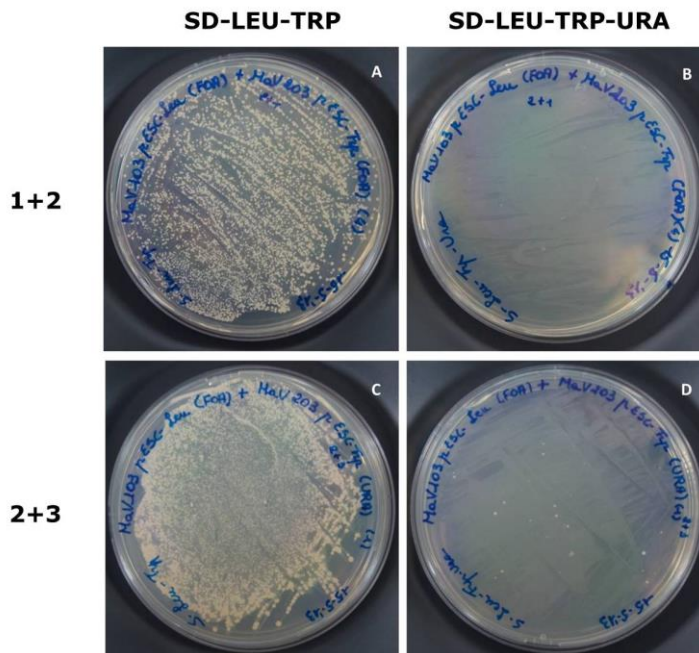
Figure 4.7. Diploid cells resulting from **1+2** plated on (A) SD-LEU-TRP and (B) SD-LEU-TRP-URA, diploid cells resulting from **2+3** plated on (C) SD-LEU-TRP and (D) SD-LEU-TRP-URA. The results obtained here suggest that the mutation, responsible for the unwanted URA3 expression, is recessive in nature. In case the mutation would be dominant, an equal amount of colonies would be present on (C) and (D).

Figures 4.7(A) and (B) show the result of mating MaV103 and MaV203 (**1**+**2**, both grown on YPD+0.1% 5-FOA). Figure 4.7A shows the diploid cells plated on SD-LEU-TRP. Colony formation confirms that the mating between the haploid cells was successful, since only diploid cells contain the pESC-LEU vector as well as the pESC-TRP vector. Figure 4.7B shows the same diploid cells on SD-LEU-TRP-URA, where no colonies are visible. The absence of colony formation on SD-LEU-TRP-URA indicates that the URA3 gene is actively repressed in all the diploid cells. The reproducibility of the results was confirmed for all four experiments with this combination of cells.

In general, it is assumed that the mutation rate in diploid cells is identical to haploid cells. The fact that no colonies are formed on the SD-LEU-TRP-URA plates (Figure 4.7B) is an indication that the mutation that occurs is recessive in nature. If a recessive mutation occurs in one allele, the diploid cells still have one wild type allele to complement for this mutation, and keep the URA3 gene repressed (Figure 4.6). If on the other hand the mutation is dominant in nature, colonies would also have formed on the SD-LEU-TRP-URA plate. Figures 4.7 (C) and (D) show the results of mating MaV103 (**2** from Table 4.1, grown on YPD+0.1% 5-FOA) with MaV203 (**3** from Table 4.1, grown on SD-URA). Again the large number of cells on SD-LEU-TRP (Figure 4.7C) confirms that the diploid cells are successfully formed. On SD-LEU-TRP-URA (Figure 4.7D) a small number of colonies (about 30 colonies) is visible. In case the mutation causing growth on SD-URA was dominant, the same amount of colonies was expected to form on SD-LEU-TRP-URA (D) as on SD-LEU-TRP (C), since all the diploid cells contain one mutated allele of the repressor gene to start with. The fact that only a small number of colonies is formed on SD-LEU-TRP-URA confirms that the mutation responsible for the derepression of the URA3 gene, is recessive. The diploid cells (**2**+**3**) contain only one wild type copy of the gene (GENE) responsible for repressing the URA3 gene, since one of the haploids (**3**) was able to grow on SD-URA. The few colonies that are formed on SD-LEU-TRP-URA are the diploid cells that obtained a mutation in the second allele of the repressor gene during their growth.

4.2.4 The SPAL10 promoter remains stable/inducible in diploid MaV203 cells

The results above indicate that recessive mutations are responsible for the spontaneous expression of the URA3 gene. Therefore, it was investigated if diploid cells can be used to prevent the spontaneous URA3 derepression. The stability of the SPAL10 promoter in diploid cells was compared with its stability in haploid cells. Both diploid cells and haploid (MaV103 and MaV203) cells were plated on YPD+ 0.1% 5-FOA to select for the cells with active repression of the URA3 gene. Hereto, liquid cultures were inoculated with single colonies in YPD and grown overnight at 30 °C. The cultures were allowed to grow to an absorbance of 7 (OD_{600}) and subsequently diluted to an absorbance of 0.1. This process was repeated for four days. On the fifth day the cultures were diluted to an absorbance of 0.5 and plated on both YPD and SD-URA plates. Since the diploid cells as well as the haploid cells were diluted from an absorbance of 7 to an absorbance of 0.1 on a daily basis, both had undergone the same amount of cell divisions. Moreover, by diluting to an absorbance of 0.5 on the fifth day, the same amount of cells was plated for the diploid and haploid cells.
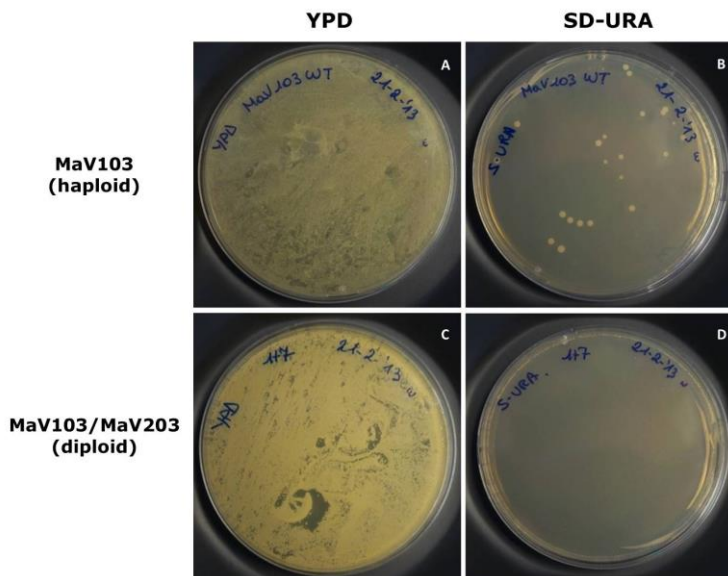
Figure 4.8. Verifying the stability of the SPAL10 promoter in diploid cells. Haploid MaV103 cells plated on (A) YPD and (B) SD-URA. Diploid cells plated on (C) YPD and (D) SD-URA. It is clear that only the haploid cells regain the ability to grow on SD-URA after a number of cell divisions (B). Diploid cells keep repressing the URA3 gene and no colony formation is observed on SD-URA (D).

From Figure 4.8 it is clear that after a few cell divisions, a small fraction
of the haploid cells (**2** from Table 4.1) are capable of growing on SD-
URA (Figure 4.8B). The diploid cells (mating of **1** and **2** Table 4.1) on
the other hand, did not form any colonies (Figure 4.8D), meaning that
the URA3 gene remains actively repressed. This indicates that the
SPAL10 promoter is stable in diploid cells and that diploid cells can be
used to circumvent the problem of unwanted URA3 expression. This is
also an additional proof that the mutation causing the activation of the
URA3 gene is recessive. A dominant mutation would result in colony
formation of the diploid cells on SD-URA as well.

4.2.5 The spontaneous derepression is the result of mutations in a single gene

After showing that the mutations, responsible for the activation of the
URA3 gene, are recessive, it is of interest to know whether the
mutations occur in one single gene or if a mutation in different genes
results in a URA3$^+$ phenotype. Therefore, a complementation analysis
was performed. In a complementation analysis two haploid cells,
containing a recessive mutation (*i.e.* capable of growing on SD-URA),
are allowed to mate and form diploid cells. If both haploid cells contain
the mutation in the same gene, the resulting diploid cells will have the
same mutant phenotype. However, if the haploid cells have a mutation
in different genes, these will complement for each other's recessive
mutation and the resulting diploid cells will have a wildtype phenotype.
In this context, four independent colonies of MaV103, capable of
growing on SD-URA, were each allowed to mate with four different
clones of MaV203, also capable of growing on SD-URA. Thus, both
mating types have an active URA3 gene. If the mutation is present in
different genes, at least in some combinations diploid cells should be
formed that contain different mutations in opposite alleles. Such
mutations would complement each other and repress the URA3 gene
resulting in a URA3$^-$ phenotype. Figure 4.9 shows an example of the
diploid cells formed. Identical results were obtained for all the mating
experiments. It was noted that for all combinations the diploid cells
retained a URA3$^+$ phenotype. This indicates that they all contain two
copies of the same mutant gene, suggesting that mutations responsible
for the URA3$^+$ phenotype are present in a single gene.
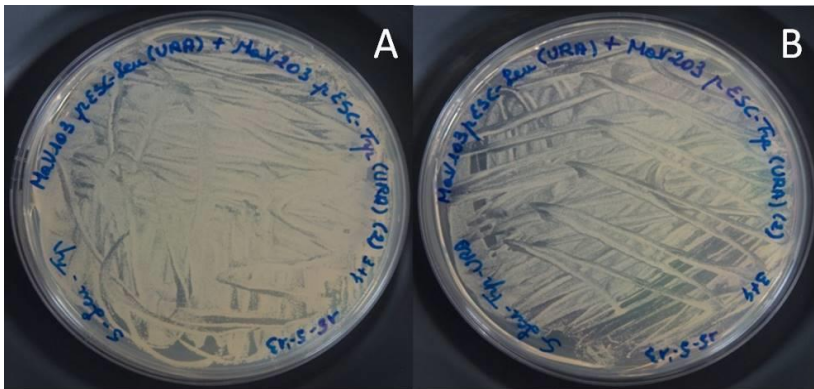
Figure 4.9. Complementation analysis. Example of the diploid cells resulting from the complementation analysis, plated on (A) SD-LEU-TRP and (B) SD-LEU-TRP-URA. Growth was clearly visible on both SD-LEU-TRP and SD-LEU-TRP-URA indicating that the mutation causing the derepression of the URA3 gene is present in a single gene. The same results were obtained for all the combinations of MaV103 and MaV203.

4.2.6 Diploid MaV103/MaV203 cells allow for selection of active EcTyrRS: proof of principle assay

In this section it is verified if diploid cells (formed by the mating of haploid MaV103 (a) and MaV203 (α) cells)[17] can be used instead of haploid cells for the screening of libraries of mutant EcTyrRSs. For this proof of principle, an experiment was performed with the same conditions as the positive selection step of the double-sieve screening method, using the wildtype EcTyrRS/tRNA$_{CUA}$^{Tyr} pair (Figure 4.10). It is examined whether the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair can suppress amber codons in GAL4 and thus activate the URA3 gene in diploid cells. In literature it has been demonstrated that the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair is orthogonal in *S. cerevisae*, suppressing amber codons by incorporating tyrosine[2,18,19]. The orthogonality of the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair is confirmed in this experiment.

For this proof of principle assay five vectors were used:

I. **pEcTyrRS/tRNA$_{CUA}$^{Tyr}**, pESC-TRP vector containing the *wildtype E. coli* tyrosyl-tRNA synthetase (EcTyrRS) gene and the amber suppressor tRNA$_{CUA}$^{Tyr} gene (*TRP1* as auxotrophic marker → permits growth on SD-TRP).

II. **ptRNA$_{CUA}$^{Tyr}**: pESC-TRP vector containing the *E. coli* tRNA$_{CUA}$^{Tyr} gene (*TRP1* as auxotrophic marker).

III. **pEcTyrRS**: pESC-TRP vector containing the wildtype EcTyrRS gene behind the constitutive ADH1 promoter (*TRP1* as auxotrophic marker).

IV. **pADH1/GAL4_T44_R110**: pESC-LEU vector with the GAL4 gene containing two amber mutations (T44TAG and R110TAG) (*LEU2* as auxotrophic marker → permits growth on SD-LEU).

V. **pECS-LEU**: Empty vector (*LEU2* as auxotrophic marker)

The construction and details on the composition of these vectors are discussed in chapter 3. In order to test if diploid cells can be used for the library screening, each of the vectors (**I**), (**II**) and (**III**) were cotransformed with vector (**IV**). As shown in Figure 4.10, successful suppression of the amber codons in GAL4 results in the expression of full-length GAL4 and the activation of the URA3 gene. This means that, on growth medium without uracil, colony formation should only occur when vector (**I**) (containing the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair) is combined with vector (**IV**) (containing the GAL4 gene with the two amber codons). In contrast, no cell growth is expected for the other vector combinations since the absence of either tRNA$_{CUA}$^{Tyr} or EcTyrRS should result in a non-successful amber suppression, leading to a premature termination of the GAL4 gene and the absence of activation of the URA3 gene.
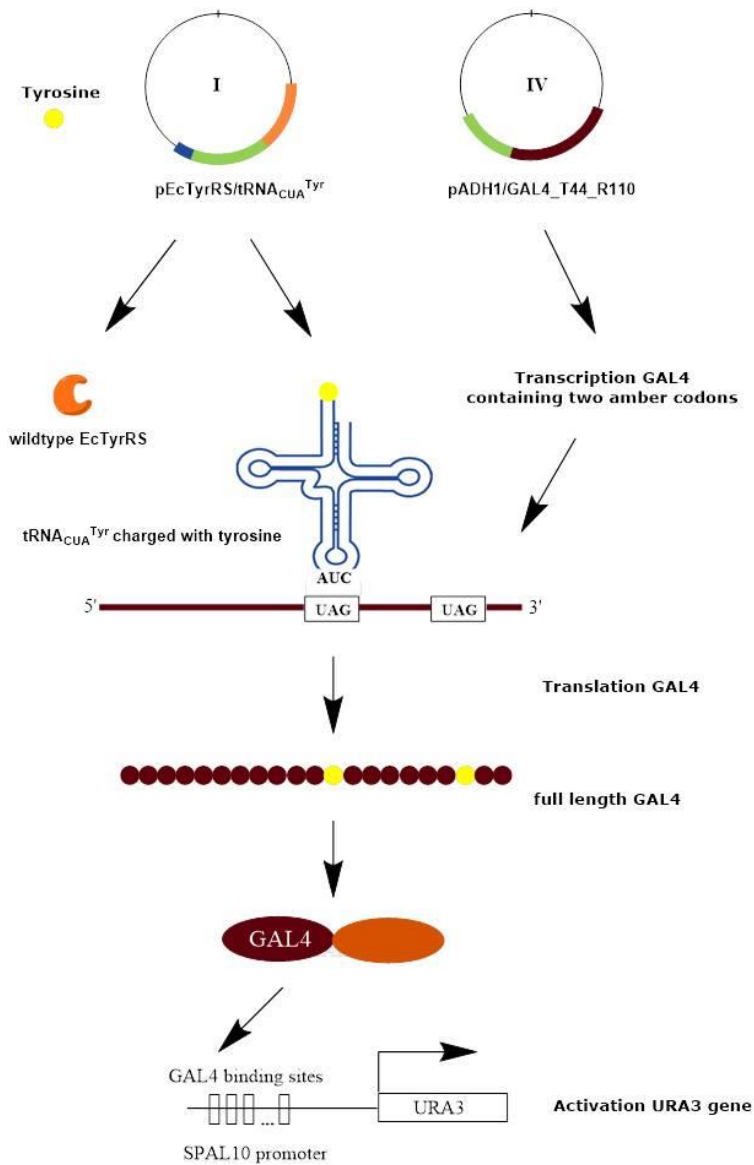
Figure 4.10. Overview of the assay to demonstrate the proof of principle. The diploid yeast cells contain both vector (**I**) (pEcTyrRS/tRNA$_{CUA}^{Tyr}$) and vector (**IV**) (pADH1/GAL4_T44_R110) resulting in a successful amber suppression and expression of full-length GAL4, which will activate the URA3 gene. In case either the tRNA$_{CUA}^{Tyr}$ or EcTyrRS is absent, no amber suppression takes place and the URA3 gene remains inactive.

Before the mating process, the haploid MaV103 (a) cells were transformed with either:

    I.   pEcTyrRS/tRNA$_{CUA}$^{Tyr} or

   II.   ptRNA$_{CUA}$^{Tyr} or

 III.   pEcTyrRS

The haploid MaV203 (α) cells were transformed with either:

   IV.   pADH1/GAL4_T44_R110   or

    V.   pESC-LEU (negative control)

Successful mating between MaV103 and MaV203 results in diploid cells capable of forming colonies on SD-LEU-TRP, since only diploid cells contain both vectors. In case of successful amber suppression in the diploid cells, full-length GAL4 is expressed activating the URA3 gene and thus promoting growth on SD-LEU-TRP-URA.
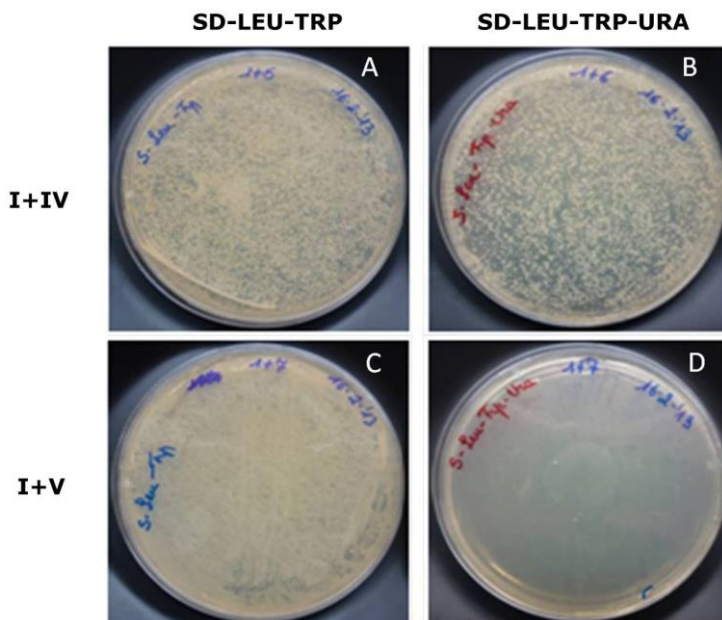
Figure 4.11. Proof of principle assay. Diploid cells formed by mating of MaV103(**I**):(pEcTyrRS/tRNA$_{CUA}$^{Tyr}) + MaV203(**IV**):(pADH1/GAL4_T44_R110) plated on (A) SD-LEU-TRP and (B) SD-LEU-TRP-URA and of MaV103(**I**):(pEcTyrRS/tRNA$_{CUA}$^{Tyr}) + MaV203(**V**):(pESC-LEU) plated on (C) SD-LEU-TRP and (D) SD-LEU-TRP-URA. Colony formation on (B) indicates that the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair is able to suppress the amber codons in GAL4, causing the expression of full-length GAL4, which activates the URA3 gene.

Figure 4.11 shows growth behavior of the diploid cells after the mating of: MaV103(**I**) with MaV203(**IV**) and MaV103(**I**) with MaV203(**V**). It can be seen that both the mating between MaV103(**I**) and MaV203(**IV**) (Figure 4.11A) as well as the mating between MaV103(**I**) and MaV203(**V**) (Figure 4.11C) resulted in diploid cells. Only diploid cells contain both vectors and are able to form colonies on SD-LEU-TRP.

Also, it is clear that only diploid cells containing vector **I** and vector **IV** are capable of growing on SD-LEU-TRP-URA (Figure 4.11B). The diploid cells that contain vector **I** and **V** are not able to grow on SD-LEU-TRP-URA (Figure 4.11D). This demonstrates that the EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair is able to suppress the amber codons in GAL4, causing the expression of full-length GAL4, activating the URA3 gene. Figure 4.12 shows growth of the diploid cells after the mating of MAV103(**II**) with MaV203(**IV**) and MAV103(**II**) with MaV203(**V**).
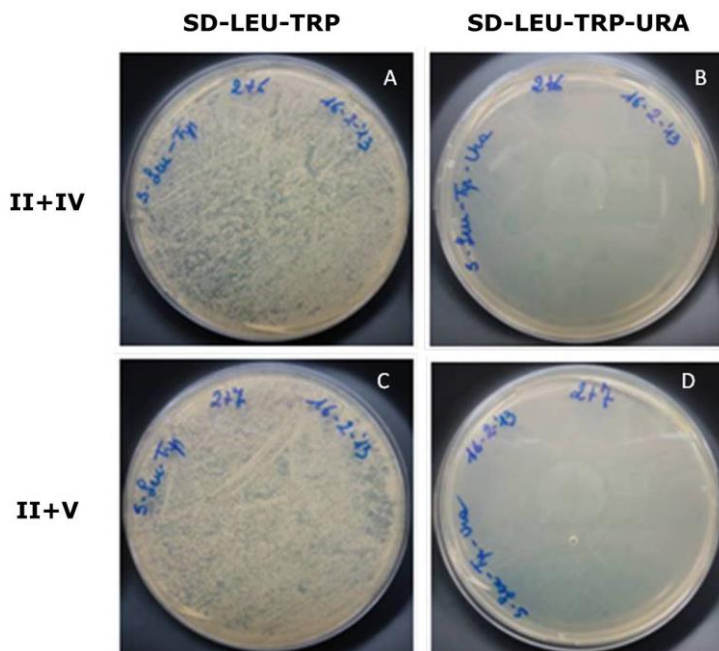
Figure 4.12. Proof of principle assay. Diploid cells formed by mating of: MaV103(**II**):(ptRNA$_{CUA}^{Tyr}$) + MaV203(**IV**): (pADH1/GAL4_T44_R110) plated on (A) SD-LEU-TRP and (B) SD-LEU-TRP-URA and of MaV103(**II**):(ptRNA$_{CUA}^{Tyr}$) + MaV203(**V**):(pESC-LEU) plated on (C) SD-LEU-TRP and (D) SD-LEU-TRP-URA. The lack of colony formation on (B) indicates that tRNA$_{CUA}^{Tyr}$ is not able to suppress the amber codons in GAL4, if EcTyrRS is not present.

From Figure 4.12, it can be observed that both the mating between
MaV103(**II**) and MaV203(**IV**) (Figure 4.12A) as well as the mating
between MaV103(**II**) and MaV203(**V**) (Figure 4.12C) resulted in diploid
cells. It is also clear that both types of diploid cells are not capable of
growing on SD-LEU-TRP-URA (Figure 4.12B and D). This means that the
amber codons are not suppressed in the GAL4 gene. Thus, no full-length
GAL4 is formed and the URA3 gene remains repressed.

SD-LEU-TRP   **SD-LEU-TRP-URA**

III+IV

III+V
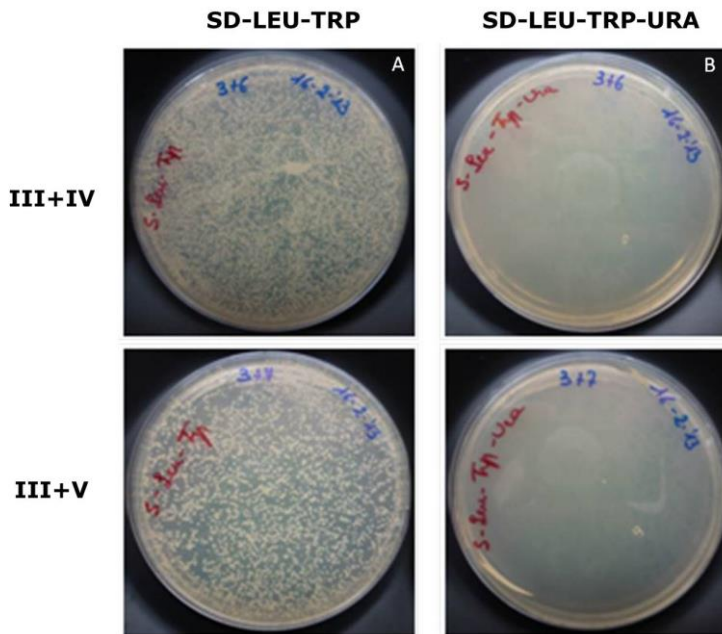
Figure 4.13. Proof of principle assay. Diploid cells formed by mating: MaV103(**III**):pEcTyrRS +
MaV203(**IV**): pADH1/GAL4_T44_R110 and plated on (A) SD-LEU-TRP and (B) SD-LEU-TRP-URA,
MaV103(**III**):pEcTyrRS + MaV203(**V**):pESC-LEU plated on (C) SD-LEU-TRP and (D) SD-LEU-TRP-URA.
The lack of colony formation on (B) indicates that EcTyrRS is not able to suppress the amber codons in
GAL4, in the absence of tRNA$_{CUA}^{Tyr}$.

Moreover, Figure 4.13 illustrates that both the mating between MAV103
(**III**) with MaV203 (**IV**) (Figure 4.13A) and between MAV103 (**III**) with
MaV203 (**V**) resulted in diploid cells (Figure 4.13C). Again both diploid
cells are not capable of growing on SD-LEU-TRP-URA (Figure 4.13B and
D). This means that also in this case, the amber codons are not
suppressed in the GAL4 gene. Only diploid cells that contain the
tRNA$_{CUA}^{Tyr}$ as well as the EcTyrRS gene are capable of suppressing the
amber codons present in GAL4. As a conclusion, it is clear that the
EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair successfully suppresses the amber codons
present in the GAL4 gene in diploid cells. When either EcTyrRS or
tRNA$_{CUA}^{Tyr}$ is expressed separately, no amber suppression takes place.
This proves that the diploid cells can be used together with vectors **I**

(pEcTyrRS/tRNA$_{CUA}^{Tyr}$) and **IV** (pADH1/GAL4_T44_R110) for the screening of libraries of mutant EcTyrRSs.

4.3 Discussion

In general terms, MaV203 cells have been widely used for the screening of mutant aminoacyl-tRNA synthetase (aaRS) libraries[2,3,6,20]. Most yeast promoters fused to URA3, confer a URA$^+$ phenotype when no specific transcription factor is present[9,10]. Moreover, basal levels of URA3 expression are sufficient to promote growth on a growth medium lacking uracil as well as sensitivity for 5-fluoorotic acid (5-FOA). The SPAL10 promoter was developed to actively repress URA3 gene expression[1]. It is assumed that the SPAL10::URA3 fusion resulted in an active repression of the URA3 gene and thus a stable URA3$^-$ phenotype[1-7]. However, the obtained results described above prove that this is not the case.

Our results clearly demonstrate that after a number of cell divisions spontaneous, recessive mutations occur in a single gene of a protein involved in the repression of the SPAL10 promoter, resulting in a URA3$^+$ phenotype. This mutation has a serious impact on the screening of mutant aaRS libraries, especially since in the negative screening step, all MaV203 cells with a URA$^+$ phenotype will be removed. As a result, also mutant aaRSs with an affinity for the non-natural amino acid will be removed if present in a MaV203 cell with a spontaneous derepression of the URA3 gene. Therefore the real library size will be smaller than theoretically calculated.

The SPAL10 promoter is derived from the SPO13 promoter, and just like the SPO13 promoter it is actively repressed by the Rpd3 Histone DeAcetylase Complex (Rpd3 HDAC)[21-25]. The obtained results confirm the hypothesis of Prof. Dr. Vidal and Prof. Dr. Boeke, who developed the MaV203 strain in 1996. In a personal communication, these authors also hypothesized that spontaneous mutations in a component of the Rpd3 HDAC is causing the spontaneous URA3 activation. When the Rpd3 HDAC fails to repress the URA3 gene, it will be expressed on a basal level, resulting in a URA$^+$ and consequent 5-FOA$^-$ phenotype.

When looking at the Rpd3 HDAC, a mutation in the Sin3 and Rpd3 encoding regions seems to be most plausible. However, Sin3 and Rpd3 can be excluded since the presence of loss-of-function mutations in either one of these non-essential genes results in severe mating

deficiency[26]. The results presented here show that the haploid cells containing the mutation are not mating deficient and are still capable of forming diploid cells.

In order to pinpoint the exact gene in which a mutation responsible for the derepression of the URA3 gene occurs, further investigation is needed. Especially because the exact mechanism of Rpd3 HDAC regulation is not known. Many proteins, involved in the Rpd3 HDAC are regulated by kinases and phosphatases. Therefore, not only a mutation in a protein involved in the Rpd3 HDAC might be the cause, also a mutation in the regulatory kinase/phosphatase mechanisms might result in derepression of the URA3 gene. However, research related to the exploration of the exact mechanism of the Rpd3 HDAC regulation is beyond the scope of this thesis.

Nevertheless, even without identifying the exact gene in which the mutations occur, unwanted URA3 activation can be avoided by the use of diploid cells.

4.4 Conclusions

In this chapter the cause of the spontaneous and unwanted URA3 expression in the MaV203 cells was elucidated. We were the first to notice this spontaneous URA3 derepression in MaV203 cells since it has never been reported in literature before. The experiments accomplished indicate that spontaneous, recessive mutations in a single gene are responsible for the unwanted activation of the URA3 gene. This unwanted URA3 gene expression can have a negative impact on techniques that require a strict URA$^-$ phenotype (e.g. the screening of a library of mutant EcTyrRSs, reverse yeast-two-hybrid). Therefore, the real library size will be smaller than theoretically calculated. The mutation most likely occurs in a protein involved in the Rpd3 Histone DeAcetylase Complex (Rpd3 HDAC) or in a kinase/phosphatase involved in the regulation of this complex.

The plausibility of these findings was further confirmed in a personal communication with Prof. Dr. Vidal and Prof. Dr. Boeke, who developed the MaV203 strain in 1996[1]. In order to pinpoint the exact gene in which a mutation responsible for the derepression of the URA3 gene occurs, further investigation is clearly needed, especially because the exact mechanism of Rpd3 HDAC regulation is not fully known. However,

research related to the exploration of the exact mechanism of the Rpd3 HDAC regulation is beyond the scope of this thesis.

Since the mutation is recessive, this unwanted URA3 activation can be avoided using diploid cells. Proof of principle experiments have been performed with the same conditions as the positive selection step of the double-sieve screening method, using the wildtype EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair. It was verified whether the EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair can suppress amber codons in GAL4 and thus activate the URA3 gene in diploid cells. It is clear that the EcTyrRS/ tRNA$_{CUA}^{Tyr}$ pair, expressed by a yeast episomal plasmid (Yep), successfully suppresses the amber codons present in the GAL4 gene in diploid cells. When either EcTyrRS or tRNA$_{CUA}^{Tyr}$ are absent, no amber suppression takes place. The results of this experiment demonstrate that the diploid cells can be used together with the plasmids pEcTyrRS/tRNA$_{CUA}^{Tyr}$ and pADH1/GAL4_T44_R110 (described in chapter 3) for the screening of libraries of mutant EcTyrRSs.

4.5 References

1. Vidal, M. *et al.,* Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions, *Proc. Natl. Acad. Sci. U.S.A.* **93,** 10315–20 (1996).

2. Chin, J. W. *et al.,* Progress toward an expanded eukaryotic Genetic Code, *Chem Biol.* **10,** 511–519 (2003).

3. Xie, J. *et al.,* Adding amino acids to the genetic repertoire, *Curr. Opin. Chem. Biol.* **9,** 548–54 (2005).

4. Deiters, A. *et al.,* Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*, *J. Am. Chem. Soc.* **125,** 11782–3 (2003).

5. Chin, J. W. *et al.,* An expanded eukaryotic genetic code, *Science* **301,** 964–7 (2003).

6. Cropp, T. A. *et al.,* Reprogramming the amino-acid substrate specificity of orthogonal aminoacyl-tRNA synthetases to expand the genetic code of eukaryotic cells, *Nat. Protoc.* 2590–600 (2007).

7. Hancock, S. M. *et al.,,* Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair, *J. Am. Chem. Soc.* **132,** 14819–24 (2010).

8. Vidal, M. *et al.,* Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system, *Proc. Natl. Acad. Sci. U.S.A.* **93,** 10321–10326 (1996).

9. Losson, R. *et al., In vivo* transcription of a eukaryotic regulatory gene, *EMBO Rep.* **2,** 2179–2184 (1983).

10. Aparicio, O. M. *et al.,* Overcoming telomeric silencing: a trans-activator competes to establish gene expression in a cell cycle-dependent way, *Genes Dev.* **8,** 1133–1146 (1994).

11. Wang, H. T. *et al.,* Developmental regulation of SPO13, a gene required for separation of homologous chromosomes at meiosis I, *Mol. Cell. Biol.* **7,** 1425–35 (1987).

12.   Luche, R. M. *et al.,* A cis-acting element present in multiple genes serves as a repressor protein binding site for the yeast CAR1 gene, *Mol. Cell. Biol.* **10,** 3884–95 (1990).

13.   Park, H. O. *et al.,* Transcriptional regulation of a yeast HSP70 gene by heat shock factor and an upstream repression site-binding factor, *Genes Dev.* **5,** 1299–1308 (1991).

14.   Bowdish, K. S. *et al.,*Bipartite structure of an early meiotic upstream activation sequence from *Saccharomyces cerevisiae*, *Mol. Cell. Biol.* **13,** 2172-2181 (1993).

15.   John, M. *et al.,*The INO1 promoter of *saccharomyces cerevisiae* includes upstream repressor sequence (URS1) common to a diverse set of yeast genes. **175,** 4235–4238 (1993).

16.   Vidal, M. *et al.,*Identification of essential nucleotides in an upstream repressing sequence of *Saccharomyces cerevisiae* by selection for increased expression of TRK2, *Proc. Natl. Acad. Sci. U. S. A.* **92,** 2370–4 (1995).

17.   Yeast Protocols Handbook, Clontech Laboratories, (2009).

18.   Young, T. S. *et al.,*Beyond the canonical 20 amino acids: expanding the genetic lexicon, *J. Biol. Chem.* **285,** 11039–44 (2010).

19.   Edwards, H. . *et al.,* A bacterial amber suppressor in *Saccharomyces cerevisiae* is selectively recognized by a bacterial aminoacyl-tRNA synthetase, *Mol. Cell. Biol.* **10,** 1633–41 (1990).

20.   Deiters, A. *et al.,* Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*, *J. Am. Chem. Soc.* **125,** Supporting Information. 1–8 (2003).

21.   Vidal, M. *et al.,* RPD1 (SIN3/UME4) is required for maximal activation and repression of diverse yeast genes. **11,** 6306–6316 (1991).

22.   Vidal, M. *et al.,* RPD3 encodes a second factor required to achieve maximum positive and negative transcriptional states in *saccharomyces cerevisiae*, **11,** 6317–6327 (1991).

23. Kadosh, D. *et al.,* Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters, *Cell* **89,** 365–71 (1997).

24. Washburn, B. K. *et al.,* Identification of the Sin3-binding site in Ume6 defines a two-step process for conversion of Ume6 from a transcriptional repressor to an activator in yeast, **21,** 2057–2069 (2001).

25. Grzenda, A. *et al.,* Sin3: master scaffold and transcriptional corepressor, *Biochim. Biophys. Acta* **1789,** 443–50 (2009).

26. Lenstra, T. L. *et al.,* The specificity and topology of chromatin interaction pathways in yeast, *Mol. Cell* **42,** 536–549 (2011).

Chapter 5

Construction of a library of EcTyrRS mutants

5.1 Introduction

In order to obtain a mutant *E. coli* tyrosyl-tRNA synthetase (EcTyrRS) with affinity for *p*-azidophenylalanine (*p*-azidoPhe) instead of tyrosine, a library needs to be constructed and screened for the appropriate mutant EcTyrRS[1–5]. In this chapter, the steps needed to construct a library of mutants EcTyrRSs will be described.

Expanding the eukaryotic genetic code with an *E. coli* aminoacyl tRNA synthetase (aaRS)/tRNA pair to incorporate non-natural amino acids is a very powerful tool, not only for manipulating protein functions in eukaryotic cells but also for the development of new diagnostics and therapeutics. In literature it can be found that mutant aaRS libraries have been constructed and screened against several non-natural amino acids[1,6,7–14]. This resulted in a wide variety of mutant aaRSs capable of incorporating a large range of non-natural amino acids[15]. In particular, Deiters *et al.*[5] described the construction and screening of a mutant *E. coli* tyrosyl-tRNA synthetase (EcTyrRS) library for the selective and efficient incorporation of *p*-azidoPhe and *p*-propargyloxyphenylalanine (*p*-propargyloxyPhe). Although mutant EcTyrRSs capable of incorporating modified amino acids have been reported in literature, it is still a benefit to possess a library since it can be screened against a wide variety of other non-natural amino acids.

The protein engineering studies on EcTyrRS that were performed in the past, and are described in literature, were based on the crystal structure of the TyrRS of *Bacillus stearothermophilus* and not on the structure of EcTyrRS. This was done since the crystal structure of EcTyrRS was not available at the time. However, more information is currently available about the 3D structure of EcTyrRS and differences between the two

synthetases are found which can have an impact on the recognition of tyrosine analogues[16]. The crystal structure of EcTyrRS provides improved information about the amino acid recognition in its binding pocket and better suggestions can be made for modifying it to recognize non-natural amino acids[16]. Figure 5.1 illustrates the protein residues proximal to L-tyrosine in the amino acid binding site of TyrRS of *E.coli* (A) and *B. staerothermophilus (B)*.
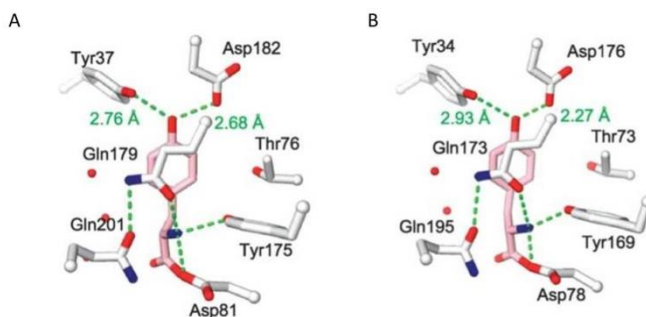
Figure 5.1. Protein residues proximal to the L-tyrosine[16]. (A) The amino acid binding site of the *E. coli* TyrRS·L-tyrosine complex and (B) *B. staerothermophilus* TyrRS·L-tyrosine complex.

The library screening performed by Deiters *et al*.[5] against both *p*-azidoPhe and *p*-propargyloxyPhe did not result in a unique mutant EcTyrRS for the azido-amino acid or a unique one for the alkyne-amino acid. Instead, around five mutants were obtained for both amino acids of which most of them were capable of incorporating *p*-azidoPhe as well as *p*-propargyloxyPhe, indicating that they are not highly specific. Therefore, by using the crystal structure of EcTyrRS instead of *Bacillus stearothermophilus* the amino acids Tyr37, Asp81, Tyr175, Gln179 and Asp182 in the binding pocket seem better candidates to be randomized, which could lead to a more specific library and thus more selective mutants. It was chosen to construct and screen an own library of mutant EcTyrRS, not only to obtain more selective EcTyrRS mutants, but also to apply it later on other aaRSs.

An advantage of having a library is that its use is not limited to one tyrosyl analogue but can be used for the screening against a wide variety of other analogues. An additional benefit of using the yeast *S. cerevisiae* for the screening of the mutant EcTyrRSs library is that the translational machinery of eukaryotes is very well conserved[17,18]. AaRS/tRNA pairs developed in yeast for the incorporation of non-natural amino acids can be used in higher eukaryotic organisms as well[19,20]. This

means that the tRNA aminoacyl synthetase/tRNA$_{CUA}$ pairs that are orthogonal in *S. cerevisiae*[21,22] are also orthogonal in higher eukaryotes such as mammalian cells[19,23].

The new mutant library is generated by randomly mutating key amino acids present in the binding pocket of the EcTyrRS[3,5]. Figure 5.2 shows the methodology used for the synthesis of site-specifically modified nanobodies with the library construction marked by the red frame.
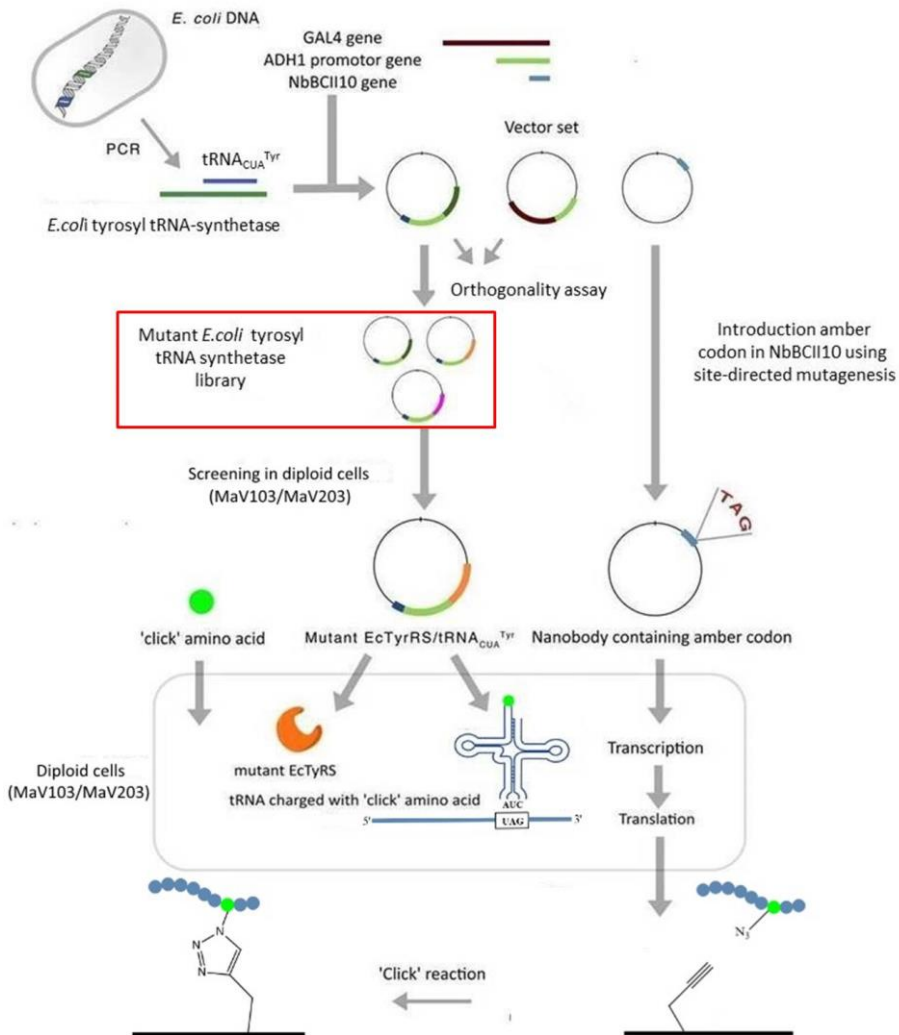
Figure 5.2. Methodology used for the synthesis of site-specifically modified nanobodies. The library construction is marked by the red frame.

5.2 Construction of a library of mutant EcTyrRSs

The 3D crystal structure of the EcTyrRS[16] is used to identify the amino acids that will be randomly mutated in the binding pocket of EcTyrRS. For the mutagenesis degenerate codon-containing primers are used (degenerate codon NNK, where N can be A, G, C, T and K can be G, T or MNN on the complementary DNA strand with M = C, A). The degenerate codon-containing primers are a mixture of primers including different combinations of nucleotides at the position of the degenerate codon.

The amino acids that were identified to be mutated were Tyr37, Asp81, Tyr175, Gln179 and Asp182. These amino acids were chosen based on the studies performed by Kobayashi *et al.*[16], as can be seen in Figure 5.1. The Quickchange Multi Site-Directed Mutagenesis Kit (Stratagene, Agilent Technologies, Brussels, Belgium) could not be used for generating these random mutations. This kit is only compatible with plasmids up to 8 kbp and since the plasmid required for this experiment is around 8.6 kbp, the performance of the kit would be on its operational limit. Besides, the Quickchange Multi Site-Directed Mutagenesis Kit only allows the introduction of a degenerate codon at one site, making it not suitable for the library construction.

Many reports in literature describe the use and screening of a library of mutant aminoacyl-tRNA synthetases (aaRS)[1,3,4,20], but very few reports can be found on the construction of such a library. Cropp *et al.*[3] describe the overall method and protocol for the construction of a library of mutant EcTyrRSs using Enzymatic Inverse PCR (EIPCR). With EIPCR one round of PCR based mutagenesis is needed per amino acid to be mutated. This means that five rounds of PCR based mutagenesis are needed in order to construct the mutant library, making it a very time-consuming procedure. However, the protocol in this article describes the random mutation of only three and not of five amino acids. Therefore, many practical problems concerning the library construction remain unsolved.

A potentially faster alternative could be found in the recently developed GeneArt® Site-Directed Mutagenesis PLUS Kit (Life Technologies). According to the manufacturer this kit allows the random mutation of all five amino acids in one round of mutagenesis.

5.2.1 Mutagenesis using the GeneArt® Site-Directed Mutagenesis PLUS Kit

According to Life Technologies, the new GeneArt® Site-Directed Mutagenesis PLUS Kit, optimized in combination with the Accuprime™ *Pfx* DNA polymerase, provides a simple and highly efficient tool for *in vitro* multi site-directed mutagenesis in DNA plasmids of up to 14 kbp from any *E. coli* plasmid. Although the protocol states that only three degenerate nucleotides at three different sites can be introduced, the R&D department of Life Technologies assured that it can be used to introduce five degenerate codons as well. This should be possible since the three residues Tyr175, Gln179 and Asp182 are located in close proximity of each other allowing their mutation with a single primer. As this is a newly developed mutagenesis kit, no information can be found about it in literature yet. To the best of our knowledge, this is the first time that this kit is used for the construction of a mutant aaRS library.

For the mutagenesis three sets of complementary, degenerate primers were used. These primers were designed using the web-based GeneArt® Primer and Construct Design tool. Figure 5.3 gives an overview of the mutagenesis procedure. The complete process of mutagenesis consists of: 1) a methylation and mutagenesis reaction, 2) a recombination reaction and, 3) a transformation in One Shot® MAX Efficiency® DH5α™-T1R competent *E. coli* cells (Appendix page 185). These experimental steps start from pEcTyrRS/tRNA$_{CUA}^{Tyr}$ as a template and are described below.

Step 1: Methylation and mutagenesis reaction. In the first step the template DNA is methylated and amplified in a mutagenesic (PCR) reaction. The template DNA is added to a master mix (see Table A13 in appendix) and divided equally into three tubes, to which 1 µl of the mutagenic primer mix (10 µM) is added. The master mix does not only contain the components to perform the mutagenesis reaction, but also DNA methylase and S-adenosine methionine (SAM) for the methylation reaction. SAM is the methyl donor for the DNA methylase in order to methylate cytosine in the template DNA. This methylated template DNA will be digested by the *McrBC* endonuclease present in the One Shot® MAX Efficiency® DH5α™-T1R competent *E. coli* cells (abbreviated as DH5α cells) (step 3).

Step 1:
Methylation &
Mutagenesis

Methylation

Mutagenesis

Step2:
In vitro
recombination reaction

Step 3:
Transformation
in DH5α cells
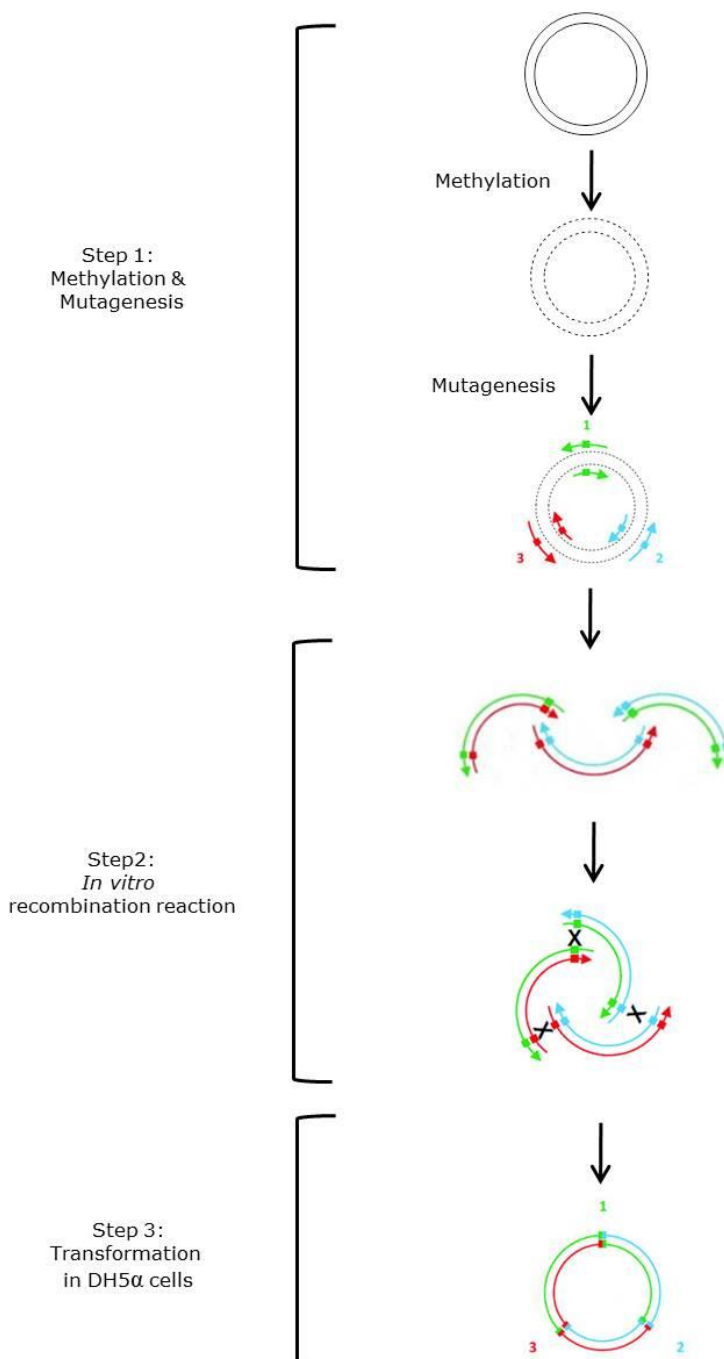
Figure 5.3. Overview of the mutagenesis procedure using the GeneArt® Site-Directed Mutagenesis PLUS Kit. The complete process of mutagenesis consists of: 1) a methylation and mutagenesis reaction, 2) a recombination reaction and, 3) a transformation in One Shot® MAX Efficiency® DH5α™-T1R competent *E. coli* cells.

The mutagenesic primers are listed in Table 5.1 and the composition of the primer mixtures is listed in Table 5.2. Each primer mix contains a forward and a reverse primer that corresponds to a different mutation site, in order to prevent the primers from annealing to each other. After adding the primer mix to the master mix, the methylation and mutagenesis reaction are performed according to the parameters presented in Table A14 of the appendix. After the PCR reaction, each tube (each with a different primer mix) contains only a fragment of the original template DNA. These fragments are assembled by means of a recombination reaction (step 2).

Table 5.1. List of mutagenesic primers with the degenerate codon depicted in bold in the sequence.

Target amino acid	Primer	Primer sequence
Tyr37	Forward (F1)	GGCCCGATCGCGCTC**NNK**TGCGGCTTCGATCCT
	Reverse (R1)	AGGATCGAAGCCGCA**MNN**GAGCGCGATCGGGCC
Asp81	Forward (F2)	ACGGGTCTGATTGGC**NNK**CCGAGCTTCAAAGCT
	Reverse (R2)	AGCTTTGAAGCTCGG**MNN**GCCAATCAGACCCGT
Tyr175, Gln179, Asp182	Forward (F3)	TTTCGTTCACTGAGTTTTCC**NNK**AACCTGTTG**NNK**GGTTAT**NNK**TTCGCCTGTCTGAACAAACA
	Reverse (R3)	TGTTTGTTCAGACAGGCGAA**MNN**ATAACC**MNN**CAACAGGTT**MNN**GGAAAACTCAGTGAACGAAA

Table 5.2. Composition of the primer mixtures used for the mutagenesis reaction in the binding pocket of EcTyrRS.

Primer Mix 1	R1 + F3
Primer Mix 2	R2 + F1
Primer Mix 3	R3 + F2

Step 2: Recombination reaction. The recombination reaction is performed to assemble the PCR products acquired in Step 1. For the recombination reaction the mixture in Table A15 of the appendix was made. After mixing well and incubating at room temperature for 15 minutes, the reaction was stopped by the inhibition of the DNA methylase by adding EDTA (0.5 M) and placing the mixture on ice.

Step 3: Transforming the mutated DNA in the DH5α cells. Step 3 was initiated immediately after step 2. This is the DNA transformation using the DH5α cells. After finishing the recombination reaction, the DNA mixture contains mutated DNA, besides non-mutated (methylated) template DNA. Therefore, the DH5α cells are used to eliminate the unwanted, template DNA. DNA methylase was added in step 1 during the PCR amplification, resulting in the methylation of cytosine of the template DNA. The DH5α cells are *McrBC*[+], meaning that they express the *McrBC* endonuclease which digests the methylated template DNA. The linear unmethylated, mutated DNA is circularized by the DH5α cells. The transformation is done according to the manufacturer's protocol.

5.2.1.1 Results and discussion of the mutagenesis via the GeneArt® Site-Directed Mutagenesis PLUS Kit

In the protocol of the GeneArt® Site-Directed Mutagenesis PLUS Kit it is noted that the optimal amount of template DNA is between 20 and 25 ng. This is because higher concentrations can reduce the efficiency of the methylation reaction and mutagenesis. Despite the fact that the manufacturer proposes a DNA range of 20 ng/µl – 40 ng/µl for the template DNA, a tenfold excess was also tested to verify if this was the ideal range. Therefore, three different concentrations of template DNA were used: 20 ng/µl, 40 ng/µl and an excess of 375 ng/µl.

After transforming the DH5α cells, tenfold dilutions were made before plating the cells on Luria-Bertani /ampicillin (LB[amp]) plates. Samples of

100 µl cell suspension were taken from all the dilutions and subsequently plated, except for the undiluted cell suspension. For the undiluted cell suspension 30 µl, 50 µl, 60 µl and 100 µl were plated independently.

It was observed that only the undiluted DH5α cells, transformed with the mutant DNA resulting from 375 ng/µl template DNA, were able to form a small amount of colonies as can be seen on Figure 5.4. No other cells were able to form colonies. Since only colonies are formed when 375 ng/µl of template DNA was used, the question arose whether this was the result of a successful mutagenesis or if it was caused by an inefficient methylation of such a large quantity of template DNA. If not all template DNA is methylated, the unmethylated template DNA will not be degraded by the DH5α cells after transformation, allowing the cells to grow on LB^{amp}. To be sure that all template DNA is methylated, a methylation control was simultaneously performed with the mutagenesis reaction, using the same excessive amount of template DNA (375 ng/µl).
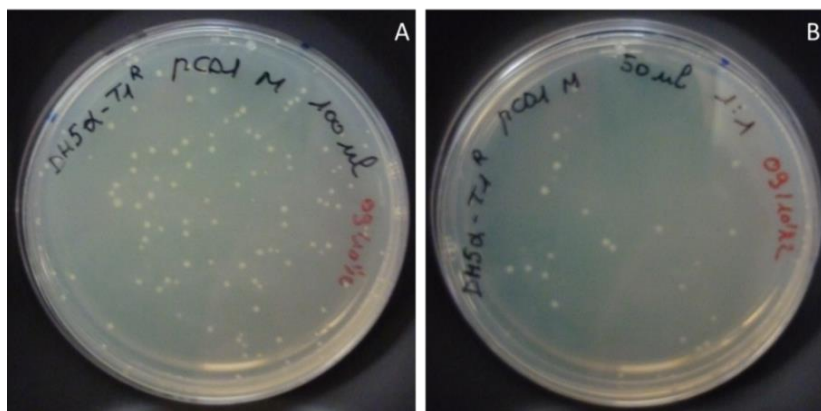
Figure 5.4. Colony formation on LB^{amp} plates after transforming the DH5α cells with the mutant DNA resulting from 375 ng/µl template DNA. (A) 100 µl of undiluted cell suspension and, (B) 50 µl of undiluted cell suspension.

Methylation efficiency control. To verify the methylation efficiency, an additional control sample was prepared. This sample contained all the previously mentioned PCR components except S-adenosyl methionine (SAM). The template DNA without SAM will not be methylated and hence not be degraded in the DH5α cells.

By comparing the amount of colonies formed after transformation of DNA with and without methylation, it was possible to determine the

methylation efficiency. Approximately 2 minutes before the end of the incubation period at 37 °C, the PCR program is paused and 2 μl of the two PCR samples (one with SAM and one without SAM) was transferred to sterile tubes containing 8 μl of 20 mM EDTA. Afterwards the samples were transformed in the DH5α cells according to the manufacturer's protocol and plated on LB^{amp} plates (Figure 5.5).
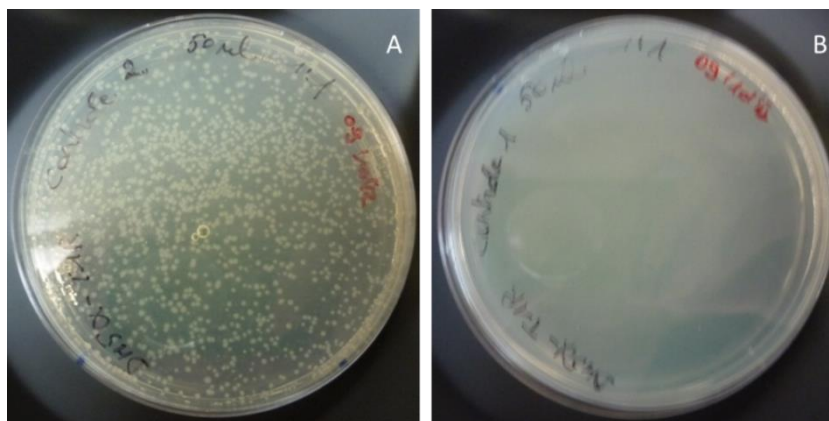
Figure 5.5. Methylation efficiency control. DH5α cells plated on LB^{amp} after transformation with the PCR samples, incubated in (A) the absence of SAM and, (B) the presence of SAM. It can be observed that if SAM is present, the digestion of the template DNA is 100% efficient: All template DNA was successfully digested, resulting in the absence of colony formation.

The methylation efficiency is determined by the following equation, given in the manufacturer's protocol:

$$\% \ methylation = \frac{\# \ colonies \ (without \ SAM) - \# \ colonies \ (with \ SAM)}{\# \ colonies \ (without \ SAM)} \ x \ 100$$

Figure 5.5 shows that no colonies are present for the cells transformed with the PCR/EDTA mix containing SAM. However, the cells transformed with the PCR/EDTA mix without SAM do form colonies. This means that the methylation reaction was 100% successful. Even with a large excess of template DNA (375 ng/μl), the DNA methylase was able to methylate all DNA. This also means that the colonies formed after transforming the DH5α cells with the amplified DNA resulting from the 375 ng/μl template DNA, are not the result of non-digested template DNA but must originate from amplified mutant DNA.

Verification of the transformation efficiency. Transformation competence of the DH5α cells was verified by transforming the cells with the supercoiled pUC19 plasmid included in the kit. The transformation control was performed according to the manufacturer's protocol. After transformation and plating the cells on LB^amp agar plates, the transformation efficiency could be calculated using the following equation, given in the manufacturer's protocol:

$$\frac{\# \text{ of colonies}}{50 \text{ pg plasmid DNA}} \, x \, \frac{300 \text{ μl}(total\ transformation\ volume)}{30 \text{ μl } (plated)} \, x \, 100(dilution\ factor)$$

$$= \frac{\# \text{ of colonies } (taking\ the\ dilution\ into\ account\ )}{pg\ plasmid\ DNA}$$

$$= \frac{\# \text{ of colonies } (taking\ the\ dilution\ into\ account) \, x \, 10^6}{\text{μg } plasmid\ DNA}$$

Eleven colonies were visible on the LB^amp plates after transformation. Introducing this in the above equation results in a transformation efficiency of $2.2*10^8$ colonies per μg plasmid DNA.
This result was slightly lower than expected since, according to the manufacturer, these cells should have a transformation efficiency of at least 10^9 colonies per μg plasmid DNA. However, this cannot account for the fact that no or very few colonies are formed after the mutagenesis reactions.

Mutagenesis Efficiency Control. The mutagenesis efficiency was verified using the pMSDM-White vector included in the kit. This vector is a 5 kbp pUC19-based plasmid containing the spectinomycin-resistance gene fused to the first 1815 bp of *E. coli*'s *lacZ* gene. The gene fusion contains three point mutations that inactivate the gene's ability to perform α-complementation when introduced into DH5α cells. The three point mutations are:

   i.   A replacement of cytosine at location 256 by thymine, generating a premature stop codon,

ii.   A deletion at position 480, generating a frameshift,

iii.   A guanine insertion at location 1062, also generating a frameshift.

The pMSDM-White vector produces white colonies on LB plates containing 30-100 µg/ml X-gal. Three control primer mixtures are included in the kit, which are designed to revert the mutations present in the pMSDM-White vector back to the wildtype sequence. Only when all three mutations are reverted to the wildtype sequence, blue colonies will grow on the X-gal containing LB agar plates. By comparing the amount of blue colonies to the amount of white colonies, the mutagenesis efficiency can be derived. The mutagenesis control was performed according to the manufacturer's protocol.

After performing the recombination reaction and transformation into DH5α cells according to the manufacturer's protocol, the cells were plated on LB plates containing 100 µg/ml X-gal (Figure 5.6). It was observed that ten colonies were blue, indicating that the mutagenesis reaction successfully reverted the point mutations back to the wild-type sequence in these cells. Only two white colonies were present on the plate, meaning that the mutagenesis efficiency is 83%, which is the expected efficiency for this kit.
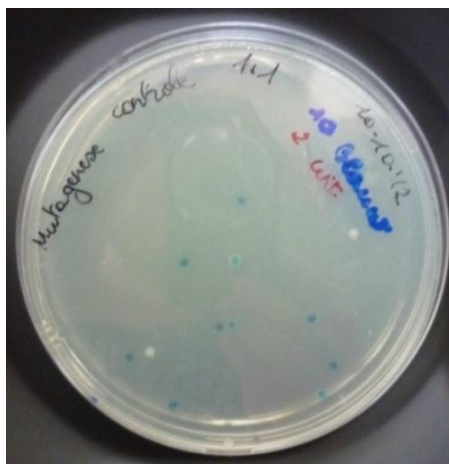
Figure 5.6. Mutagenesis efficiency control. DH5α cells plated on LB plates containing 100 µg/ml X-gal. Ten out of twelve colonies turned blue, indicating a transformation efficiency of 83%.

Analysis of the PCR products obtained after mutagenesis. The mutagenesis reactions were unsuccessful despite the fact that the methylation reaction, mutagenesis control reaction and the transformation work properly. Therefore, it was opted to analyze the PCR products obtained after the methylation and amplification. According to the manufacturer's protocol, faint or even multiple bands can still result in high mutagenesis efficiencies. The expected lengths of the PCR products were: 8254 bp for primer mix 1 (primers R1 + F3); 165 bp for primer mix 2 (primers R2 + F1) and 341 bp for primer mix 3 (primers R3 + F2).

On Figure 5.7 it can be seen that when 20 ng/µl (A) or 40 ng/µl (B) template DNA is used, bands appear at the correct position for the mutagenesis products from primer mix 2 (165 bp) and primer mix 3 (341 bp) (lanes 2 and 3, respectively). For primer mix 1 no band appears at the correct position (not shown on gel). However, when 375 ng/µl template DNA is used (C), no bands were visible at the correct position for the mutagenesis products, neither for primer mix 1, 2 or 3. No PCR product of the correct length was obtained using primer mix 1 (primers R1 + F3), for all the different concentrations of template DNA. Therefore, it was hypothesized that either primer R1 or primer F3 failed to hybridize correctly to the template DNA.
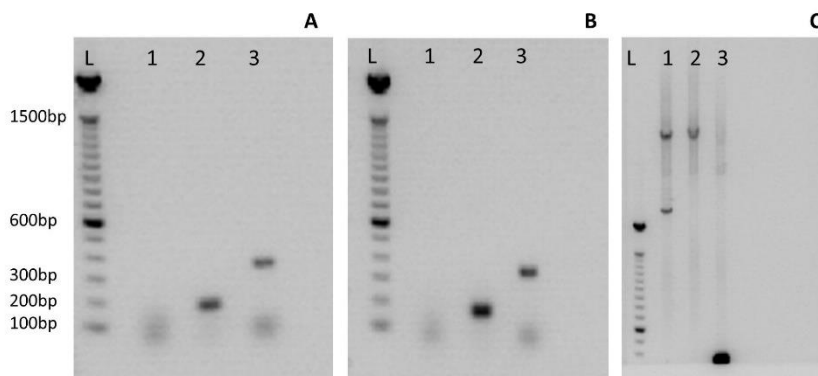
Figure 5.7. Analysis of the PCR products after the methylation and mutagenesis reaction (0.8% agarose) using (A) 20 ng/µl, (B) 40 ng/µl and, (C) 375 ng/µl template DNA. It can be seen that when 20 ng/µl (A) or 40 ng/µl (B) template DNA is used, bands appear at the correct position for the mutagenesis products from primer mix 2 (165 bp) and 3 (341 bp) (lanes 2 and 3, respectively). However when 375 ng/µl template DNA is used, no bands appear at the correct position. In all cases no band appears at the correct position using primer mix 1 (lane 1). L = 100 bp ladder (Invitrogen).

Control of mutagenesis primers. To evaluate the mutagenesis primers, a PCR reaction was performed combining the mutagenesis primers with the forward and reverse primers used to amplify the EcTyrRS gene of *E. coli* BL21 cells (Table A.1 of the appendix). If the mutagenesis primers anneal properly to the template DNA, it will result in the amplification of fragments of the EcTyrRS gene. The primer combinations, as well as the expected length of the EcTyrRS fragments are listed in Table 5.3. The composition of the PCR mix and the parameters used for the PCR reaction are given in Table A2 and A16 of the appendix.

The PCR products were analyzed on a 1.5% agarose gel (Figure 5.8), which clearly shows that all the PCR amplifications were successful. In all cases bands are visible at the correct position, indicating that the mutagenesis primers are hybridizing at the correct position of the template DNA and so cannot be responsible for the lack of colony formation after the mutagenesis reaction.

Table 5.3. List of the primer combinations used to validate the functionality of the mutagenesis primers, together with the expected lengths of the PCR products.

	Primers	length PCR fragment
1	EcTyrRS_forw + R1	141 bp
2	EcTyrRS_forw + R2	273 bp
3	EcTyrRS_forw + R3	581 bp
4	EcTyrRS_rev + F1	1188 bp
5	EcTyrRS_rev + F2	1056 bp
6	EcTyrRS_rev + F3	779 bp
7	EcTyrRS_for+ EcTyrRS_rev	1275 bp
8	negative control 1	/
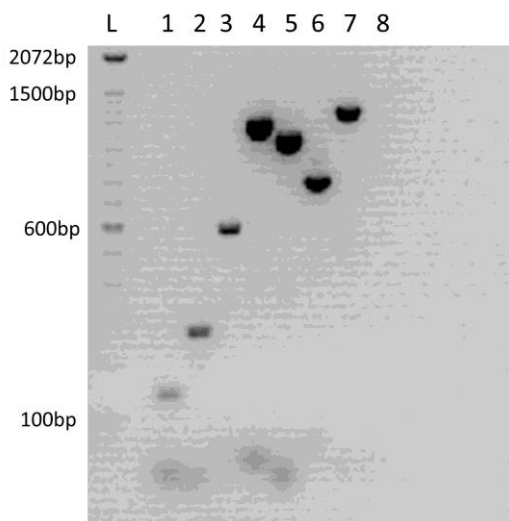
Figure 5.8. Analysis of the PCR performed to validate the mutagenesis primers (1.5% agarose). The lane numbers match with the numbers in Table 5.3 in which also the expected lengths of the PCR products are given. In all cases the PCR products have the correct length, indicating that the mutagenesis primers hybridize correctly to the template DNA.

Introducing the degenerate codons in two rounds. It can be concluded thus far that the boundaries of the mutagenesis kit were being pushed and therefore it was attempted to introduce the degenerate mutations in two successive rounds of mutagenesis. In the first round the codons of Tyr37 and Asp81 are randomized. Two primer mixtures (shown in Table 5.4) were used for the amplification reaction. The sequences of the primers are listed in Table 5.1.

Table 5.4. Composition of the primer mixtures used for the mutagenesis reaction of Tyr37 and Asp81 in the binding pocket of EcTyrRS.

Primer Mix 1	F1+ R2
Primer Mix 2	F2+ R1

For this experiment, 20 ng/μl of template DNA was used and the methylation and amplification reaction were performed according to the manufacturer's protocol. PCR analysis of the amplification products (Figure 5.9) shows a band at the correct position (165 bp) for primer mixture 1 (lane 1), but no band was visible at the position of 8509 bp for primer mixture 2 (Figure not shown).
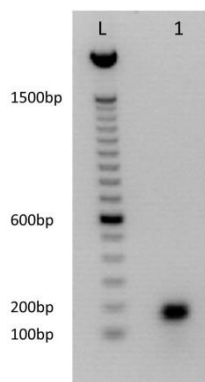
Figure 5.9. Analysis of the PCR products of the mutagenesis of Tyr37 and Asp81 (1.5% agarose). L: 100 bp ladder (Invitrogen), lane 1: PCR product using primer mix 1. In lane 1, a band is visible at he correct position of 165 bp, indicating that the PCR reaction was successful.

After the recombination reaction and transformation into DH5α cells, no colonies were visible on LB agar plates containing 100 µg/ml ampicillin. Due to the fact that the kit failed in introducing a random mutation in two codons, and the mutagenesis primers have been shown to hybridize properly, it was concluded that the protocol or the kit of the manufacturer is not working properly.

Adapted mutagenesis protocol from Life Technologies. Despite the fact that the mutagenesis primers and the methylation reaction work properly and the transformation efficiency was good, the mutagenesis reaction was not successful. Originally it was decided to use this GeneArt mutagenesis kit, since it claimed to be able to introduce all the mutations in a single mutagenesis reaction. However, the results obtained in this work show that this kit is not able to introduce the mutations, not even in two rounds. For that reason, this issue was addressed to the manufacturer (Life Technologies). It was also pointed out by us that some errors were present in their protocol. Because of this, the manufacturer provided an adapted protocol, optimized for the introduction of five random mutations. According to their recommendation, it was necessary to change from the GeneArt® Site-Directed Mutagenesis PLUS Kit to the GeneArt® Site-Directed Mutagenesis Kit, combined with the adapted protocol that was provided. The manufacturer's R&D team stated that the work method that was performed is correct but that the library construction was pushing the limits of the PLUS kit.

Apparently, the all-in-one formula of the GeneArt® 2x Enzyme mix provided in the PLUS kit, results in a decrease of mutagenesis efficiency

compared to the enzyme present in the older GeneArt® Site-Directed Mutagenesis Kit. The GeneArt® Site-Directed Mutagenesis Kit has some minor differences in its protocol compared to the PLUS kit. The methylation and amplification reaction is identical and therefore the methylation and mutagenesis reaction was performed as described in step 1, section 5.2.1. However, the recombination reaction is different since the two mutagenesis kits contain a different enzyme mix for this reaction. For the recombination reaction using the adapted protocol, the mixture shown in Table A17 of the appendix was prepared. It was also advised to use the complete recombination reaction mixture to transform the DH5α cells.

However, also using the adapted protocol in combination with the GeneArt® Site-Directed Mutagenesis Kit did not lead to any colony formation after transforming the DH5α cells. This indicates that the GeneArt® 2x Enzyme mix was not causing the problems that were faced during the library construction. Clearly, our findings indicate that the GeneArt® Site-Directed Mutagenesis (PLUS) Kit is not capable of introducing degenerate codons on multiple locations in a gene and is not fit to be used for library construction. Therefore other methods to construct a library of mutant EcTyrRSs were investigated.

5.2.2 Enzymatic Inverse PCR (EIPCR)

An efficient way to produce libraries of mutant proteins is the so-called Enzymatic Inverse PCR[25]. In the past, this method has already been successfully used to produce libraries of mutant aaRS[3]. In this method the EcTyrRS gene is split into two fragments, EcTyrRS_Nterm and EcTyrRS_Cterm. Each fragment is cloned into separate small vectors (e.g. pUC19) and mutated by parallel rounds of Enzymatic Inverse PCR. Splitting up the EcTyrRS gene to create two sub-libraries results in a lower complexity and makes it easier to construct them using PCR-based mutagenesis.

Vector construction needed for EIPCR. To create the two sublibraries, the EcTyrRS gene needs to be split into two fragments that have to be ligated into separate small vectors. The N-terminal fragment of EcTyrRS (EcTyrRS_Nterm) is ligated in the pUC19 vector between the restriction sites EcoRI and PstI. The C-terminal fragment of EcTyrRS (EcTyrRS_Cterm) is ligated in the pUC19 vector between the restriction sites PstI and NotI. In order to accomplish this a PstI site is introduced in the EcTyrRS gene and a NotI site is inserted in the pUC19 vector.

Site-Directed Mutagenesis for the introduction of the PstI restriction site in EcTyrRS. A site-directed mutagenesis was performed using the QuickChange Lightning Site-Directed Mutagenesis kit (Stratagene, Agilent Technologies) according to the manufacturer's protocol. The vector pEcTyrRS/tRNA$_{CUA}^{Tyr}$, containing the EcTyrRS gene, was used as a template. The forward and reverse primers are displayed in Table A18 of the appendix. For the introduction of the PstI restriction site, adenine at position 585 was mutated into guanine (a585g). This is a silent mutation, meaning that after the a585g mutation, the codon will still encode the same amino acid and therefore the amino acid sequence of EcTyrRS is not affected by the mutagenesis. After the site-directed mutagenesis, restriction analysis and sequencing (S12 of the appendix) confirmed that the adenine at location 585 had been successfully changed to guanine, resulting in a PstI restriction site in the middle of the EcTyrRS gene. This results in the vector pEcTyrRS/tRNA$_{CUA}^{Tyr}$_PstI.

Introduction of a NotI restriction site in the pUC19 vector.
Figure 5.10 gives an overview of the procedure to introduce a NotI restriction site in the pUC19 vector. First the pUC19 vector was cut using PstI and HindIII restriction enzymes, resulting in a linear vector with a PstI overhang and a HindIII overhang (Figure 5.10A). Two oligonucleotides (IDT, Leuven, Belgium) (Figure 5.10B) were designed to be partially each other's reverse complement.
Hybridization results in a double stranded oligonucleotide with the NotI restriction site in the center (marked in green) and at the ends overhangs (marked in red) that are complementary to these resulting from cutting a vector with PstI and HindIII (Figure 5.10C). The double stranded oligonucleotide is ligated between the PstI and HindIII overhangs and in this way a NotI restriction site is introduced between the PstI and HindIII restriction sites of the pUC19 vector (Figure 5.10D). The ligation was performed using the T4 ligase (Thermoscientific) according to the manufacturer's protocol. The correct insertion of the NotI restriction site was confirmed by restriction analysis.

Oligo 1: GATTCATGCGGCCGCA

Oligo 2: AGCTTGCGGCCGCATGAATCTGCA

B

A

C

5'CTGCA   AGCTT 3'
3'G   A 5'

5' CTGCA   GATTCATGCGGCCGCA   AGCTT 3'
3' G   ACGTCTAAGTACGCCGGCGTTCGA   A 5'

PstI   HindIII

D

5' CTGCAGATTCATGCGGCCGCAAGCTT 3'
3' GACGTCTAAGTACGCCGGCGTTCGAA 5'
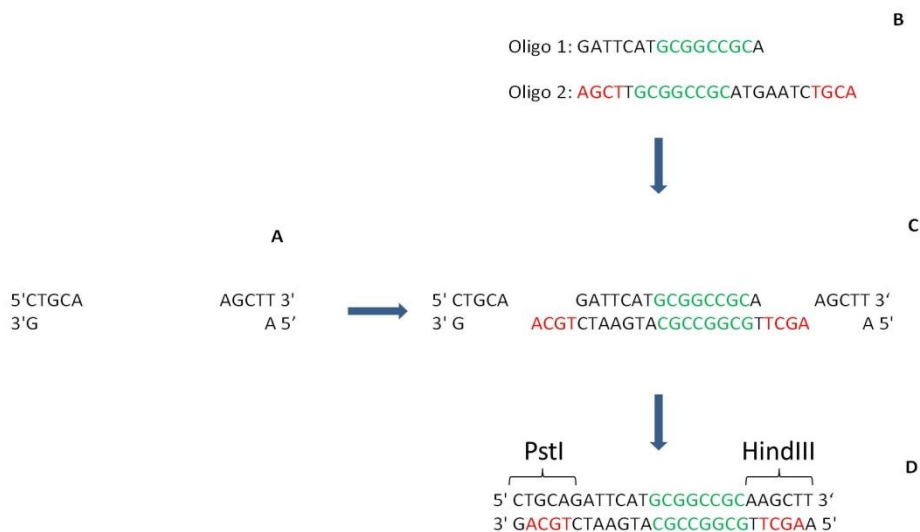
Figure 5.10. Introduction of a NotI restriction site between the PstI and the HindIII restriction sites. The overhangs present in the double stranded oligonucleotides to bind with the PstI and HindIII overhangs are marked in red. The NotI restriction site is marked in green.

Ligation of EcTyrRS-Nterm and EcTyrRS-Cterm in pUC19.

EcTyrRS-Nterm was excised from the pEcTyrRS/tRNA$_{CUA}$_PstI vector by means of the restriction enzymes EcoRI and PstI, resulting in a fragment of 585 bp (EcTyrRS-Nterm). Restriction digestion of pUC19 with these restriction enzymes generates two fragments: i) a fragment of 2647 bp (the open vector) and ii) a fragment of approximately 39 bp (not visible on the agarose gel). The EcTyrRS-Cterm was obtained by cutting the pEcTyrRS/tRNA$_{CUA}$_PstI vector with the restriction enzymes PstI and NotI, resulting in a fragment of 690 bp (EcTyrRS-Cterm). The pUC19 vector was cut with the same enzymes, resulting in a fragment of about 2675 bp (open vector) and a fragment of 10 bp (not visible on the agarose gel). After restriction digestion, the reaction mixtures were analyzed using agarose gel electrophoresis.
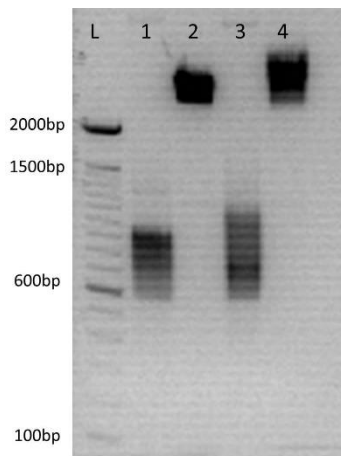
Figure 5.11. Restriction digestion of pEcTyrRS/tRNA$_{CUA}^{Tyr}$_PstI and pUC19 with respectively EcoRI + PstI (lane 2 & 3), and PstI + NotI (lane 4 & 5) (1% agarose). In all cases the bands are at the correct position. The wide bands in lanes 1 & 3 are the result of loading a high concentration of sample. L: 100bp ladder (Invitrogen).

Bands appeared at the expected position on the agarose gel (Figure 5.11), indicating that the restriction digestion was successful. EcTyrRS_Nterm, EcTyrRS_Cterm and the open pUC19 vector were excised from the agarose gel. After purification of the DNA, the EcTyrRS_Nterm and EcTyrRS_Cterm were cloned into the pUC19 vector and subsequently transformed in competent Top10 F' *E. coli* cells. This results in the pUC19-EcTyrRS_Nterm and pUC19-EcTyrRS_Cterm vectors, respectively. Plasmid DNA from five colonies of each was isolated and verified with restriction analysis.

5.3 Conclusions

As a conclusion it can be stated that constructing a library of EcTyrRS mutants involves a number of challenging steps. In this case five randomized NNK codons are needed to be introduced in the active site of EcTyrRS, with N = A/C/G/T and K = G/T. This means that NNK represents 32 different codons (4 x 4 x 2 = 32). When such a randomized codon is introduced at five locations in the binding pocket of EcTyrRS, it will result in a theoretical diversity of 32^5 (meaning 32^5 different mutants). In this chapter it has been shown that commercial kits such as the Quickchange Multi Site-Directed Mutagenesis kit (Agilent) or the GeneArt® Site-Directed Mutagenesis (PLUS) Kit in our hands were not successful in generating mutant libraries. A potentially successful way of generating such a library of mutant EcTyrRSs is by

means of EIPCR. This method is well documented in literature and has been used for the construction of mutant libraries before[3]. However, EIPCR requires a round of mutagenesis for every amino acid that needs to be mutated, making it a very time-consuming method. Preliminary work has already been done to construct the library of mutant EcTyrRS by means of EIPCR, *i.e.* the vectors have been constructed but more investigation is needed.

Another option is to make use of "custom library construction services" with a cost of about €13500. In this case the library is constructed and delivered as a glycerol stock in *E. coli*. This means that the library still needs to be extracted from the *E. coli* cells and transformed in the MaV203 yeast strain for screening.

It is possible to calculate the amount of yeast clones needed for a library of 32^5 different mutants to be complete with 99% confidence. The following equation given by Cropp *et al.*[3], with N being the number of clones needed and size being the theoretical diversity (32^5), suggests that a library of this size requires $1.55*10^8$ clones.

$$\frac{(ln(1 - confidence))}{\ln(1 - \frac{1}{size})} = N$$

$$\frac{(\ln(1 - 0.99))}{\ln\left(1 - \frac{1}{32^5}\right)} = 1.55*10^8$$

Knowing that a large-scale transformation of a good transforming yeast strain will generate only up to 10^6 transformants, this means that about 150 large-scale transformations need to be performed in order to transform the complete library into the correct yeast strain. The construction and the transformation of a mutant library in yeast is time-consuming and there is no guarantee of success. Therefore, the use of a mutant EcTyrRS, of which the affinity for *p*-azidoPhe has been described in literature[5], was evaluated for the site-specific modification of NbBCII10 (described in chapter 6).

5.4 References

1. Chin, J. W. *et al.,* Progress toward an expanded eukaryotic genetic code, *Chem Biol.* **10,** 511–519 (2003).

2. Chin, J. W. *et al.* An expanded eukaryotic genetic code. *Science* **301,** 964–7 (2003).

3. Cropp, T. A. *et al.*, Reprogramming the amino-acid substrate specificity of orthogonal aminoacyl-tRNA synthetases to expand the genetic code of eukaryotic cells, *Nat. Protoc.* 2590–600 (2007).

4. Deiters, A. *et al.,* Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*, *J. Am. Chem. Soc.* **125,** Supporting Information. 1–8 (2003).

5. Deiters, A. *et al.,* Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*, *J. Am. Chem. Soc.* **125,** 11782–3 (2003).

6. Xie, J. *et al.*, Adding amino acids to the genetic repertoire, *Curr. Opin. Chem. Biol.* **9,** 548–54 (2005).

7. Kowal, A K. *et al.*, Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria, *Proc. Natl. Acad. Sci. U.S.A.* **98,** 2268–2273 (2001).

8. Wang, L. *et al.*, Addition of the keto functional group to the genetic code of *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S.A.* **100,** 56–61 (2003).

9. Chin, J. W. *et al.,* Addition of *p*-azido-L-phenylalanine to the genetic code of *Escherichia coli*, *J. Am. Chem. Soc.* **124,** 9026–7 (2002).

10. Zhang, Z. *et al.,* A new strategy for the site-specific modification of proteins in vivo, *Biochemistry* **42,** 6735–46 (2003).

11. Deiters, A. *et al.*, Site-specific PEGylation of proteins containing unnatural amino acids, *Bioorganic Med. Chem. Lett.* **14,** 5743–5745 (2004).

12. Wang, J. *et al.*, A genetically encoded fluorescent amino acid, *PNAS*. **103,** 8738–8739 (2006).

13. Kim, C. H. *et al.*, Protein conjugation with genetically encoded unnatural amino acids, *Curr. Opin. Chem. Biol.* **17,** 412–9 (2013).

14. Xie, J. *et al.*, A chemical toolkit for proteins-an expanded genetic

code, *Nat. Rev. Mol. Cell Biol.* **7,** 775–82 (2006).

15. Liu, C. C. *et al.*, Adding new chemistries to the genetic code, *Annu. Rev. Biochem.* **79,** 413–44 (2010).

16. Kobayashi, T. *et al.,* Structural snapshots of the KMSKS loop rearrangement for amino acid activation by bacterial tyrosyl-tRNA synthetase, *J. Mol. Biol.* **346,** 105–17 (2005).

17. Kwok, Y. *et al.*, Evolutionary relationship between *Halobacterium cutirubrum* and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes, *Can. J. Biochem.* **58,** 213–218 (1980).

18. Bhattacharya, A. *et al.*, Nonsense-mediated mRNA decay in *Saccharomyces cerevisiae*, **274,** 15–25 (2001).

19. Sakamoto, K. *et al.,* Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells, *Nucleic Acids Res.* **30,** 4692–9 (2002).

20. Köhrer, C. *et al.*, Import of amber and ochre suppressor tRNAs into mammalian cells: a general approach to site-specific insertion of amino acid analogues into proteins, *Proc. Natl. Acad. Sci. U.S.A.* **98,** 14310–5 (2001).

21. Edwards, H. *et al.*, A bacterial amber suppressor in *Saccharomyces cerevisiae* is selectively recognized by a bacterial aminoacyl-tRNA synthetase, *Mol. Cell. Biol.* **10,** 1633–41 (1990).

22. Edwards, H. *et al.*, An *Escherichia coli* tyrosine transfer RNA is a leucine-specific transfer RNA in the yeast *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. U.S.A.* **88,** 1153–1156 (1991).

23. Liu, W. *et al.*, Genetic incorporation of unnatural amino acids into proteins in mammalian cells, *Nat. Methods* **4,** 239–244 (2007).

24. Young, T. S. *et al.*, Beyond the canonical 20 amino acids: expanding the genetic lexicon, *J. Biol. Chem.* **285,** 11039–44 (2010).

25. Skinner, D. Z. *et al.* Enzymatic Inverse PCR: A restriction site independent, single fragment method for high-efficiency, site-directed mutagenesis, *Biofeedback* **13,** (1992).

Chapter 6

The synthesis of 'click' functionalized nanobodies by amber suppression

6.1 Introduction

In this chapter the azide modification of the nanobody[1–6] BCII10 (NbBCII10[7]) in *S. cerevisiae* will be discussed. This part of the study is marked by the red frame in Figure 6.1. The site-specific modification of the NbBCII10 is based on the incorporation of *p*-azidoPhe by means of amber suppression[8–11]. This is done by expanding the genetic code of *S. cerevisiae*[12–17] with an orthogonal, mutant *E. coli* tyrosyl-tRNA synthetase (EcTyrRS)/tRNA$_{CUA}$^{Tyr} pair. The EcTyrRS/tRNA$_{CUA}$^{Tyr} pair is added to the MaV203 yeast strain[18,19] as part of the yeast episomal plasmid pEcTyrRS(mutant)/tRNA$_{CUA}$^{Tyr}, as described in section 3.2.1, while the methodology of amber suppression was described in detail in section 2.2. When amber suppression is mentioned in this chapter, it always refers to amber suppression performed by the *E. coli* EcTyrRS/tRNA$_{CUA}$^{Tyr} pair. Only when "endogenous amber suppression in yeast" is explicitly mentioned, it refers to the suppression performed by the yeast synthetase and yeast tRNA.

It was first verified whether the MaV203 cells are capable of expressing the wildtype nanobody, and the stability of the protein expression was assessed. The expression of full-length functionalized Green Fluorescent Protein (GFP) was used as a proof of principle to find out whether or not the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair can be used to produce modified proteins in relevant concentrations.
The construction of the vectors that are used in the expression experiments of this chapter are explained in detail in chapter 3.
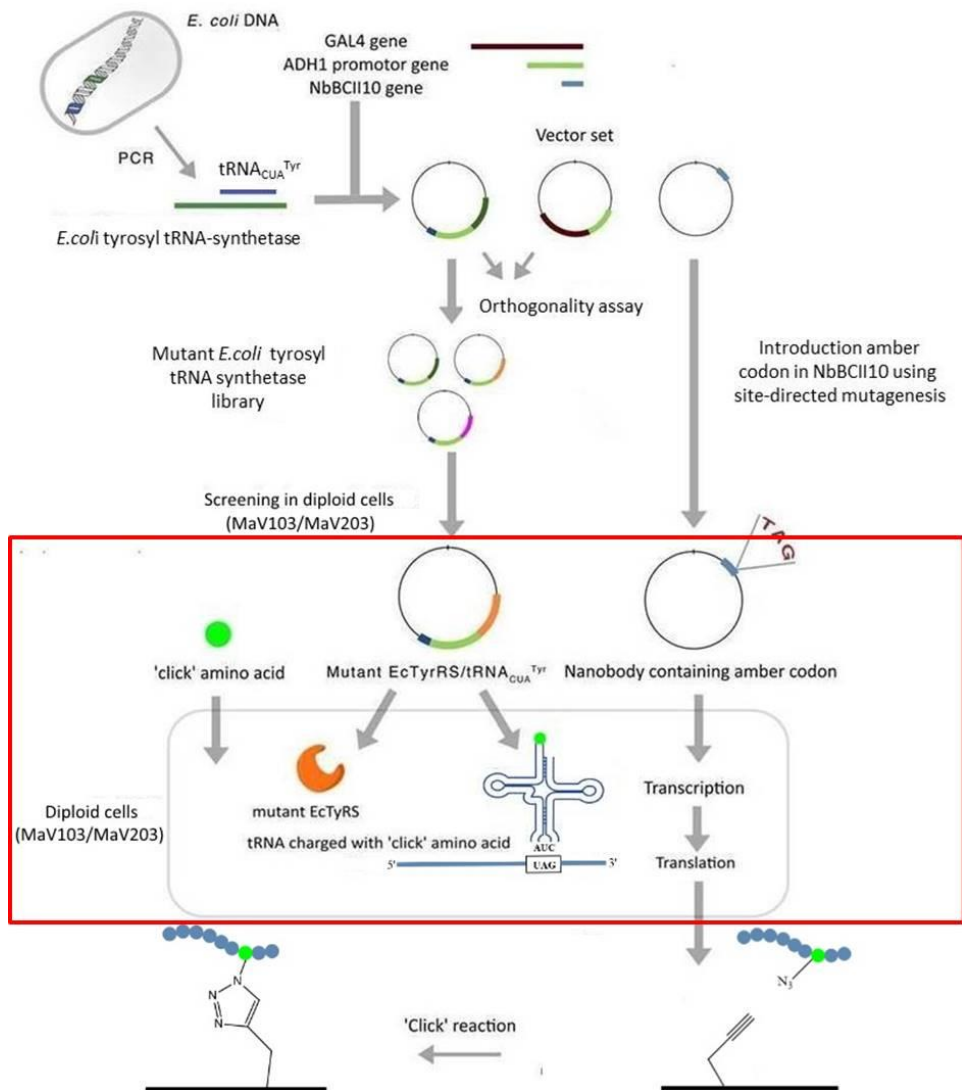
Figure 6.1. Overview of the thesis with highlight on the synthesis of the site-specifically azidified NbBCII10.

6.2 Results and discussion

6.2.1 Expression of unmodified BCII10 nanobody

As it is important to verify if the MaV203 cells are capable of expressing and secreting the unmodified BCII10 nanobody (NbBCII10_His$_6$) to the growth medium, MaV203 cells were transformed with the pTEF-MF/NbBCII10_His$_6$ vector according the LiAc/SS-DNA/PEG procedure[20]. In case the MaV203 cells would not be able to do this, another yeast strain would be required.

The construction and composition of the pTEF-MF/NbBCII10_His$_6$ vector is described in detail in section 3.2.2. It contains the wildtype NbBCII10 gene with a C-terminal His$_6$-tag, under the control of the constitutive TEF1 promoter[21]. MF is the yeast mating factor α leader sequence, which is fused to the N-terminus of NbBCII10_His$_6$, causing it to be secreted to the medium. During NbBCII10_His$_6$ secretion, the MF α leader sequence will be cleaved.

After transformation, three colonies were selected and liquid cultures were grown overnight at 30 °C. An extra liquid culture was made of untransformed MaV203 cells as a negative control for the SDS-PAGE and Western Blot analysis. After centrifugation (5min, 3000 r.p.m.), the pellet consisted of the MaV203 cells, while the supernatant contained the NbBCII10_His$_6$.

SDS-PAGE and Western Blot analysis of NbBCII10_His$_6$. The NbBCII10_His$_6$ has a molecular weight of approximately 15 kDa. SDS-PAGE was performed to separate the proteins using the Mini Protein Tetra System (Biorad). After the separation of the proteins (15% separating gel and 4% stacking gel), a western blot analysis was performed.

The NbBCII10_His$_6$ was transferred to a polyvinylidene fluoride (PVDF) membrane and detected using primary mouse IgG monoclonal anti-histidine antibody (1 µg/ml) in combination with the secondary antibody anti-mouse IgG-alkaline phosphatase (300 ng/ml).

The NbBCII10_His$_6$ was visualized by the conversion of 5-bromo-4-chloro-3-indolylphosphate/nitrobluetetrazolium(NBT/BCIP) into a colored precipitate by the alkaline phosphatase (AP) that is conjugated to the secondary anti-mouse IgG antibody. Figure 6.2 shows that the MaV203 cells transformed with the pTEF-MF/NbBCII10_His$_6$ vector (lanes 1-3) express NbBCII10_His$_6$ and successfully export it to the growth medium.

The negative control (untransformed MaV203 cells: lane 4) shows no band at the position of 15 kDa. This indicates that the MaV203 cells can be used to produce NbBCII10_His$_6$.
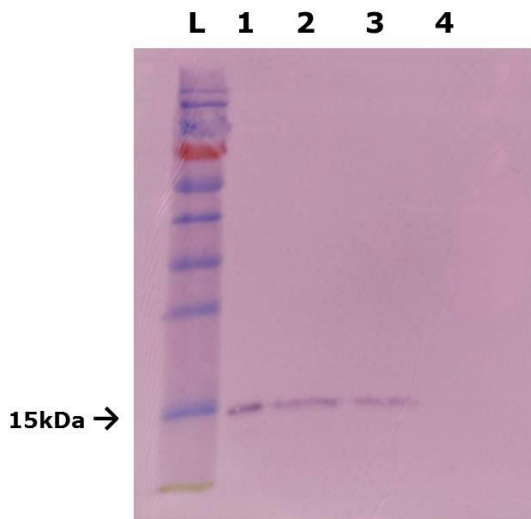
Figure 6.2. Western blot analysis of the NbBCII10_His$_6$ expressed by the MaV203 cells. The C-terminal His$_6$-tag is detected using the mouse IgG monoclonal anti-histidine antibody. A clear band of NbBCII10_His$_6$ is visible at 15 kDa in lanes 1, 2 and 3. L=Prestained Protein Ladder, lane 1: MaV203 pTEF-MF/NbBCII10_His$_6$ (clone 1), lane 2: MaV203 pTEF-MF/NbBCII10_His$_6$ (clone 2), lane 3: MaV203 pTEF-MF/NbBCII10_His$_6$ (clone 3), lane 4: untransformed MaV203.

6.2.2 Determination of the stability of heterologous protein expression in MaV203 cells

Sometimes protein expression tends to decrease in yeast cells after a number of cell divisions. This was also noticed for the expression of NbBCII10_His$_6$. In order to investigate this, a series of western blot analyses was performed. As was indicated in Figure 6.2, the clones 1, 2 and 3 successfully express NbBCII10_His$_6$. Therefore, clones 2 and 3 were used to further investigate the stability of the NbBCII10_His$_6$ expression. These two clones were replated four and eight times on single drop-out SD-URA plates. Afterwards, a liquid culture was made of both clones and a western blot analysis was performed on the growth medium. Figure 6.3 illustrates the change of protein expression after a number of cell divisions. It shows the comparison of the NbBCII10_His$_6$ expression between the freshly transformed clones and after replating these clones four and eight times.
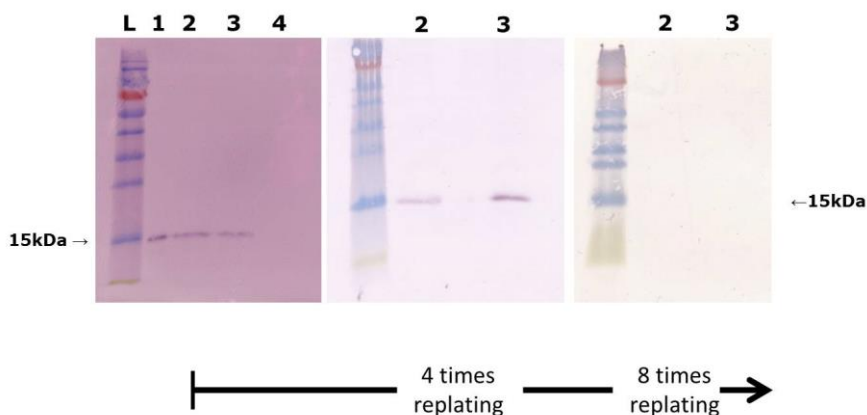
Figure 6.3. Western blot analysis to investigate the stability of the NbBCII10_His$_6$ expression in MaV203 cells. It shows the comparison of the NbBCII10_His$_6$ expression between freshly transformed clones and after replating these clones four and eight times. After replating the cells, it is seen that none of the clones is able to express NbBCII10_His$_6$.

These experiments indicate that the expression of NbBCII10_His$_6$ decreases as a function of the number of cell divisions. Therefore, only freshly transformed cells will be used for future expression experiments.

6.2.3 Suppression of the amber codon in NbBCII10_His$_6$

In this section, the protocol to verify amber suppression in NbBCII10_His$_6$ is explained. First the MaV203 cells are transformed with the vectors containing the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair and NbBCII10_TAG (NbBCII10_His$_6$ with an amber codon located between the nanobody and the C-terminal His$_6$-tag at). Secondly, liquid cultures of these cells are grown for 48 hours, ensuring the cells to reach the stationary phase. In case of successful amber suppression, the cells will produce modified NbBCII10_His$_6$ and, because of the MF α leader sequence located upstream of the NbBCII10_His$_6$, secrete it into the growth medium. Thus, the longer the cells are grown, the higher the concentration of modified NbBCII10_His$_6$ in the growth medium. In a last step, the cell culture is centrifuged and the supernatant (growth medium) is analyzed by means of western blot.

Western Blot analysis of modified NbBCII10_His$_6$. For the synthesis of modified NbBCII10_His$_6$, the glycine 132 codon was mutated into an amber codon. As a result, the amber codon is located between the nanobody and the C-terminal His$_6$-tag, as shown in Figure 3.21 of

section 3.2.2. In case of successful amber suppression, the produced nanobodies will contain a C-terminal His_6-tag, allowing for their detection with mouse IgG monoclonal anti-histidine antibody. However, if the amber codon is not suppressed, no His_6-tag will be present and the nanobodies will not be detected by Western Blot analysis.

For the protein expression experiments, a mutant EcTyrRS of which the affinity for p-azidoPhe was described in literature[11], was used. To obtain this mutant, three amino acids located in the binding pocket had to be mutated, namely Y37T; D182S and F183A[22]. The construction of this mutant EcTyrRS has been described of section 3.2.1.

As a first experiment, it was verified that the EcTyrRS/$tRNA_{CUA}^{Tyr}$ pair is capable of incorporating natural tyrosine in NbBCII10_His_6 in detectable levels. The mutant EcTyrRS/$tRNA_{CUA}^{Tyr}$ pair was used in the absence of p-azidoPhe as a negative control. In this context, the following vectors of which the composition and construction was described in chapter 3, were used:

> I. **pTEF-MF/NbBCII10_His_6**: containing the wildtype NbBCII10_His_6 gene.
>
> II. **pTEF-MF/NbBCII10_TAG**: containing the NbBCII10_His_6 gene with an amber codon located between the C-terminus of the nanobody and the His_6-tag.
>
> III. **pEcTyrRS/$tRNA_{CUA}^{Tyr}$**: containing the genes of the wildtype EcTyrRS/$tRNA_{CUA}^{Tyr}$ pair.
>
> IV. **pEcTyrRS(mutant)/$tRNA_{CUA}^{Tyr}$**: containing the genes of the mutant EcTyrRS/$tRNA_{CUA}^{Tyr}$ pair with affinity for p-azidoPhe.

MaV203 cells were transformed[20] with the following vector combinations:

> 1. pTEF-MF/NbBCII10_His_6 (**I**) (positive control)
>
> 2. pTEF-MF/NbBCII10_TAG (**II**) + pEcTyrRS/$tRNA_{CUA}^{Tyr}$ (**III**)
>
> 3. pTEF-MF/NbBCII10_TAG(**II**)+ pEcTyrRS(mutant)/$tRNA_{CUA}^{Tyr}$ (**IV**) (negative control in the absence of p-azidoPhe)

After transformation, three colonies of each of the combinations were selected and liquid cultures were made in the appropriate selective media. The MaV203 cells with pTEF-MF/NbBCII10_His$_6$ (**I**) served as positive control. The MaV203 cells containing both pTEF-MF/NbBCII10_TAG and pEcTyrRS(mutant)/tRNA$_{CUA}$^{Tyr} (**II** + **IV**) served as a negative control in the absence of *p*-azidoPhe. After centrifugation of the grown cultures the supernatant (growth medium) was analyzed with Western Blot. Table 6.1 gives an overview of the samples analyzed with Western Blot in Figure 6.4.

Table 6.1. Overview of the samples analyzed with Western Blot. The numbers in the table correlate with the lane numbers of the Western blot shown in Figure 6.4. The vector combinations are described in the text.

Lane	MaV203 transformed with vectors
1	Ladder
2	**II + IV** (clone 1) (negative control)
3	**II + IV** (clone 2) (negative control)
4	**II + IV** (clone 3) (negative control)
5	/
6	**II + III** (clone 1)
7	**II + III** (clone 2)
8	**II + III** (clone 3)
9	/
10	**I** (positive control)

From Figure 6.4 it can be observed that only the positive control shows a clear band at 15 kDa (lane 10). In case the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair was able to suppress the amber codon by incorporating natural tyrosine, clear bands were expected in lanes 6, 7 and 8. These MaV203 cells contained both the pTEF-MF/NbBCII10_TAG vector as well as the pEcTyrRS/tRNA$_{CUA}$^{Tyr} vector (**II + III**). The faint bands present in lanes 6 and 7 are probably the result of only endogenous amber suppression in yeast[23–25] and not by suppression performed by the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair, since similar faint bands were also observed in

lanes 2, 3 and 4. These MaV203 cells contained both the pTEF-MF/BCII10_TAG vector and the pEcTyrRS (mutant)/tRNA$_{CUA}^{Tyr}$ vector (**II + IV**) and served as a negative control.
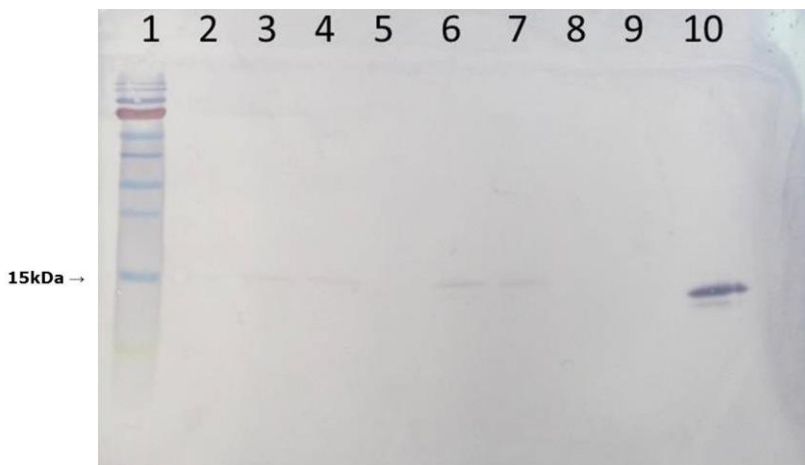
Figure 6.4. Western Blot analysis of NbBCII10_His$_6$. The numbering of the lanes matches with the cells described in Table 6.1. A clear band is only visible for MaV203 pTEF-MF/NbBCII10_His$_6$ (positive control) (lane 10). The faint bands present in the other lanes are probably a result of endogenous amber suppression in yeast.

Western Blot analysis after concentrating NbBCII10_His$_6$. Since no amber suppression (performed by the EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair) was detected in the previous experiment, it was decided to concentrate the NbBCII10_His$_6$ by freeze-drying before SDS-PAGE and Western Blot analysis. Since *S. cerevisiae* does not secrete any proteins to the growth medium besides the NbBCII10_His$_6$, no other proteins were expected to interfere with the analysis.

As in the previous experiment, the MaV203 cells were transformed with the following vector combinations:

1. pTEF-MF/NbBCII10_His$_6$ (**I**) (positive control)

2. pTEF-MF/NbBCII10_TAG(**II**) + pEcTyrRS/tRNA$_{CUA}^{Tyr}$ (**III**)

3. pTEF-MF/NbBCII10_TAG (**II**)+pEcTyrRS(mutant)/tRNA$_{CUA}^{Tyr}$ (**IV**)

After transformation, three colonies of each of the vector combinations
were selected and liquid cultures were made in the appropriate selective
media and grown at 30 °C. In this experiment it was also verified if *p*-
azidoPhe is incorporated. Therefore, after 24 hours, the cultures of the
MaV203 cells containing vectors (**II + IV)** were split into two fractions.
To one of the fractions, *p*-azidoPhe was added (1 mM), while fraction
two was kept on selective media without this modified amino acid. When
reaching the stationary phase the cells were centrifuged and the
supernatant was freeze-dried. The freeze-dried residue was analyzed
with Western Blot (in duplicate). Table 6.2 gives an overview of the
samples shown in the Western Blot in Figure 6.5.

Table 6.2. Overview of the freeze-dried samples (growth medium of the MaV203 cells
transformed with the vector combinations **I**, **II**+**III** and **II**+**IV**) analyzed with Western Blot.
The numbers in the table correlate with the lane numbers in Figure 6.5.

Lane	MaV203
1	Ladder
2	**II + IV** + 1mM *p*-azidoPhe (clone 1)
3	**II + IV** + 1mM *p*-azidoPhe (clone 2)
4	**II + IV** (clone 1) (negative control)
5	**II + IV** (clone 2) (negative control)
6	**II + III** (clone 1)
7	**II + III** (clone 2)
8	**II** (clone 1) (negative control)
9	**II** (clone 2) (negative control)
10	**I** (positive control)
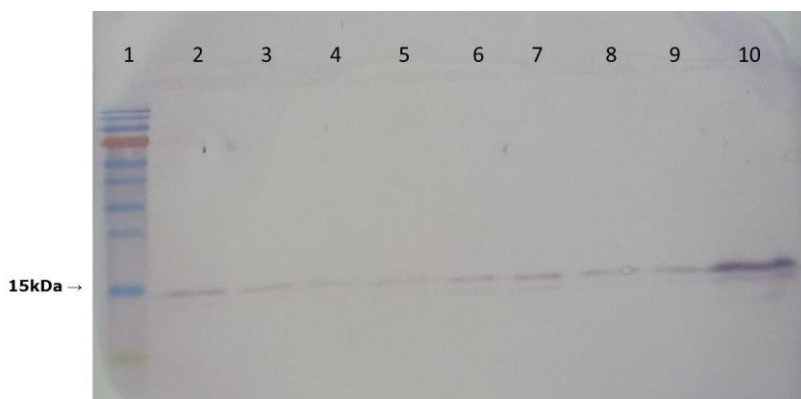
Figure 6.5. Western Blot analysis of the concentrated NbBCII10_His$_6$. The numbering on the figure matches with the lane numbers in Table 6.2. A clear band is visible only for MaV203 pTEF-MF/NbBCII10_His$_6$ (**I**) (positive control, lane 10). The faint bands present in the other lanes are probably the result of endogenous amber suppression in yeast.

Figure 6.5 only shows a clear band at the position of 15 kDa for the positive control (**I**) (lane 10). Furthermore, in case the pEcTyrRS(mutant)/tRNA$_{CUA}^{Tyr}$ pair was capable of suppressing the amber codon by introducing the modified amino acid, clear bands were expected in lanes 2 and 3. Moreover, if the wildtype EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair was capable of suppressing the amber codon by introducing natural tyrosine, clear bands should be present in lanes 6 and 7. The faint bands seen in lanes 2, 3, 6 and 7 are probably the result of endogenous amber suppression in yeast and not from the amber suppression by the EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair, since similar faint bands are visible in lanes 8 and 9 for cells only containing the pTEF-MF/NbBCII10_TAG vector (**IV**) (negative control).

A possible explanation for the inefficient amber suppression might be that the expression of NbBCII10_His$_6$ is decreased as a result of a recombination of the plasmids or a drop in copy number of the plasmids (discussed in a personal communication with Dr. Aliona Bogdanova, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany). This might be a consequence of the pressure the cells have to cope with for propagating two types of plasmids. The sequence of the pTEF-MF/NbBCII10_His$_6$ vector has been send for verification to Edward Dolk from QVQ BV (Utrecht, the Netherlands), who confirmed us in a personal communication that all the necessary components were present in the vector and that the sequence of NbBCII10_His$_6$ (S8)

should not cause any problems for expression. However, it was noticed that the NbBCII10_His$_6$ contained the amino acids methionine and alanine at the N-terminus (S13 of the appendix). Although these amino acids should not be present, their presence will not interfere with the expression of NbBCII10_His$_6$ (personal communication Prof. Dr. Muyldermans, VUB).

Another possible explanation is that one of the components of the suppression system (EcTyrRS or tRNA$_{CUA}^{Tyr}$) is not expressed properly. So far, tRNA$_{CUA}^{Tyr}$ expression was performed on a basal level. This proved to be enough for the proof of principle assay, performed in section 4.2.6, but it might not be sufficient for the production of detectable levels of modified proteins. It is known that prokaryotic tRNA is expressed in a different way than eukaryotic tRNA[26]. As mentioned in section 3.2.1 the *E. coli* tRNA$_{CUA}^{Tyr}$ can be efficiently expressed in eukaryotic cells if an external promoter is added, containing an A- and B-box in front of the gene. In the following study the eukaryotic promoter sequence of SNR52 (small nuclear RNA), combined with the SUP4 terminator (SUP4 = yeast tRNA^{Tyr} gene) sequence were added to elevate the activity of the *E. coli* tRNA$_{CUA}^{Tyr}$[27,28]. This resulted in the constructs pEcTyrRS/ SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t (**V**) and pEcTyrRS(mutant)/ SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t (**VI**). The expected elevated activity of *E. coli* tRNA$_{CUA}^{Tyr}$ was evaluated using Green Fluorescent Protein (GFP), containing an amber codon at triplet location 39.

6.2.4 Amber suppression in Green Fluorescent Protein: proof of principle

As mentioned in Chapter 3, Green Fluorescent Protein (GFP) containing an amber codon at location 39 and a C-terminal His$_6$-tag (GFP_Y39TAG) was used as a proof of principle experiment. If the amber codon is successfully suppressed, full-length GFP is expressed, resulting in fluorescence when exposed to UV-light of 365 nm. If amber suppression is not successful, the premature termination will result in a non-fluorescent truncated GFP molecule. Therefore, by expressing GFP_Y39TAG in the cytosol of MaV203 cells, successful amber suppression can be verified by monitoring the fluorescence of the cells. Two types of vectors were used in these experiments:

- pADH1/GFP (**VII**) and pADH1/GFP_Y39TAG (**VIII**)

- pTEF/GFP (**IX**) and PTEF/GFP_Y39TAG (**X**)

*Cytosolic expression of GFP using pADH1/GFP (**VII**) and pADH1/GFP_Y39TAG (**VIII**).* In these experiments, GFP containing a C-terminal His$_6$-tag was expressed in the cytosol of MaV203 cells using the vectors pADH1/GFP and pADH1/GFP_Y39TAG. These constructs have been discussed in section 3.3.2.

The following vectors were used for the cytosolic expression of GFP:

V. pEcTyrRS/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t

VI. pEcTyrRS(mutant)/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t

VII. pADH1/GFP

VIII. pADH1/GFP_Y39TAG

MaV203 cells were transformed with the following vector combinations:

- pADH1/GFP (**VII**) (positive control)

- pADH1/GFP_Y39TAG (**VIII**) (negative control)

- pEcTyrRS/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t (**V**) + pADH1/GFP_Y39TAG (**VIII**)

- pEcTyrRS(mutant)/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t (**VI**) + pADH1/GFP_Y39TAG (**VIII**)

After transformation, three colonies of each combination were selected and liquid cultures were made in the appropriate selective media. After 24 hours incubation, the liquid MaV203 cultures containing vectors **VI** + **VIII** were split into two fractions. To one of the fractions, *p*-azidoPhe (1 mM) was added, while the other fraction was kept on selective medium without the modified amino acid. When the cells reached the stationary phase both the cell density and fluorescence were measured. The resulting fluorescence data per cell density of 1 A (absorbance at OD$_{600}$)

is plotted in Figure 6.6, after normalization against the fluorescence
value of the positive control (MaV203 containing pADH1/GFP (**VII**)).
Table 6.3 gives an overview of the vector combinations used and the
resulting fluorescence values.

Table 6.3. Overview of the MaV203 cells of which the fluorescence was measured. The
fluorescence data are normalized against the fluorescence of the positive control
(expression of wildtype GFP). The numbers correspond with those used in Figure 6.6.

	MaV203	%
1	**VII** (positive control)	100.0
2	**VIII** (negative control)	33.7
3	**V** + **VIII** (clone 1)	33.5
4	**V** + **VIII** (clone 2)	37.7
5	**VI** + **VIII** + 1 mM *p*-azidoPhe (clone 1)	40.0
6	**VI** + **VIII** + 1 mM *p*-azidoPhe (clone 2)	46.0
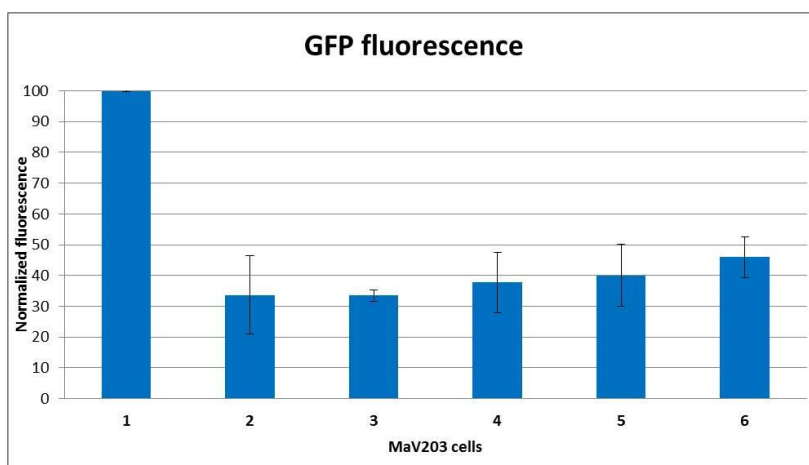
Figure 6.6. Graphical representation of the fluorescence data of the MaV203 cells listed in Table
6.3, normalized to the fluorescence of the positive control.

The normalized fluorescence values of the different MaV203 cells ranged from 30% to 45%, and there is no significant difference between the negative control (2) and the other MaV203 cells. The addition of the modified amino acid and use of the mutant EcTyrRS does not cause a significant increase in fluorescence.

*Cytosolic expression of GFP using pTEF/GFP (**IX**) and pTEF/GFP_Y39TAG (**X**).* Another vector for the cytosolic expression of GFP containing a C-terminal His$_6$-tag was evaluated. The GFP gene was cloned into the pTEF-MF vector from which the mating factor α leader sequence had been removed, resulting in the cytosolic expression of GFP. The construction of the vectors is discussed in section 3.3.1. The following vectors were used for the cytosolic expression of GFP:

V.   pEcTyrRS/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t

VI.   pEcTyrRS(mutant)/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t

IX.   pTEF/GFP

X.   pTEF/GFP_Y39TAG

MaV203 cells were transformed with the following vector combinations:

- pEcTyrRS/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t (**V**) + pTEF/GFP (**IX**) (positive control)

- pEcTyrRS(mutant)/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t (**VI**) + pTEF/GFP_Y39TAG (**X**)

- pTEF/GFP_Y39TAG (**X**) (negative control)

As a positive control the vector combination (**V**+**IX**) was used. The pTEF/GFP vector (**IX**) expresses full-length GFP while the pEcTyrRS/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t (**V**) vector has no direct influence on the GFP expression. However, in order to make sure that the positive control experiences the same amount of pressure as the other cells (propagation of two plasmids), also two types of plasmids (**V**+**IX**) were introduced. After transformation, colonies of each combination were selected and liquid cultures were made in the appropriate selective media. After 24 hours of incubation, the liquid cultures of the MaV203

cells containing the vectors (**VI**+**X**) were split into two fractions. To one of the fractions *p*-azidoPhe (1 mM) was added, while the other fraction was kept on selective medium without the modified amino acid. When the cells reached the stationary phase, 1 ml of cell suspension was centrifuged (5 minutes at 3000 r.p.m.), and the cell pellet was irradiated with UV-light of 365nm. In case of successful amber suppression, full-length GFP is produced and the cells will show green fluorescence. Table 6.4 gives an overview of the cell pellets shown in Figure 6.7. The remaining part of the cell culture was used to extract the proteins using the Y-PER Yeast Protein Extraction Reagent (Life Technologies).

Table 6.4. Overview of the MaV203 cells used to verify amber suppression in GFP. The numbers in this table match with the numbers of the "eppendorf" tubes presented in Figure 6.7.

	MaV203
1	**V + IX** (positive control)
2	**VI + X** (clone 1)
3	**VI + X** (clone 1) +1mM *p*-azidoPhe
4	**VI + X** (clone 2)
5	**VI + X** (clone 2) +1mM *p*-azidoPhe
6	**VI + X** (clone 3)
7	**VI + X** (clone 3) +1mM *p*-azidoPhe
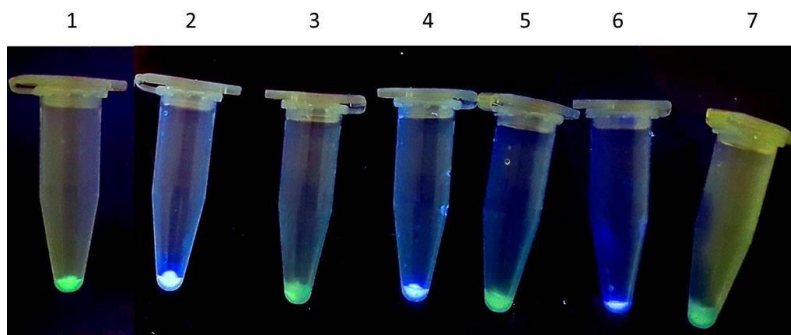
Figure 6.7. Cell pellets of MaV203 cells described in Table 6.4, exposed to UV-light. In case of successful amber suppression, full-length GFP is expressed in the cytosol and the cells will show fluorescence when exposed to UV-light. The numbers used in this figure match with the numbers in Table 6.4. Number 1 is the positive control. It is clear that amber suppression is successful in the presence of *p*-azidoPhe (numbers 3, 5 and 7), while it is not occurring in the absence of the modified amino acid (numbers 2, 4 and 6). This indicates that the EcTyrRS(mutant)/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t pair selectively incorporates *p*-azidoPhe in response to an amber codon.

From Figure 6.7, it can be seen that the clones containing pTEF/GFP_Y39TAG + pEcTyrRS(mutant)/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t (**IV**+**X**), present efficient amber suppression in the presence of *p*-azidoPhe. They produce full-length GFP and show fluorescence when exposed to UV-light (Figure 6.7: numbers 3, 5 and 7). When *p*-azidoPhe was absent in the growth medium containing all natural amino acids, no fluorescence was visible (Figure 6.7: numbers 2, 4 and 6). This shows that the amber suppression system is successful in these cells and that the mutant EcTyrRS[11,22] selectively incorporates the modified amino acid and not the natural tyrosine. For protein extraction purposes, the cells listed in Table 6.4 were plated on appropriate selective media, in order to preserve them for further experiments. This was done, since it is very difficult to select a clone capable of performing successful amber suppression. Even for MaV203 cells containing pTEF/GFP + pEcTyrRS/SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t (**V**+**IX**) (positive control), the levels of GFP expression varied strongly between different clones. Some clones expressed high levels of GFP (as shown in Figure 6.7 number 1), while others showed intermediate levels or no GFP expression at all. A plausible reason for this is that protein expression levels are very clone-specific in *S. cerevisiae*[29] (discussed in a personal communication with Dr. Aliona Bogdanova, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany). Therefore, selecting a clone that

efficiently brings both the GFP protein as well as the EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair to expression is very difficult.

Native PAGE. A protein extraction of the cells listed in Table 6.4 and shown in Figure 6.7 was performed using the Y-PER Yeast Protein Extraction Reagent. These extracts were analyzed by means of native PAGE (15% separating gel and 4% stacking gel), taking into account that GFP_His$_6$ has a molecular weight of 28.5 kDa and a pI of 5.7. Figure 6.8 shows the native PAGE under UV-light exposure.
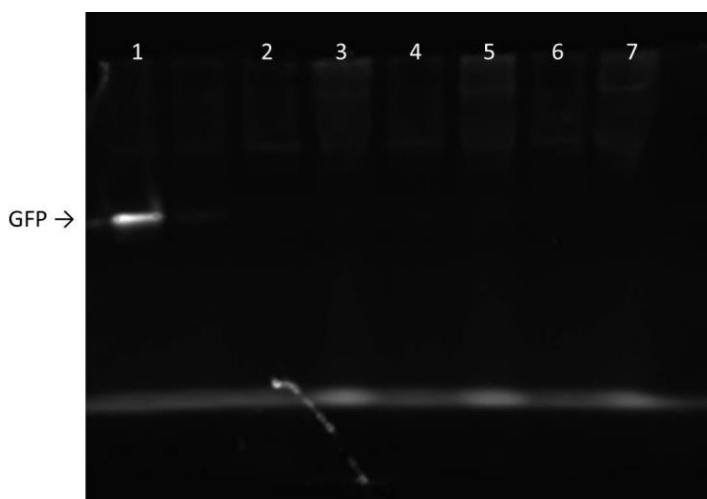
Figure 6.8. Native PAGE performed on the cell extracts of the MaV203 cells listed in Table 6.4. Only a clear fluorescent band is visible for the cell extract of the MaV203 cells expressing the wildtype GFP (lane 1). However, a thicker fluorescent band can be observed at the position of the running front in lanes 3, 5 and 7. These lanes contain the extracts of the MaV203 cells that proved to successfully incorporate *p*-azidoPhe in GFP.

A clear fluorescent band is only visible for the positive control (lane 1). A fluorescent band was also expected for lanes 3, 5 and 7 since these samples showed fluorescence before being loaded on the native PAGE gel. However, thicker fluorescent bands can be observed at the position of the running front in lanes 3, 5 and 7. It is not expected that changing one amino acid (tyrosine → *p*-azidophenylalanine) can have such a large effect on the pI of a protein. However, the modified amino acid is introduced in the interior of the GFP structure. This might have an influence on the tertiary structure of GFP causing it to expose other amino acid side chains on the exterior, compared to the wildtype GFP,

while maintaining its fluorescence. Further investigation is needed to confirm this hypothesis.

Western Blot analysis of the cytosolically expressed GFP_His$_6$. In the final step of the protein extraction the mixture is centrifuged and the resulting pellet should contain the cell debris and DNA, while the cytosolic proteins (including GFP_His$_6$) should be present in the supernatant. However, GFP was still partially present in the pellet after a first protein extraction. Therefore a second extraction was performed on the same cell pellets. The supernatant from the first and the second extraction as well as the pellet were analyzed by Western Blot. Since the full-length GFP contains a C-terminal His$_6$-tag, it can be detected with mouse IgG monoclonal anti-histidine antibody. Table 6.5 gives an overview of the samples analyzed by Western Blot. Figure 6.9 shows the result of the Western Blot.

Table 6.5. Overview of the samples analyzed by Western Blot. The numbering of the samples matches with the lanes on Figure 6.9.

	MaV203
1	Ladder
2	**V + IX** (positive control) (extract 1)
3	**V + IX** (positive control) (extract 2)
4	**VI + X** (clone 1)+1mM *p*-azidoPhe (extract 1)
5	**VI + X** (clone 1)+1mM *p*-azidoPhe (extraction 2)
6	**VI + X** (clone 1)+1mM *p*-azidoPhe (pellet)
7	**VI + X** (clone 2)+1mM *p*-azidoPhe (extraction 2)
8	**VI + X** (clone 2)+1mM *p*-azidoPhe (pellet)
9	**VI + X** (clone 3)+1mM *p*-azidoPhe (pellet)

Figure 6.9. Western Blot analysis of the protein extracts of the cells mentioned in Table 6.5. A clear band is visible for the first extract of the positive control (lane 2) and a faint band for the second extract of the positive control (lane 3). GFP does not seem to be present in the extraction products of the other cells.

The results obtained with this Western Blot (Figure 6.9) are similar to what was seen on the native PAGE (Figure 6.8). Again, only a band of GFP is visible for the positive control (lane 2: extraction 1 and lane 3: extraction 2). The other extracts did not seem to contain full-length GFP. Also here, bands were expected in lanes 4, 5 and 7 since the extracts showed fluorescence before being loaded on the gel. Fluorescence indicates that the amber codon has been successfully suppressed and that full-length GFP (including the His_6-tag) should have been expressed. In order to clarify this, further investigation was carried out. Hereto, the cells listed in Table 6.4 were plated on appropriate selective media. Liquid cultures of these cells were made in order to further investigate the amber suppression in GFP. Surprisingly, the cells seemed to have lost the ability to express full-length GFP after being replated once. Even the positive control cells containing pEcTyrRS/ $SNR52\_tRNA_{CUA}^{Tyr}\_SUP4t$ + pTEF/GFP (**V**+**IX**) had lost the ability to express GFP and did not show any fluorescence anymore. This indicates that the suppression system is not stable. Even though clones that initially had the ability to successfully suppress the amber codon were isolated, this ability seems to disappear rather fast. Even after repeating the transformation of MaV203 cells with these vector combinations and selecting numerous clones, it was not possible to select clones capable

of suppressing the amber codon, confirming the difficulty of selecting a
clone that expresses all the components needed for amber suppression.
Cytosolic expression of GFP in the yeast strain INVSc1. In order to verify
if the MaV203 yeast strain is responsible for the lack of stability of the
components of the amber suppression system, the INVSc1 yeast strain
(Life Technologies) was used for the cytosolic expression of GFP. The
INVSC1 strain is a fast-growing, diploid strain optimized for protein
expression. The INVSc1 strain was transformed with the following vector
combinations:

1. pEcTyrRS/SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t (**V**) + pTEF/GFP (**IX**)
   (positive control)

2. pEcTyrRS(mutant)/SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t(**VI**) + pTEF_GFP_Y39TAG
   (**X**)

3. pTEF_GFP_Y39TAG (**X**) (negative control)

However, after transformation, it was not possible to select a clone with
the ability to suppress the amber codon in GFP. Even the positive control
cells, transformed with pEcTyrRS/ SNR52_tRNA$_{CUA}^{Tyr}$_SUP4t + pTEF/GFP
(**V**+**IX**), showed only expression of GFP in about 25% of the selected
cells.

6.3 Conclusions

The results discussed in this chapter indicate that an *in vivo* system
based on expanding the genetic code of *S. cerevisiae* with an orthogonal
EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair, using two yeast episomal plasmids, is not
efficient. It has been observed, as described in literature[29], that protein
expression is clone-specific in *S. cerevisiae*, making it very difficult to
select clones that express all the components of the suppression system
at sufficient levels. When yeast cells are transformed with a yeast
episomal plasmid for the expression of a heterologous protein, the level
of protein expression seems to vary strongly from high, over
intermediate to no expression at all for different clones (even though all
of them contain the plasmid). In order to obtain a successful amber
suppression, three components need to be expressed in sufficient levels:

    i.   the EcTyrRS,

    ii.   the *E. coli* $tRNA_{CUA}^{Tyr}$ and,

    iii.   the target gene containing the amber codon.

If any of these components is not properly expressed, amber suppression will not occur. In this study, very few clones expressed all three components in sufficiently high amounts to produce detectable levels of modified proteins. This explains why it was necessary to select a large number of clones to find a clone that was able to successfully supress the amber codon in GFP.

Moreover, the results show that even when a suited clone is found, the expression of the components of the suppression system is very unstable. It was observed that the ability of the MaV203 cells to express NbBCII10 decreased over time, until the NbBCII10 expression stopped. For the suppression of the amber codon in GFP, these findings were even more pronounced, *i.e.* replating the cells once resulted in a complete loss of amber suppression.

A possible explanation might be that, in the case of wildtype NbBCII10_His_6, the expression is performed using a single yeast episomal plasmid (pTEF-MF/NbBCII10_His_6). The propagation of the plasmid and the expression of the heterologous protein induces pressure on the cells. To cope with this pressure, the cells might slowly decrease the expression of the heterologous protein (e.g. NbBCII10 of GFP) (discussed in a personal communication with Dr. Aliona Bogdanova, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany).

In case of amber suppression, three components are expressed by means of two yeast episomal plasmids. The propagation of two plasmids and the expression of the EcTyrRS, the $tRNA_{CUA}^{Tyr}$ and the NbBCII10_His_6 probably result in a much higher pressure on the cells. This might result in a faster down-regulation of the expression of the components needed for a successful amber suppression.

Although more investigation is needed to elucidate the exact mechanism of this plasmid loss, a recombination of the plasmids or a drop in copy number of the plasmids might be an explanation. It is known that e.g. basal expression of the URA3 gene is enough for the cells to survive on SD-URA. Therefore, low copy numbers of the plasmids might be

sufficient for the cells to survive on the selective media, but not for the expression of detectable amounts of heterologous proteins.

In general, it can therefore be concluded that more stable alternatives than yeast episomal plasmids have to be used to expand the genetic code of *S. cerevisiae*. A possible alternative could be the use of inducible promoters to express the components in combination with so-called Yeast Integrating Plasmids (YIp). These YIps get stably integrated into the chromosomal DNA of yeast and do not need any auxotrophic marker to be propagated in the cell. However, more investigation is clearly required.

6.4 References

1.   Hamers-Casterman, C. *et al.,* Naturally occurring antibodies devoid of light chains, *Nature* **363,** 446–448 (1993).

2.   Muyldermans, S. *et al.,* Camelid immunoglobulins and nanobody technology, *Vet. Immunol. Immunopathol.* **128,** 178–83 (2009).

3.   Woolven, B. P. *et al.*, The structure of the Ilama heavy chain constant genes reveals a mechanism for heavy-chain antibody formation, *Immunogenetics.* **50**, 98–101 (1999).

4.   Muyldermans, S. *et al.*, Single domain camel antibodies: Current status, *Rev. Mol. Biotechnol.* **74,** 277–302 (2001).

5.   Muyldermans, S. *et al.*, Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains, *Protein Eng.* **7,** 1129–1135 (1994).

6.   Deffar, K. *et al.*, Nanobodies - the new concept in antibody engineering, *J. Biotechnol.* **8,** 2645–2652 (2009).

7.   Dumoulin, M. *et al.,* Single-domain antibody fragments with high conformational stability, *Protein Science.* **11,** 500–515 (2002).

8.   Herrington, M. B. *et al.*Nonsense Mutations and Suppression Suppression In: eLS. John Wiley & Sons, Ltd: Chichester (2013).

9.   Chin, J. W. *et al.,* Progress toward an expanded eukaryotic genetic code, *Chem Biol.* **10,** 511–519 (2003).

10.   Young, T. S. *et al.*, Beyond the canonical 20 amino acids: expanding the genetic lexicon, *J. Biol. Chem.* **285,** 11039–44 (2010).

11.   Deiters, A. *et al.,* Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*, *J. Am. Chem. Soc.* **125,** 11782–3 (2003).

12. Cereghino, G. P. *et al.*, Applications of yeast in biotechnology:
     protein production and genetic analysis, *Curr. Opin. Biotechnol.*
     10, 422–7 (1999).

13. Demain, A. L. *et al.*, Production of recombinant proteins by
     microbes and higher organisms, *Biotechnol. Adv.* **27,** 297–306
     (2009).

14. Huang, C. *et al.*, Recombinant immunotherapeutics: Current state
     and perspectives regarding the feasibility and market, *Appl.
     Microbiol. Biotechnol.* **87,** 401–410 (2010).

15. Çelik, E. *et al.*, Production of recombinant proteins by yeast cells,
     Biotechnol. Adv. **30,** 1108–1118 (2012).

16. Eckart, M. R. *et al.*, Quality and authenticity of heterologous
     proteins synthesized in yeast, *Curr. Opin. Biotechnol.* **7,** 525–30
     (1996).

17. Yin, J. *et al.*, Select what you need: A comparative evaluation of
     the advantages and limitations of frequently used expression
     systems for foreign genes, *J. Biotechnol.* **127,** 335–347 (2007).

18. Vidal, M. *et al.*, Reverse two-hybrid and one-hybrid systems to
     detect dissociation of protein-protein and DNA-protein
     interactions, *Proc. Natl. Acad. Sci. U.S.A.* **93,** 10315–20 (1996).

19. Vidal, M. *et al.*, Genetic characterization of a mammalian protein-
     protein interaction domain by using a yeast reverse two-hybrid
     system, *Proc. Natl. Acad. Sci. U.S.A.* **93,** 10321–10326 (1996).

20. Gietz, R. D. *et al.*, Studies on the transformation of intact yeast
     cells by the LiAc/SS-DNA/PEG procedure, *Yeast* **11,** 355–60
     (1995).

21. Partow, S. *et al.,* Characterization of different promoters for
     designing a new expression vector in *Saccharomyces cerevisiae*,
     Yeast. **21,** 955–964 (2010).

22. Deiters, A. *et al.,* Adding amino acids with novel reactivity to the
     genetic code of *Saccharomyces cerevisiae*, *J. Am. Chem. Soc.*
     125, Supporting Information. 1–8 (2003).

23. Gesteland, R. F. *et al.,* Yeast suppressors of UAA and UAG
     nonsense codons work efficiently in vitro via tRNA, *Cell* **7,** 381–90
     (1976).

24. Ohno, R. *et al.*, Co-expression of yeast amber suppressor tRNA^{Tyr}
     and tyrosyl-tRNA synthetase in *Escherichia coli*: Possibility to
     expand the genetic code, *J. Biochem.* **1068,** 1065–1068 (1998).

25. Liu, D. R. *et al.*, Progress toward the evolution of an organism
     with an expanded genetic code, *Proc. Natl. Acad. Sci. U. S. A.* **96,**
     4780–5 (1999).

26. Chen, S. *et al.*, An improved system for the generation and
     analysis of mutant proteins containing unnatural amino acids in
     Saccharomyces cerevisiae, *J. Mol. Biol.* **371,** 112–22 (2007).

27. Wang, Q. *et al.*, New methods enabling efficient incorporation of
     unnatural amino acids in yeast, *J. Am. Chem. Soc.* **130,** 6066–7
     (2008).

28. Wang, Q. *et al.*, Expanding the genetic code for biological studies,
     Chem. Biol. **16,** 323–36 (2009).

29. Gorlani, A. *et al.*, Expression of VHHs in *Saccharomyces
     cerevisiae*, *Methods Mol Biol.* **911,** 277–286 (2012).

Chapter 7

General discussion and conclusions

The aim of this work was the evaluation and application of the amber suppression approach for the site-specific modification of the nanobody BCII10 (NbBCII10) for the covalent and oriented coupling to solid surfaces. The controlled immobilization of proteins is an important step towards the development and miniaturization of devices with a wide range of applications, e.g. biomedical implants and biosensing systems. All these applications demand high sensitivity and selectivity combined with high reproducibility, properties that can only be achieved by a covalent and oriented coupling of the protein to the surface. Even though the use of enzymes such as sortase A or methods such as EPL can result in an oriented and covalent coupling of the protein, the modification/immobilization using these techniques is limited to the N- or C- terminus of the protein.

The site-specific modification of the NbBCII10 presented here is based on the incorporation of the non-natural amino acid *p*-azidophenylalanine (*p*-azidoPhe) by means of amber suppression in the yeast *S. cerevisiae*. In this context, the amber suppressor *E. coli* tyrosine $tRNA_{CUA}$ ($tRNA_{CUA}^{Tyr}$) is used in combination with a mutant *E. coli* tyrosyl-tRNA synthetase (EcTyrRS). This mutant EcTyrRS has affinity for *p*-azidoPhe instead of the natural tyrosine. This strategy allows the introduction of a bioorthogonal functional group at any location in a protein, which is unique compared to other existing modification strategies, e.g. EPL.

It was opted to use the yeast *S. cerevisiae* for the production of the modified nanobodies. The advantage of yeast is that it combines the ease of microbial growth with an eukaryotic environment, allowing also the functionalization of more complex proteins that require post-translational modifications. *S. cerevisiae* has been recognized as a "generally regarded as safe" (GRAS) strain, making it a safe organism to produce recombinant proteins.

Nanobodies were used in this work as a protein model system, since they are small, easy to produce and very stable. These characteristics make them very suited for applications requiring a long shelf-life, e.g. biosensors.

The first objective was the construction of the DNA vectors necessary for this work. These vectors were needed for the production of azide-modified nanobodies as well as to perform the orthogonality assay/proof of principle experiment for diploid cells. The construction and composition of the vectors was discussed in detail in chapter 3. All the vectors needed for these objectives have been successfully constructed and have been verified by restriction analysis and DNA sequencing.

After the construction of the vectors, a mutant EcTyrRS with affinity for *p*-azidoPhe was required. For this purpose, a library of mutant EcTyrRSs should be constructed and screened for the appropriate mutant. Such a library is generated by the random mutation of key amino acids present in the binding pocket of the EcTyrRS. In all literature reports available until now, the screening of such a mutant library is performed by means of a double-sieve method in the haploid MaV203 yeast strain. This double-sieve method is based on the activation/repression of the GAL4 inducible URA3 gene present in the MaV203 cells (see section 4.1.1).

In general, the amber suppression approach is based on the use of MaV203 cells to screen a library of mutant aminoacyl-tRNA synthetases (aaRSs) and the use of yeast episomal plasmids to produce site-specifically modified proteins have been extensively described in literature. However, a number of inconsistencies were revealed for this approach in this PhD study.

The first inconsistency observed was that a fraction of MaV203 cells is capable of growing on SD-URA (single drop-out medium lacking uracil) in the absence of GAL4. This has never been reported before, since nothing can be found about it in literature. The results obtained in this work indicate that spontaneous recessive mutations in a single gene occur in the MaV203 cells after a number of cell divisions. These spontaneous mutations result in an unwanted activation of the URA3 gene (chapter 4). The unwanted URA3 gene activation can have a serious unfavorable impact on the screening of mutant aaRS libraries, since the screening is based on the activation/repression of the GAL4 inducible URA3 gene. In the negative screening step, all MaV203 cells

with a URA$^+$ phenotype will not survive. As a result, also mutant aaRSs with an affinity for the non-natural amino acid will not be selected if present in a MaV203 cell with a spontaneous expression of the URA3 gene. Consequently, the real library size will be smaller than theoretically calculated.

The above-described spontaneous mutation most likely occurs in a protein involved in the Rpd3 Histone Deacetylase Complex (Rpd3 HDAC) or in a kinase/phosphatase involved in the regulation of this complex. The plausibility of these findings was further confirmed in a personal communication with Prof. Dr. Vidal and Prof. Dr. Boeke, who developed the MaV203 strain in 1996.

Furthermore, in this work it was demonstrated that the mutation is recessive and that the unwanted URA3 activation can be avoided by using diploid MaV103/MaV203 cells, instead of haploid MaV203. The diploid cells were formed by mating of MaV103(a) and MaV203(α) cells and look promising for the screening of a library of mutant EcTyrRSs. Moreover, in a proof of principle experiment it was shown that the diploid cells can be used together with the yeast episomal plasmids pEcTyrRS/tRNA$_{CUA}$^{Tyr} and pADH1/GAL4_T44_R110 for the screening of libraries of mutant EcTyrRSs. This proof of principle experiment was performed using the same conditions as the positive selection step of the double-sieve screening method. The results of this experiment show clearly that the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair is able to suppress the amber codons present in the GAL4 gene in diploid cells. When either EcTyrRS or tRNA$_{CUA}$^{Tyr} are absent, no amber suppression takes place and a Ura$^-$ phenotype is maintained. It can therefore be concluded that diploid MaV103/MaV203 cells should be used instead of haploid MaV203 cells for the screening of libraries of mutant aminoacyl-tRNA synthetases.

After solving the problem of the unwanted URA3 expression in the haploid MaV203 strain, the construction of a library of mutant EcTyrRSs was the next objective (chapter 5). The EcTyrRS was the first aaRS to be modified and introduced in *S. cerevisiae* for the amber suppression approach and has been extensively described. All protein-engineering studies accomplished on EcTyrRS in the past were based on the crystal structure of the TyrRS of *Bacillus stearothermophilus* and not on the structure of EcTyrRS, resulting in mutant EcTyrRSs capable of incorporating both *p*-azidophenylalanine as well as *p*-propargyloxyphenylalanine. The reason was that the crystal structure of

EcTyrRS was not available at the time. However, nowadays the crystal structure of EcTyrRS is available, together with improved information regarding the amino acid recognition in its binding pocket. This results in more accurate information to predict the recognition of the non-natural amino acid in the binding pocket. This allows for better suggestions concerning its modification in order to recognize non-natural amino acids, resulting in more selective mutants. It was opted to construct and screen an own library of mutant EcTyrRSs, not only to acquire more selective EcTyrRS mutants, but also to apply it later on other aaRSs. The library is constructed by the random mutation of five key amino acids (Tyr37, Asp81, Tyr175, Gln179 and Asp182) present in the binding pocket based on the crystal structure of EcTyrRS. For this purpose, the GeneArt® site-directed mutagenesis PLUS kit was evaluated. According to the manufacturer, this kit allows the random mutation of these five amino acids in one round of mutagenesis. However, the results obtained in this work show that the GeneArt® site-directed mutagenesis PLUS kit is not able to introduce the random mutations in the binding pocket of EcTyrRS and therefore cannot be used for the construction of a mutant library. Since Enzymatic Inverse PCR (EIPCR) might be an alternative to construct the library of mutant EcTyrRSs, preliminary work has been started. Although the vectors necessary for this EIPCR method have been constructed successfully, more research is needed.

Generally spoken, the construction and screening process of a library of mutant EcTyrRSs is time-consuming and there is no guarantee of success. For this reason, protein expression experiments for the production of modified NbBCII10 were performed in parallel with the library construction. For these protein expression experiments, a mutant EcTyrRS of which the affinity for *p*-azidoPhe was described in literature, was evaluated (chapter 6).

However, when performing the protein expression experiments, a second inconsistency of the amber suppression approach was uncovered. The results of the expression experiments indicate that yeast episomal plasmids are not suited to create a stable protein modification system. The results obtained in this thesis first of all indicate that it is very difficult to select a clone capable of expressing all the necessary components needed for amber suppression, namely:

   i.   the EcTyrRS,
   ii.   the *E. coli* tRNA$_{CUA}^{Tyr}$ and,
  iii.   the target gene containing the amber codon.

It has been observed in this work, as it is described in literature, that heterologous protein expression in *S. cerevisiae* is clone-specific. Even though the cells contain the vector after transformation (survival on selective medium), the level of protein expression of different clones can vary strongly from high, over intermediate to absence of expression.

Therefore, only very few clones will express all three components in sufficient amounts to obtain detectable levels of modified proteins. If one of the components is not properly expressed, amber suppression will not occur. This explains why it was so difficult to find a clone capable of suppressing the amber codon in GFP.

Furthermore, it became clear that even if suited clones were found, the expression of the components needed for amber suppression is not stable. It was observed that the ability of MaV203 cells to express wildtype NbBCII10 slowly decreased over time, until NbBCII10 expression stopped. For the suppression of the amber codon in GFP, these findings were even more pronounced, *i.e.* replating the cells once resulted already in a complete loss of amber suppression.

A possible explanation might be that, in case of wildtype NbBCII10_His$_6$, the expression is performed using a single yeast episomal plasmid (pTEF-MF/NbBCII10_His$_6$). The propagation of the plasmid and the expression of the heterologous protein induces pressure on the cells. To cope with this pressure, the cells might slowly decrease the expression of the heterologous protein (e.g. NbBCII10 or GFP). In case of amber suppression, three components have to be expressed by means of two yeast episomal plasmids. The propagation of two plasmids and the expression of EcTyrRS, tRNA$_{CUA}^{Tyr}$ and NbBCII10_His$_6$ probably results in a much higher pressure on the cells. This probably results in a faster down-regulation of the expression of the components needed for successful amber suppression.

Although more investigation is needed to elucidate the exact plasmid loss mechanism, recombination of the plasmids or a drop in copy number of the plasmids might be an explanation (discussed in a personal communication with Dr. Aliona Bogdanova, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany).

It can be concluded that yeast episomal plasmids (YEps) are not suited to create a stable protein modification system and that alternatives should be investigated.

Summary

The site-specific modification of proteins is of great importance for the development and improvement of many biotechnological applications (e.g. biosensors). The focus of this work is the *in vivo* site-specific modification of proteins by means of amber suppression. As a result, a non-natural amino acid can be introduced to the protein, which can serve as a unique chemical "handle" to immobilize proteins in an oriented and covalent way, resulting in a bioactive surface.

The model protein used in this work is the nanobody BCII10. Nanobodies are very stable single-domain antibody fragments derived from camelid heavy-chain antibodies (HCAbs). They are relatively small proteins encoded by a single gene with an activity comparable to classical antibodies.

In this work the yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) is used for the production of the modified nanobody. Yeast combines the ease of microbial growth with an eukaryotic environment, enabling the possibility to introduce eukaryotic specific modifications, such as protein folding, disulfide bond formation, proteolytic processing and glycosylation. Besides, the translational machinery of eukaryotes is very well conserved, allowing genes involved in the site-specific incorporation of non-natural amino acids in yeast, to be duplicated and used in higher eukaryotes (e.g. mammalian cells) or other yeast species such as *Pichia pastoris, Kluyveromyces lactis, Schizoschaccharomyces pombe* and more.

Amber suppression allows the incorporation of bioorthogonal functionalities (e.g. azide) in the form of a modified (non-natural) amino acid in response to an amber codon. For this purpose, the translational machinery of *S. cerevisiae* is expanded with a genetically encoded *E. coli* suppressor tRNA$_{CUA}^{Tyr}$ and a mutant *E. coli* tyrosyl-tRNA synthethase (EcTyrRS) with an affinity for *p*-azidophenylalanine (*p*-azidoPhe) instead of tyrosine. This will result in the incorporation of *p*-azidoPhe in response of the amber codon. Such a mutant EcTyrRS is obtained by the construction and screening of a library of mutant EcTyrRSs. Both the benefits as well as the difficulties of constructing such a library is discussed in chapter 5. The main advantage of the amber suppression approach is that it allows the production of proteins that contain a

genetically encoded orthogonal functional group (*i.e.* azide) on a single, strategically chosen position in the protein. This then allows the oriented coupling of the 'click' functionalized nanobody to an alkynylated surface by means of the Huisgen 1,3-dipolar cycloaddition.

Nevertheless, during the development of this work, two inconsistencies were encountered: one for the library screening and one for the protein expression. The problems will be briefly explained as follows.

First, in order to obtain a mutant EcTyrRS, a library needs to be generated by a random mutation of certain amino acids present in the binding pocket of EcTyrRS. This library is then screened by means of a double-sieve method in the haploid yeast strain MaV203 to obtain the mutant EcTyrRS selective for the modified amino acid. The double-sieve method is based on the activation and repression of the URA3 gene in MaV203. The *SPAL10::URA3* fusion in MaV203 cells was developed by Vidal *et al.* to actively repress URA3 gene expression unless full-length GAL4 is present.

However, this work showed that spontaneous, recessive mutations occur in a single repressor gene of the SPAL10 promoter in MaV203 cells, resulting in a URA3$^+$ phenotype *without* the presence of GAL4. This unwanted URA3 expression has a negative impact on the library screening. In this work evidence is provided that diploid cells can be used for the library screening to circumvent the unwanted activation of the URA3 gene (chapter 4).

A second inconsistency was encountered during the protein expression experiments. Literature reports that yeast episomal plasmids (YEp) can be used to express both the EcTyrRS/tRNA$_{CUA}$^{Tyr} pair as well as the protein of interest, containing the amber codon. However, the results discussed in chapter 6 of this work suggest that yeast episomal plasmids are not suited for the production of modified proteins and that more stable expression systems (e.g. yeast integrating plasmids) need to be investigated.

Future perspectives

Amber suppression remains an interesting and promising technique for the site-specific modification of proteins, as it allows more or less a free and rational selection of the location of modification. However, it is clear that more research is needed in order to produce modified heterologous proteins in sufficient levels to become interesting for commercial applications.

In the first place, it is important to find more stable alternatives to efficiently express the necessary components of the amber suppression system. A possible alternative is the use of inducible promoters to express the components in combination with so-called Yeast Integrating plasmids (YIp). These YIps integrate stably in the chromosomal DNA of yeast and do not need an auxotrophic marker to propagate. Once integrated, these YIps remain part of the chromosomal DNA of yeast. Therefore, integrating multiple copies of an expression cassette (containing all the components for amber suppression) into the yeast's chromosome might result in a more stable expression and higher yield of modified proteins. Another possibility could be to combine the use of YEps and YIps e.g. by expressing the EcTyrRS/tRNA$_{CUA}^{Tyr}$ pair using a yeast integrating plasmid (YIp) while expressing the nanobody using a yeast episomal plasmid (Yep). These alternative expression systems can be investigated using the same mutant EcTyrRS as was used in the amber suppression experiments discussed in chapter 6 of this thesis.

Once a stable expression/modification system is generated, new and more selective mutant synthetases can be developed. For this, bioinformatical analyses of the binding pocket can be performed to identify the specific amino acids to be modified in order to recognize non-natural amino acids. Once these amino acids are identified, a library can be constructed and screened as described in section 4.1.1. Yeast episomal plasmids can be used for the construction and screening of a library of mutant aaRSs, as was demonstrated by the proof of principle assay for the use of diploid cells in section 4.2.6. Afterwards, the obtained mutant aaRS can be transferred to a YIp and be used for the production of site-specifically modified proteins.

Nederlandstalige samenvatting

De plaats-specifieke modificatie van proteïnen speelt een belangrijke rol in de ontwikkeling en verbetering van vele biotechnologische toepassingen (b.v. biosensoren). In dit werk wordt de focus gelegd op de plaats-specifieke modificatie van proteïnen met behulp van *in vivo* amber suppressie. Hiermee kan een niet-natuurlijk aminozuur ingebouwd worden waarmee het proteïne georiënteerd en covalent aan een vaste drager gekoppeld kan worden voor de productie van een bioactief oppervlak.

Het modelproteïne dat hiervoor gebruikt zal worden is de nanobody BCII10, gericht tegen β-lactamase. Nanobodies zijn zeer stabiele antilichaamfragmenten afgeleid van zogenaamde zware-keten antilichamen ("heavy-chain antibodies" of HCAbs) van kameelachtigen. Het zijn relatief kleine proteïnen, gecodeerd door één enkel gen, met een activiteit gelijkaardig aan die van klassieke antilichamen.

In dit werk wordt de gist *Saccharomyces cerevisiae* (*S. cerevisiae*) gebruikt voor de productie van het gemodificeerde NbBCII10 nanobody. Gist combineert een microbiële groei met een eukaryote omgeving die post-translationele modificaties zoals eiwitvouwing, vorming van disulfidebruggen, proteolytische aanpassingen en glycosylering toelaat. Een bijkomend voordeel is dat de translationele machinerie van eukaryoten sterk geconserveerd is. Hierdoor kunnen genen die betrokken zijn bij de plaats-specifieke incorporatie van niet-natuurlijke aminozuren in gist, ook gebruikt worden in hogere eukaryoten zoals bijvoorbeeld zoogdiercellen. Daarnaast kunnen ze ook gebruikt worden in andere giststammen zoals *Pichia pastoris, Kluyveromyces lactis, Schizoschaccharomyces pombe* en dergelijke.

Met behulp van ambersuppressie kunnen bio-orthogonale functionele groepen (b.v. een azide) in de vorm van een niet-natuurlijk aminozuur ingebouwd worden als respons op een ambercodon. Hiervoor dient de translationele machinerie van *S. cerevisiae* uitgebreid te worden met een genetisch gecodeerde *E. coli* suppressor tRNA$_{CUA}^{Tyr}$ en een mutant *E. coli* tyrosyl-tRNA synthetase (EcTyrRS) met een affiniteit voor *p*-azidofenylalanine (*p*-azidoPhe) in plaats van tyrosine. Dit leidt tot de incorporatie van *p*-azidoPhe als reactie op een amber codon. Dit mutant EcTyrRS wordt verkregen door de constructie en het screenen van een

bibliotheek van mutante EcTyrRSs. Zowel de voordelen als de moeilijkheden van het construeren van een bibliotheek van mutante EcTyrRS wordt besproken in hoofdstuk 5. Het voordeel van deze strategie is dat het mogelijk is om nanobodies te produceren die een genetisch gecodeerde azidegroep bezitten op een strategisch gekozen locatie. Dit maakt de georiënteerde en covalente koppeling mogelijk van het 'click' gemodificeerde nanobody op een gealkynyleerd oppervlak met behulp van de Huisgen 1,3-dipolaire cycloadditie.

Hoewel de methodologie uitgebreid beschreven is in de literatuur, werden in dit werk tekortkomingen aan het licht gebracht met betrekking tot het screenen van de bibliotheek van EcTyrRS-mutanten en de expressie van gemodificeerde nanobodies.

Om een geschikte EcTyrRS mutant te vinden, dient een bibliotheek geconstrueerd te worden door de willekeurige mutatie van een aantal essentiële aminozuren aanwezig in de bindingssite van EcTyrRS. Deze bibliotheek wordt vervolgens gescreend in de haploïde giststam MaV203. Dit screenen is gebaseerd op de activatie en repressie van het URA3 gen aanwezig in MaV203. De *SPAL10::URA3* fusie in MaV203 cellen werd ontwikkeld door Vidal *et al.* om de expressie van het URA3-gen actief te onderdrukken, tenzij GAL4 aanwezig is. In dit werk werd echter aangetoond dat een spontane, recessieve mutatie optreedt in één enkele repressor van de SPAL10-promoter. Deze mutatie leidt tot een URA3$^+$-fenotype, zelfs indien GAL4 *niet* aanwezig is. Dit heeft een negatieve invloed op het screenen van de bibliotheek. In dit werk werd bovendien bewezen dat diploïde cellen gebruikt kunnen worden om de bibliotheek van mutant EcTyrRS te screenen en dat hiermee de spontane URA3-expressie vermeden kan worden (hoofdstuk 4).

Een tweede probleem werd ontdekt tijdens de experimenten voor de expressie van gemodificeerd NbBCII10. In de literatuur staat beschreven dat "yeast episomal plasmids" (YEp) gebruikt kunnen worden om zowel het EcTyrRS/tRNA$_{CUA}^{Tyr}$-paar als het nanobody (met het ambercodon) tot expressie te brengen. De resultaten die besproken worden in hoofdstuk 6 van dit werk tonen echter aan dat een eiwitmodificatiesysteem gebaseerd op YEps niet stabiel genoeg is voor de productie van gemodificeerde nanobodies. Verder onderzoek naar stabielere expressiesystemen zoals "yeast integrated plasmids" (YIp) is zeker nodig.

Publications:

- E. Steen Redeker, T. T. Duy, <u>D. Cortens</u>, B. Billen, W. Guedens and P. Adriaensens. "Protein Engineering For Directed Immobilization." Bioconjugate Chemistry, 24 (11), p. 1761-1777, 2013.

- K.L Jiménez-Monroy, N. Renaud, J. Drijkoningen, K. Schouteden, <u>D. Cortens</u>, G. Degutis, W. Guedens, C. van Haesendonck, J. Manca, L. D. A. Siebbeles, F. C. Grozema and P. Wagner. "High Electronic Conductance Through Double Helix DNA Molecules with Fullerene Anchoring Groups", article submitted in ACS Nano (IF:12.881).

- <u>D. Cortens</u>, P. Adriaensens, W. Guedens and E. Steen Redeker, "Spontaneous derepression of the URA3 gene in the yeast MaV203 strain", article submitted for publication in PLOS ONE (IF: 3.234).

Poster contibutions:

- <u>D. Cortens</u>, E. Steen Redeker, P. Adriaensens and W. Guedens. "*In vivo* synthesis of click functionalized nanobodies for advanced biosensing platforms". Biomedica 2015 The European Life Sciences Summit, Genk (2015).

- <u>D. Cortens</u>, E. Steen Redeker, P. Adriaensens and W. Guedens. "Clickable nanobodies for the development of improved bio-active surfaces". BPG (Belgian Polymer Group) Annual Meeting, Ghent, Belgium (2014).

- K.L. Jiménez-Monroy, J. Drijkoningen, J.P. Noben, <u>D. Cortens</u>, G. Degutis, P. Losada-Pérez, A.Y. Atalay, J. Manca, H.-G. Boyen, P. Wagner. "Electrical characterization of molecular wires based on DNA". Engineering of Functional Interfaces (EnFI), Aachen, Germany (2014).

- K.L. Jiménez-Monroy, J. Drijkoningen, <u>D. Cortens</u>, P. Losada-Pérez, A.Y. Atalay, J. Manca, P. Wagner. "DNA electrical behaviour at room temperature: I-V measurements on single molecules". Belgian Physical Society, Leuven, Belgium (2014).

- <u>D. Cortens</u>, P. Adriaensens, W. Guedens and E. Steen Redeker. "The development of click functionalized nanobodies by nonsense suppression in *S.cerevisiae* for oriented coupling to surfaces". Engineering of Functional Interfaces (EnFI), Hasselt, Belgium (2013).

- E. Steen Redeker, G. Reekmans, <u>D. Cortens</u>, W. Guedens and P. Adriaensens. "Designing clickable proteins". Belgian-German (Macro)Molecular Meeting, Advanced Materials by Modular Strategies: From Synthesis to Industrial Applications, Houffalize, Belgium (2012).

- <u>D. Cortens</u>, E. Steen Redeker, P. Adriaensens and W. Guedens. "*In vivo* functionalization of nanobodies toward 'Click'-chemistry coupling to surfaces". BPG (Belgian Polymer Group): Polymers for a Sustainable Society, Blankenberge, Belgium (2012).

- <u>D. Cortens</u>, E. Steen Redeker, P. Adriaensens and W. Guedens. "Development of an *in vivo* system for the expression of clickable nanobodies". 2nd NanoSensEU Symposium on Biosensor Development "Trends and Technology", Hasselt, Belgium (2012).

- <u>D. Cortens</u>, E. Steen Redeker, P. Adriaensens and W. Guedens. "*In vivo* synthesis of clickable nanobodies". KVCV ChemCYS, Blankenberge, Belgium (2012).

Appendix

List of bacterial and yeast strains

E.coli	Genotype
BL21(DE3)pLysS	F⁻ *ompT hsd*S$_B$ (r$_B$⁻ m$_B$⁻) *gal dcm* (DE3) pLysS (Cam^R)
Top10F′	F′ [lacI^q Tn10 (Tet^R)] mcrA Δ(*mrr⁻hsd*RMS-mcrBC) Φ80*lacZ*ΔM15 Δ*lac*X74 recA1 araD139 Δ(*ara-leu*)7697 *gal*U *gal*K rpsL (Str^R) *end*A1 *nup*G
DH5α-T1^R	F⁻ Φ80*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *recA1 endA1 hsdR17*(r$_k$⁻, m$_k$⁺) *phoA supE44 thi*-1 *gyrA96 relA1 tonA* (confers resistance to phage T1)

S. cerevisiae	
MaV103	*MATα; leu2-3,112; trp1-901; his3Δ200; ade2-101; cyh2R; can1R; gal4Δ; gal80Δ; GAL1::lacZ; HIS3UASGAL1::HIS3@LYS2; SPAL10::URA3*
MaV203	*MATα; leu2-3,112; trp1-901; his3Δ200; ade2-101; cyh2R; can1R; gal4Δ; gal80Δ; GAL1::lacZ; HIS3UASGAL1::HIS3@LYS2; SPAL10::URA3*
INVSc1	*MATa his3D1 leu2 trp1-289 ura3-52 MA*T *his3D1 leu2 trp1-289 ura3-52*

List of kits

TA Cloning Kit	Invitrogen, Carlsbad, USA
QIAprep Spin Miniprep Kit	Qiagen, Venlo, the Netherlands
Qiagen Gel Extraction kit	
QuickChange Lightning Site-Directed Mutagenesis kit	Stratagene, Agilent Technologies, Brussels, Belgium
The Quickchange Multi Site-Directed Mutagenesis Kit	
GeneArt® Site-Directed Mutagenesis PLUS Kit	Life Technologies, Carlsbad, USA
GeneArt® Site-Directed Mutagenesis Kit	

A. Tables

Table A1: overview of the primers used for the amplification of the genes for the development of the DNA constructs necessary for the amber suppression approach.

	Gene		Primer sequence	Restriction-site
pEcTyrRS / tRNA_CUA	tRNA_CUA	Forward	GGGGGG**ACCGGT**AAGCTTCCC GATAAGGGAGGCAGGCCAGTAA AAGCATTACCCCGTGGTGGGGT TCCCGA	AgeI
		Reverse	GGCGGC**GCTAGC**GGGAAGTT CAGGGACTTTTGAAAAAATG GTGGTGGGGAAGGAT	NheI
	EcTyrRS	Forward	TCATAACGA**GAATTC**ATGGCA AGCAGTAACTTG	EcoRI
		Reverse	TTACTACGT**GCGGCCGC**TTATT TCCAGCAAATCAGAC	NotI
	ADH1-promotor	Forward	GGGGGG**ACCGGT**CGGGATCG AAGAAATGATGGTAAATGAAA TAGGAAATCAAGG	AgeI
		Reverse	GGGGGG**GGAATTC**AGTTGATT GTATGCTTGGTATAGCTTGAA ATATTGTGCAGAAAAAGAAAC	EcoRI
pADH1 / Gal4	ADH1-promotor	Forward	CGAATT**GCTAGC**CGGGATCGAA GAAATGATGGTAAATGAAATA GGAAATC	NheI
		Reverse	GGGGGG**GGGATCC**AGTTGATTG TATGCTTGGTATAGCTTGAAAT ATTGTGCAGAAAAAGAAAC	BamHI
	GAL4	Forward	GGACGG**GGATCC**ATGAAGCTA CTGTCTTCTATCGAACAAGCAT	BamHI
		Reverse	ATCACGGCTT**AGATCT**TTACTCT TTTTTGGGTTTGGTGGGTAT CTTC	BglII

Continue from **Table A1**: overview of the primers used for the amplification of the genes necessary for the development of the DNA constructs.

	Gene		Primer sequence	Restriction-site
pTEF-MF/ NbBCII10_His₆	NbBCII10	Forward	GGC**GAATTC**ATGGCCCAGGTGC	EcoRI
		Reverse	GGG**CTCGAG**TTAGTGATGGTGATGGTG	XhoI
pTEF/ GFP_Y39TAG	GFP_Y39TAG	Forward	GGGCGGC**ACTAGT**ATGGATTACAAAGATG	SpeI
		Reverse	GGCGGC**CTCGAG**TCAGTGATGGTGAT	XhoI
pADH1/ GFP(_Y39TAG)	ADH1	Forward	GAATT**GCTAGC**CGGGATGCGAAGAAATGATGGTAAATGAAATAGGAAATC	NheI
		Reverse	GGGGGGG**GGATCC**AGTTGATTGTATGCTTGGTATAGCTTGAAATATTGTGCAGAAAAAGAAAC	BamHI
	GFP	Forward	CCGAGGCG**GATCC**ATGGATTACAAAGATG	BamHI
		Reverse	GCGGCT**TAGATCT**TCAGTGATGGTGATGGTGATGCTTGT	BglII

Table A2. Reaction mixtures used for the PCR amplification of the genes needed to develop the DNA constructs. For EcTyrRS and tRNA^{Tyr} four different MgCl$_2$ concentrations were tested. The optimal MgCl$_2$ concentration was 2mM for EcTyrRS and 4mM for the tRNA^{Tyr}. For the other genes to be amplified the PCR reaction was performed with 2mM MgCl$_2$.

	2mM MgCl2 (µl)	4mM MgCl2 (µl)	6mM MgCl2 (µl)	8mM MgCl2 (µl)
PCR buffer 10x	5	5	5	5
dNTP's (2,5mM)	4	4	4	4
Forward primer (10µM)	1	1	1	1
Reverse primer (10µM)	1	1	1	1
MgCl2 (50µM, Thermoscientific)	2	4	6	8
Pfu Polymerase (Thermoscientific)	0.25	0.25	0.25	0.25
DNA (20ng/µl)	5	5	5	5
water	31.75	29.75	27.75	25.75

Table A3. Program settings for the PCR of EcTyrRS for the construction of the pEcTyrRS/tRNA$_{CUA}^{Tyr}$ vector.

Temperature	Time	# Cycles
95°C	5 min	1
95 °C	30 sec	
45-70 °C	30 sec	30
72 °C	1 min	
72 °C	5 min	1

Table A4. Program settings for the PCR of *E. coli* tRNA^{Tyr} for the construction of the pEcTyrRS/tRNA$_{CUA}^{Tyr}$ vector.

Temperature	Time	# Cycles
95 °C	5 min	1
95 °C	30 sec	
45-70 °C	30 sec	30
72 °C	1 min 30 sec	
72 °C	5 min	1

Table A5. The forward and reverse primer used for the site-directed mutagenesis of the tRNA^{Tyr} anticodon.

Forward primer	CCAAAGGGAGCAGACTCTAAATCTGCCGTCATC
Reverse primer	GATGACGGCAGATTTAGAGTCTGCTCCCTTTGG

Table A6. Program settings for the PCR of the yeast ADH1 promotor for the construction of the pEcTyrRS/tRNA$_{CUA}$^{Tyr} vector.

Temperature	Time	# Cycles
95 °C	5 min	1
95 °C	30 sec	
60 °C	30 sec	30
72 °C	4 min	
72 °C	5 min	1

Table A7. The forward and reverse primers used for the site-directed mutagenesis of the EcTyrRS gene.

Location of mutation		Primer sequence
Y37T	Forward	GATCGAAGCCGCAA**GT**GAGCGCGATCGGGC
	Reverse	GCCCGATCGCGCTC**AC**TTGCGGCTTCGATC
D182S_ **F183A**	Forward	CTGTTTGTTCAGACAGGCG**GC**G**GA**ATAACCCTGCAACAGGTTGTAGGAAAAC
	Reverse	GTTTTCCTACAACCTGTTGCAGGGTTAT**TC**C**GC**CGCCTGTCTGAACAAACAG

Table A8. Program settings for the PCR of NbBCII10_His$_6$ for the construction of the pTEF-MF/ NbBCII10_His$_6$ vector.

Temperature	Time	# Cycles
95 °C	5 min	1
95 °C	30 sec	
50 °C	30 sec	30
72 °C	1min 30 sec	
72 °C	5 min	1

Table A9. The forward and reverse primers used for the site-directed mutagenesis of the NbBCII10_His$_6$ gene, introducing the amber codon (TAG).

	Primer sequence
Forward	ACCGTCTCCTCACGC**TAG**CGCCACCACCATCAC
Reverse	GTGATGGTGGTGGCG**CTA**GCGTGAGGAGACGGT

Table A10. Program settings for the PCR amplification of GFP_Y39TAG for the construction of the pTEF/GFP_Y39TAG vector

Temperature	Time	# Cycles
95 °C	5 min	1
95 °C	30 sec	
55 °C	30 sec	30
72 °C	2 min 30 sec	
72 °C	5 min	1

Table A11. Program settings for the PCR of the ADH1 promotor for the construction of the pADH1/GAL4 vector

Temperature	Time	# Cycles
95 °C	5 min	1
95 °C	30 sec	
59 °C	30 sec	20
72 °C	6 min	
72 °C	5 min	1

Table A12. The forward and reverse primer used for the site-directed mutagenesis of GAL4, introducing the amber codon at position 44 and 110.

Location of mutation	Primer sequence	
T44TAG	Forward	AGTGTCGCTACTCTCCCAAA**TAG**AAAAGGTCTCCGCTGACTAG
	Reverse	CTAGTCAGCGGAGACCTTTT**CTA**TTTGGGAGAGTAGCGACACT
R110TAG	Forward	AAGATAATGTGAATAAAGATGCCGTCACAGAT**TAG**TTGGCTTCAGTGGAGAC
	Reverse	GTCTCCACTGAAGCCAA**CTA**ATCTGTGACGGCATCTTTATTCACATTATCTT

Table A13. Composition of the master mix used for the multi-site mutagenesis and of the reaction mix used for the methylation control.

	Mutagenesis (3reactions)	Methylation control (1 reaction)
10x Accuprime™ *Pfx* reaction buffer	6 µl	1x
10x Enhancer	6 µl	1x
Plasmid pEcTyrRS/tRNA$_{CUA}^{Tyr}$ (20ng/µl)	1.2 µl	24 ng
DNA Methylase (4U/µl)	1.2 µl	4.8 units
25x SAM (S-adenosine methionine)	2.4 µl	-
Accuprime™ *Pfx* (2.5U/µl)	0.5 µl	1.5 units
water	to 57 µl	to 57 µl

Table A14. PCR parameters for the methylation and mutagenesis reaction for the construction of the library of mutant EcTyrRS.

Temperature	Time	# Cycles
37 °C	17'	1
95 °C	2'	
95 °C	20"	18
57 °C	30"	
68 °C	Primer mix 1: 9' Primer mixtures 2 & 3: 1'	
68 °C	5'	1

Table A15. Recombination reaction mixture for the construction of the library of mutant EcTyrRS.

Component	Volume	Final Concentration
water	4 µl	-
PCR product from tube 1	2 µl	-
PCR product from tube 2	2 µl	-
PCR product from tube 3	2 µl	-
GeneArt® 2x Enzyme mix	10 µl	1x

Table A16. PCR parameters for the control of the mutagenesis primers used for the construction of the library of mutant EcTyrRS.

Temperature	Time	# Cycles
95 °C	5'	1
94 °C	20"	
50 °C	30"	30
72 °C	3'	
72 °C	5'	1

Table A17. Components of the recombination reaction mixture (20 µl) used in the adapted protocol of the GeneArt® Site-Directed Mutagenesis Kit.

Component	Volume
water	8 µl
PCR product with primer mix 1	2 µl
PCR product with primer mix 2	2 µl
PCR product with primer mix 3	2 µl
5x reaction buffer	4 µl
10x Enzyme mix	2 µl

Table A18. The forward and reverse primers used for the introduction of the PstI restriction site in EcTyrRS.

Forward primer	CGGTGTGGTGCTGCAGATTGGTGGTTCTGAC
Reverse primer	GTCAGAACCACCAATCTGCAGCACCACACCG

B. Sequence alignments

S1. Alignment of the *E. coli* BL21 (DE3) pLys TyrRS sequence with the sequence of the amplified EcTyrRS gene. It can be seen that clone 3 contains two point mutations, whereas the sequence of clone 4 is identical to the reference.

```
                          20              40              60              80
EcTyrRS BL21(DE3)  ATGGCAAGCA GTAACTTGAT TAAACAATTG CAAGAGCGGG GGCTGGTAGC CCAGGTGACG GACGAGGAAG CGTTAGCAGA  80
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... ..........  80
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... ..........  80
                          100             120             140             160
EcTyrRS BL21(DE3)  GCGACTGGCG CAAGGCCCGA TCGCGCTCTA TTGCGGCTTC GATCCTACCG CTGACAGCTT GCATTTGGGG CATCTTGTTC 160
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... .......... 160
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 160
                          180             200             220             240
EcTyrRS BL21(DE3)  CATTGTTATG CCTGAAACGC TTCCAGCAGG CGGGCCACAA GCCGGTTGCG CTGGTAGGCG GCGCGACGGG TCTGATTGGC 240
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... .......... 240
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 240
                          260             280             300             320
EcTyrRS BL21(DE3)  GACCCGAGCT TCAAAGCTGC CGAGCGTAAG CTGAACACCG AAGAAACTGT TCAGGAGTGG GTGGACAAAA TCCGTAAGCA 320
EcTyrRS clone 3    A......... .......... .......... .......... .......... .......... .......... .......... 320
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 320
                          340             360             380             400
EcTyrRS BL21(DE3)  GGTTGCCCCG TTCCTCGATT TCGACTGTGG AGAAAACTCT GCTATCGCGG CGAACAACTA TGACTGGTTC GGCAATATGA 400
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... .......... 400
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 400
                          420             440             460             480
EcTyrRS BL21(DE3)  ATGTGCTGAC CTTCCTGCGC GATATTGGCA AACACTTCTC CGTTAACCAG ATGATCAACA AAGAAGCGGT TAAGCAGCGT 480
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... .......... 480
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 480
                          500             520             540             560
EcTyrRS BL21(DE3)  CTCAACCGTG AAGATCAGGG GATTTCGTTC ACTGAGTTTT CCTACAACCT GTTGCAGGGT TATGACTTCG CCTGTCTGAA 560
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... .......... 560
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 560
                          580             600             620             640
EcTyrRS BL21(DE3)  CAAACAGTAC GGTGTGGTGC TGCAAATTGG TGGTTCTGAC CAGTGGGGTA ACATCACTTC TGGTATCGAC CTGACCCGTC 640
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... .......... 640
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 640
                          660             680             700             720
EcTyrRS BL21(DE3)  GTCTGCATCA GAATCAGGTG TTTGGCCTGA CCGTTCCGCT GATCACTAAA GCAGATGGCA CCAAATTTGG TAAAACTGAA 720
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... .......... 720
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 720
                          740             760             780             800
EcTyrRS BL21(DE3)  GGCGGCGCAG TCTGGTTGGA TCCGAAGAAA ACCAGCCCGT ACAAATTCTA CCAGTTCTGG ATCAACACTG CGGATGCCGA 800
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... .......... 800
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 800
                          820             840             860             880
EcTyrRS BL21(DE3)  CGTTTACCGC TTCCTGAAGT TCTTCACCTT TATGAGCATT GAAGAGATCA ACGCCCTGGA AGAAGAAGAT AAAAACAGCG 880
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... ...G. ..... 880
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 880
                          900             920             940             960
EcTyrRS BL21(DE3)  GTAAAGCACC GCGCGCCCAG TATGTACTGG CGGAGCAGGT GACTCGTCTG GTTCACGGTG AAGAAGGTTT ACAGGCGGCA 960
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... .......... 960
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 960
                          980             1000            1020            1040
EcTyrRS BL21(DE3)  AAACGTATTA CCGAATGCCT GTTCAGCGGT TCTTTGAGTG CGCTGAGTGA AGCGGACTTC GAACAGCTGG CGCAGGACGG 1040
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... .......... 1040
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 1040
                          1060            1080            1100            1120
EcTyrRS BL21(DE3)  CGTACCGATG GTTGAGATGG AAAAGGGCGC AGACCTGATG CAGGCACTGG TCGATTCTGA ACTGCAACCT TCCGGTGGTC 1120
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... .......... 1120
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 1120
                          1140            1160            1180            1200
EcTyrRS BL21(DE3)  AGGCACGTAA AACTATCGCC TCCAATGCCA TCACCATTAA CGGTGAAAAA CAGTCCGATC CTGAATACTT CTTTAAAGAA 1200
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... .......... 1200
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... .......... 1200
                          1220            1240            1260
EcTyrRS BL21(DE3)  GAAGAGCGTC TGTTTGGTCG TTTTACCTTA CTGCGTCGCG GTAAAAAGAA TTACTGTCTG ATTTGCTGGA AATAA 1275
EcTyrRS clone 3    .......... .......... .......... .......... .......... .......... .......... ..... 1275
EcTyrRS clone 4    .......... .......... .......... .......... .......... .......... .......... ..... 1275
```

S2. Alignment of the sequence of the amber suppressor tRNA$_{CUA}^{Tyr}$ with the sequence of the tRNA^{Tyr} gene amplified from *E. coli* BL21 (DE3) pLys (without the restriction sites introduced by the primers). There is a difference in one nucleotide in the amplified tRNA gene, matching with the anticodon.

```
                         20                        40                        60
                          |                         |                         |
    tRNACUA  AAGCTTCCCG ATAAGGGAGC AGGCCAGTAA AAGCATTACC CCGTGGTGGG GTTCCCGAGC 60

tRNA clone 5   .......... .......... .......... .......... .......... .......... 60

tRNA clone 9   .......... .......... .......... .......... .......... .......... 60

tRNA clone 10  .......... .......... .......... .......... .......... .......... 60
                         80                        100                       120
                          |                         |                         |
    tRNACUA  GGCCAAAGGG AGCAGACTCT AAATCTGCCG TCATCGACTT CGAAGGTTCG AATCCTTCCC 120

tRNA clone 5   .......... ........G. .......... .......... .......... .......... 120

tRNA clone 9   .......... ........G. .......... .......... .......... .......... 120

tRNA clone 10  .......... ........G. .......... .......... .......... .......... 120
                         140
                          |
    tRNACUA  CCACCACCAT TTTTTTCAAA AGTCCCTGAA CTTCCC 156

tRNA clone 5   .......... .......... .......... ...... 156

tRNA clone 9   .......... .......... .......... ...... 156

tRNA clone 10  .......... .......... .......... ...... 156
```

S3. Alignment of the DNA sequence of the amplified tRNA^{Tyr} gene with the amber suppressor tRNA$_{CUA}^{Tyr}$ after site-directed mutagenesis proving that the amplified tRNA^{Tyr} was successfully altered to the amber suppressor tRNA$_{CUA}^{Tyr}$.

```
                         20                        40                        60
                          |                         |                         |
    tRNACUA  AAGCTTCCCG ATAAGGGAGC AGGCCAGTAA AAGCATTACC CCGTGGTGGG GTTCCCGAGC 60

pCR2.1_tRNACUA .......... .......... .......... .......... .......... .......... 60
                         80                        100                       120
                          |                         |                         |
    tRNACUA  GGCCAAAGGG AGCAGACTCT AAATCTGCCG TCATCGACTT CGAAGGTTCG AATCCTTCCC 120

pCR2.1_tRNACUA .......... .......... .......... .......... .......... .......... 120
                         140
                          |
    tRNACUA  CCACCACCAT TTTTTTCAAA AGTCCCTGAA CTTCCC 156

pCR2.1_tRNACUA .......... .......... .......... ...... 156
```

S4. Alignment of the sequence of the amplified ADH1 promoter with the ADH1 promoter from the vector pMA424. Clone 3 and 5 are identical to the reference sequence, whereas clone 4 contains several mutations.

S5. Alignment of the protein sequence of wildtype EcTyrRS and the mutant EcTyrRS obtained after site-directed mutagenesis. As can be seen, the protein substitution Y37T, D182S and F183A were successful.

```
                                                          20                          40
                                                          |                           |
EcTyrRS BL21(DE3) translation frame +1   MASSNLIKQL  QERGLVAQVT  DEEALAERLA  QGPIALYCGF  40

    EcTyrRS_mutant translation frame +1  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . T . . .  40
                                                          60                          80
                                                          |                           |
EcTyrRS BL21(DE3) translation frame +1   DPTADSLHLG  HLVPLLCLKR  FQQAGHKPVA  LVGGATGLIG  80

    EcTyrRS_mutant translation frame +1  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  80
                                                          100                         120
                                                          |                           |
EcTyrRS BL21(DE3) translation frame +1   DPSFKAAERK  LNTEETVQEW  VDKIRKQVAP  FLDFDCGENS  120

    EcTyrRS_mutant translation frame +1  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  120
                                                          140                         160
                                                          |                           |
EcTyrRS BL21(DE3) translation frame +1   AIAANNYDWF  GNMNVLTFLR  DIGKHFSVNQ  MINKEAVKQR  160

    EcTyrRS_mutant translation frame +1  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  160
                                                          180                         200
                                                          |                           |
EcTyrRS BL21(DE3) translation frame +1   LNREDQGISF  TEFSYNLLQG  YDFACLNKQY  GVVLQIGGSD  200

    EcTyrRS_mutant translation frame +1  . . . . . . . . . .  . . . . . . . . . .  . SA . . . . . . .  . . . . . . . . . .  200
                                                          220                         240
                                                          |                           |
EcTyrRS BL21(DE3) translation frame +1   QWGNITSGID  LTRRLHQNQV  FGLTVPLITK  ADGTKFGKTE  240

    EcTyrRS_mutant translation frame +1  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  240
                                                          260                         280
                                                          |                           |
EcTyrRS BL21(DE3) translation frame +1   GGAVWLDPKK  TSPYKFYQFW  INTADADVYR  FLKFFTFMSI  280

    EcTyrRS_mutant translation frame +1  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  280
                                                          300                         320
                                                          |                           |
EcTyrRS BL21(DE3) translation frame +1   EEINALEEED  KNSGKAPRAQ  YVLAEQVTRL  VHGEEGLQAA  320

    EcTyrRS_mutant translation frame +1  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  320
                                                          340                         360
                                                          |                           |
EcTyrRS BL21(DE3) translation frame +1   KRITECLFSG  SLSALSEADF  EQLAQDGVPM  VEMEKGADLM  360

    EcTyrRS_mutant translation frame +1  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  360
                                                          380                         400
                                                          |                           |
EcTyrRS BL21(DE3) translation frame +1   QALVDSELQP  SRGQARKTIA  SNAITINGEK  QSDPEYFFKE  400

    EcTyrRS_mutant translation frame +1  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  . . . . . . . . . .  400
                                                          420
                                                          |
EcTyrRS BL21(DE3) translation frame +1   EERLFGRFTL  LRRGKKNYCL  ICWK*  425

    EcTyrRS_mutant translation frame +1  . . . . . . . . . .  . . . . . . . . . .  . . . . .  425
```

S6. The sequence and composition of the SNR52_tRNA$_{CUA}$^{Tyr}_SUP4t construct.

S7. Alignment of the sequence of the amplified NbBCII10_His$_6$ gene with the NbBCII10_His$_6$ gene of the pHEN6-BCII10 vector.

```
                                    20                      40                      60
                                    |                       |                       |
        NbBCII10_His6  ATGGCCCAGG TGCAGCTGGT GGAGTCTGGG GGAGGCTCGG TGCAGGCTGG AGGGTCCCTG 60

NbBCII10_His6 (clone 1)  .......... .......... .......... .......... .......... .......... 60

NbBCII10_His6 (clone 2)  .......... .......... .......... .......... .......... .......... 60
                                    80                      100                     120
                                    |                       |                       |
        NbBCII10_His6  AGACTCTCCT GTACAGCCTC TGGAGGCTCT GAATACAGCT ACAGTACATT TTCCTTGGGC 120

NbBCII10_His6 (clone 1)  .......... .......... .......... .......... .......... .......... 120

NbBCII10_His6 (clone 2)  .......... .......... .......... .......... .......... .......... 120
                                    140                     160                     180
                                    |                       |                       |
        NbBCII10_His6  TGGTTCCGCC AGGCTCCAGG GCAGGAGCGT GAGGCGGTCG CGGCAATTGC GAGTATGGGT 180

NbBCII10_His6 (clone 1)  .......... .......... .......... .......... .......... .......... 180

NbBCII10_His6 (clone 2)  .......... .......... .......... .......... .......... .......... 180
                                    200                     220                     240
                                    |                       |                       |
        NbBCII10_His6  GGCCTCACAT ACTACGCCGA CTCCGTGAAG GGCCGATTCA CCATCTCCCG AGACAACGCC 240

NbBCII10_His6 (clone 1)  .......... .......... .......... .......... .......... .......... 240

NbBCII10_His6 (clone 2)  .......... .......... .......... .......... .......... .......... 240
                                    260                     280                     300
                                    |                       |                       |
        NbBCII10_His6  AAGAACACGG TGACTCTGCA GATGAACAAC CTGAAACCTG AGGACACGGC CATCTATTAC 300

NbBCII10_His6 (clone 1)  .......... .......... .......... .......... .......... .......... 300

NbBCII10_His6 (clone 2)  .......... .......... .......... .......... .......... .......... 300
                                    320                     340                     360
                                    |                       |                       |
        NbBCII10_His6  TGTGCGGCGG TGCGTGGTTA TTTTATGCGA CTACCCTCGT CACATAACTT TCGCTACTGG 360

NbBCII10_His6 (clone 1)  .......... .......... .......... .......... .......... .......... 360

NbBCII10_His6 (clone 2)  .......... .......... .......... .......... .......... .......... 360
                                    380                     400                     420
                                    |                       |                       |
                                                         His6 tag
        NbBCII10_His6  GGGCAGGGGA CCCAGGTCAC CGTCTCCTCA CGCGGCCGCC ACCACCATCA CCATCACTAA 420

NbBCII10_His6 (clone 1)  .......... .......... .......... .......... .......... .......... 420

NbBCII10_His6 (clone 2)  .......... .......... .......... .......... .......... .......... 420
```

S8. Alignment of the sequence of the amplified NbBCII10_His$_6$ gene, after site-directed mutagenesis to introduce the amber codon (TAG), with the NbBCII10_His$_6$ gene of the pHEN6-BCII10 vector.

```
                                 20                    40                    60
                                 |                     |                     |
     NbBCII10_His6  ATGGCCCAGG TGCAGCTGGT GGAGTCTGGG GGAGGCTCGG TGCAGGCTGG AGGGTCCCTG 60

NbBCII10_gly132TAG  .......... .......... .......... .......... .......... .......... 60
                                 80                    100                   120
                                 |                     |                     |
     NbBCII10_His6  AGACTCTCCT GTACAGCCTC TGGAGGCTCT GAATACAGCT ACAGTACATT TTCCTTGGGC 120

NbBCII10_gly132TAG  .......... .......... .......... .......... .......... .......... 120
                                 140                   160                   180
                                 |                     |                     |
     NbBCII10_His6  TGGTTCCGCC AGGCTCCAGG GCAGGAGCGT GAGGCGGTCG CGGCAATTGC GAGTATGGGT 180

NbBCII10_gly132TAG  .......... .......... .......... .......... .......... .......... 180
                                 200                   220                   240
                                 |                     |                     |
     NbBCII10_His6  GGCCTCACAT ACTACGCCGA CTCCGTGAAG GGCCGATTCA CCATCTCCCG AGACAACGCC 240

NbBCII10_gly132TAG  .......... .......... .......... .......... .......... .......... 240
                                 260                   280                   300
                                 |                     |                     |
     NbBCII10_His6  AAGAACACGG TGACTCTGCA GATGAACAAC CTGAAACCTG AGGACACGGC CATCTATTAC 300

NbBCII10_gly132TAG  .......... .......... .......... .......... .......... .......... 300
                                 320                   340                   360
                                 |                     |                     |
     NbBCII10_His6  TGTGCGGCGG TGCGTGGTTA TTTTATGCGA CTACCCTCGT CACATAACTT TCGCTACTGG 360

NbBCII10_gly132TAG  .......... .......... .......... .......... .......... .......... 360
                                 380                   400                   420
                                 |                     |                     |
                                                     His6 tag
     NbBCII10_His6  GGGCAGGGGA CCCAGGTCAC CGTCTCCTCA CGCGGCCGCC ACCACCATCA CCATCACTAA 420

NbBCII10_gly132TAG  .......... .......... .......... ...TAG.... .......... .......... 420
```

S9. Alignment of the amplified GFP_Y39TAG gene with the sequence of the GFP gene of the pBAD_GFP-39TAG vector.

```
                                        20                       40                       60
                                         |                        |                        |
GFP_Y39TAG (pBAD_GFP-39TAG )  atggattaca aagatgatga tgataaagtg agcaagggcg aggagctgtt caccggggtg 60

            GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 60
                                        80                      100                      120
                                         |                        |                        |
GFP_Y39TAG (pBAD_GFP-39TAG )  gtgcccatcc tggtcgagct ggacggcgac gtaaacggcc acaagttcag cgtgtccggc 120

            GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 120
                                       140                      160                      180
                                         |                        |                        |
GFP_Y39TAG (pBAD_GFP-39TAG )  gagggcgagg gcgatgccac ctagggcaag ctgaccctga agttcatctg caccaccggc 180

            GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 180
                                       200                      220                      240
                                         |                        |                        |
GFP_Y39TAG (pBAD_GFP-39TAG )  aagctgcccg tgccctggcc caccctcgtg accaccctga cctacggcgt gcagtgcttc 240

            GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 240
                                       260                      280                      300
                                         |                        |                        |
GFP_Y39TAG (pBAD_GFP-39TAG )  agccgctacc ccgaccacat gaagcagcac gacttcttca agtccgccat gcccgaaggc 300

            GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 300
                                       320                      340                      360
                                         |                        |                        |
GFP_Y39TAG (pBAD_GFP-39TAG )  tacgtccagg agcgcaccat cttcttcaag gacgacggca actacaagac ccgcgccgag 360

            GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 360
                                       380                      400                      420
                                         |                        |                        |
GFP_Y39TAG (pBAD_GFP-39TAG )  gtgaagttcg agggcgacac cctggtgaac cgcatcgagc tgaagggcat cgacttcaag 420

            GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 420
                                       440                      460                      480
                                         |                        |                        |
GFP_Y39TAG (pBAD_GFP-39TAG )  gaggacggca acatcctggg gcacaagctg gagtacaact acaacagcca caacgtctat 480

            GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 480
                                       500                      520                      540
                                         |                        |                        |
GFP_Y39TAG (pBAD_GFP-39TAG )  atcatggccg acaagcagaa gaacggcatc aaggccaact tcaagatccg ccacaacatc 540

            GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 540
                                       560                      580                      600
                                         |                        |                        |
GFP_Y39TAG (pBAD_GFP-39TAG )  gaggacggca gcgtgcagct cgccgaccac taccagcaga acacccccat cggcgacggc 600

            GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 600
                                       620                      640                      660
                                         |                        |                        |
GFP_Y39TAG (pBAD_GFP-39TAG )  cccgtgctgc tgcccgacaa ccactacctg agcacccagt ccgccctgag caaagacccc 660

            GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 660
                                       680                      700                      720
                                         |                        |                        |
GFP_Y39TAG (pBAD_GFP-39TAG )  aacgagaagc gcgatcacat ggtcctgctg gagttcgtga ccgccgccgg gatcactctc 720

            GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 720
                                       740                      760
                                         |                        |
                                              [His6 tag>]
GFP_Y39TAG (pBAD_GFP-39TAG )  ggcatggacg agctgtacaa gcatcaccat caccatcact ga 762

            GFP_Y39TAG  .......... .......... .......... .......... .. 762
```

S10. Alignment of the sequence of the GFP gene with the GFP_Y39TAG gene of the pBAD_GFP-39TAG vector, after site-directed mutagenesis to convert the amber codon into a tyrosine.

```
                        20                    40                    60
                        |                     |                     |
     GFP  atggattaca aagatgatga tgataaagtg agcaagggcg aggagctgtt caccggggtg 60

GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 60
                        80                   100                   120
                        |                     |                     |
     GFP  gtgcccatcc tggtcgagct ggacggcgac gtaaacggcc acaagttcag cgtgtccggc 120

GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 120
                       140                   160                   180
                        |                     |                     |
     GFP  gagggcgagg gcgatgccac ctatggcaag ctgaccctga agttcatctg caccaccggc 180

GFP_Y39TAG  .......... .......... ...g...... .......... .......... .......... 180
                       200                   220                   240
                        |                     |                     |
     GFP  aagctgcccg tgccctggcc caccctcgtg accaccctga cctacggcgt gcagtgcttc 240

GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 240
                       260                   280                   300
                        |                     |                     |
     GFP  agccgctacc ccgaccacat gaagcagcac gacttcttca agtccgccat gcccgaaggc 300

GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 300
                       320                   340                   360
                        |                     |                     |
     GFP  tacgtccagg agcgcaccat cttcttcaag gacgacggca actacaagac ccgcgccgag 360

GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 360
                       380                   400                   420
                        |                     |                     |
     GFP  gtgaagttcg agggcgacac cctggtgaac cgcatcgagc tgaagggcat cgacttcaag 420

GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 420
                       440                   460                   480
                        |                     |                     |
     GFP  gaggacggca acatcctggg gcacaagctg gagtacaact acaacagcca caacgtctat 480

GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 480
                       500                   520                   540
                        |                     |                     |
     GFP  atcatggccg acaagcagaa gaacggcatc aaggccaact tcaagatccg ccacaacatc 540

GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 540
                       560                   580                   600
                        |                     |                     |
     GFP  gaggacggca gcgtgcagct cgccgaccac taccagcaga acacccccat cggcgacggc 600

GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 600
                       620                   640                   660
                        |                     |                     |
     GFP  cccgtgctgc tgcccgacaa ccactacctg agcacccagt ccgccctgag caaagacccc 660

GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 660
                       680                   700                   720
                        |                     |                     |
     GFP  aacgagaagc gcgatcacat ggtcctgctg gagttcgtga ccgccgccgg gatcactctc 720

GFP_Y39TAG  .......... .......... .......... .......... .......... .......... 720
                       740                   760
                        |                     |
                                    His6 tag
     GFP  ggcatggacg agctgtacaa gcatcaccat caccatcact ga 762

GFP_Y39TAG  .......... .......... .......... .......... .. 762
```

S11. Alignment of the sequence of GAL4 with the sequence of GAL4 of the pCMV-GAL4 vector, after introducing an amber codon (TAG) at location 44 and 110 by site-directed mutagenesis.

```
                                              20                        40                        60
GAL4 wildtype translation frame +1   MKLLSSIEQA CDICRLKKLK CSKEKPKCAK CLKNNWECRY SPKTKRSPLT RAHLTEVESR 60

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... ...*...... .......... 60
                                              80                       100                       120
GAL4 wildtype translation frame +1   LERLEQLFLL IFPREDLDMI LKMDSLQDIK ALLTGLFVQD NVNKDAVTDR LASVETDMPL 120

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... ........*. .......... 120
                                             140                       160                       180
GAL4 wildtype translation frame +1   TLRQHRISAT SSSEESSNKG QRQLTVSIDS AAHHDNSTIP LDFMPRDALH GFDWSEEDDM 180

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .......... .......... 180
                                             200                       220                       240
GAL4 wildtype translation frame +1   SDGLPFLKTD PNNNGFFGDG SLLCILRSIG FKPENYTNSN VNRLPTMITD RYTLASRSTT 240

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .......... .......... 240
                                             260                       280                       300
GAL4 wildtype translation frame +1   SRLLQSYLNN FHPYCPIVHS PTLMMLYNNQ IEIASKDQWQ ILFNCILAIG AWCIEGESTD 300

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .......... .......... 300
                                             320                       340                       360
GAL4 wildtype translation frame +1   IDVFYYQNAK SHLTSKVFES GSIILVTALH LLSRYTQWRQ KTNTSYNFHS FSIRMAISLG 360

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .......... .......... 360
                                             380                       400                       420
GAL4 wildtype translation frame +1   LNRDLPSSFS DSSILEQRRR IWWSVYSWEI QLSLLYGRSI QLSQNTISFP SSVDDVQRTT 420

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .......... .......... 420
                                             440                       460                       480
GAL4 wildtype translation frame +1   TGPTIYHGII ETARLLQVFT KIYELDKTVT AEKSPICAKK CLMICNEIEE VSRQAPKFLQ 480

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .......... .......... 480
                                             500                       520                       540
GAL4 wildtype translation frame +1   MDISTTALTN LLKEHPWLSF TRFELKWKQL SLIIYVLRDF FTNFTQKKSQ LEQDQNDHQS 540

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .......... .......... 540
                                             560                       580                       600
GAL4 wildtype translation frame +1   YEVKRCSIML SDAAQRTVMS VSSYMDNHNV TPYFAWNCSY YLFNAVLVPI KTLLSNSKSN 600

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .......... .......... 600
                                             620                       640                       660
GAL4 wildtype translation frame +1   AENNETAQLL QQINTVLMLL KKLATFKIQT CEKYIQVLEE VCAPFLLSQC AIPLPHISYN 660

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .......... .......... 660
                                             680                       700                       720
GAL4 wildtype translation frame +1   NSNGSAIKNI VGSATIAQYP TLPEENVNNI SVKYVSPGSV GPSPVPLKSG ASFSDLVKLL 720

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .......... .......... 720
                                             740                       760                       780
GAL4 wildtype translation frame +1   SNRPPSRNSP VTIPRSTPSH RSVTPFLGQQ QQLQSLVPLT PSALFGGANF NQSGNIADSS 780

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .......... .......... 780
                                             800                       820                       840
GAL4 wildtype translation frame +1   LSFTFTNSSN GPNLITTQTN SQALSQPIAS SNVHDNFMNN EITASKIDDG NNSKPLSPGW 840

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .......... .......... 840
                                             860                       880
GAL4 wildtype translation frame +1   TDQTAYNAFG ITTGMFNTTT MDDVYNYLFD DEDTPPNPKK E* 882

GAL4_T44_R110 translation frame +1   .......... .......... .......... .......... .. 882
```

S12. Alignment of the sequence of the EcTyrRS gene after introduction of the PstI site, with the EcTyrRS gene of *E. coli* BL21 (DE3) pLys.

```
                        20            40            60            80
                        |             |             |             |
EcTyrRS BL21(DE3) ATGGCAAGCA GTAACTTGAT TAAACAATTG CAAGAGCGGG GGCTGGTAGC CCAGGTGACG GACGAGGAAG CGTTAGCAGA 80

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 80
                        100           120           140           160
                        |             |             |             |
EcTyrRS BL21(DE3) GCGACTGGCG CAAGGCCCGA TCGCGCTCTA TTGCGGCTTC GATCCTACCG CTGACAGCTT GCATTTGGGG CATCTTGTTC 160

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 160
                        180           200           220           240
                        |             |             |             |
EcTyrRS BL21(DE3) CATTGTTATG CCTGAAACGC TTCCAGCAGG CGGGCCACAA GCCGGTTGCG CTGGTAGGCG GCGCGACGGG TCTGATTGGC 240

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 240
                        260           280           300           320
                        |             |             |             |
EcTyrRS BL21(DE3) GACCCGAGCT TCAAAGCTGC CGAGCGTAAG CTGAACACCG AAGAAACTGT TCAGGAGTGG GTGGACAAAA TCCGTAAGCA 320

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 320
                        340           360           380           400
                        |             |             |             |
EcTyrRS BL21(DE3) GGTTGCCCCG TTCCTCGATT TCGACTGTGG AGAAAACTCT GCTATCGCGG CGAACAACTA TGACTGGTTC GGCAATATGA 400

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 400
                        420           440           460           480
                        |             |             |             |
EcTyrRS BL21(DE3) ATGTGCTGAC CTTCCTGCGC GATATTGGCA AACACTTCTC CGTTAACCAG ATGATCAACA AAGAAGCGGT TAAGCAGCGT 480

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 480
                        500           520           540           560
                        |             |             |             |
EcTyrRS BL21(DE3) CTCAACCGTG AAGATCAGGG GATTTCGTTC ACTGAGTTTT CCTACAACCT GTTGCAGGGT TATGACTTCG CCTGTCTGAA 560

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 560
                        580           600           620           640
                        |             |             |             |
EcTyrRS BL21(DE3) CAAACAGTAC GGTGTGGTGC TGCAAATTGG TGGTTCTGAC CAGTGGGGTA ACATCACTTC TGGTATCGAC CTGACCCGTC 640

    EcTyrRS_PstI  .......... .......... .......... ...G...... .......... .......... .......... .......... 640
                        660           680           700           720
                        |             |             |             |
EcTyrRS BL21(DE3) GTCTGCATCA GAATCAGGTG TTTGGCCTGA CCGTTCCGCT GATCACTAAA GCAGATGGCA CCAAATTTGG TAAAACTGAA 720

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 720
                        740           760           780           800
                        |             |             |             |
EcTyrRS BL21(DE3) GGCGGCGCAG TCTGGTTGGA TCCGAAGAAA ACCAGCCCGT ACAAATTCTA CCAGTTCTGG ATCAACACTG CGGATGCCGA 800

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 800
                        820           840           860           880
                        |             |             |             |
EcTyrRS BL21(DE3) CGTTTACCGC TTCCTGAAGT TCTTCACCTT TATGAGCATT GAAGAGATCA ACGCCCTGGA AGAAGAAGAT AAAAACAGCG 880

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 880
                        900           920           940           960
                        |             |             |             |
EcTyrRS BL21(DE3) GTAAAGCACC GCGCGCCCAG TATGTACTGG CGGAGCAGGT GACTCGTCTG GTTCACGGTG AAGAAGGTTT ACAGGCGGCA 960

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 960
                        980           1.000         1.020         1.040
                        |             |             |             |
EcTyrRS BL21(DE3) AAACGTATTA CCGAATGCCT GTTCAGCGGT TCTTTGAGTG CGCTGAGTGA AGCGGACTTC GAACAGCTGG CGCAGGACGG 1040

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 1040
                        1.060         1.080         1.100         1.120
                        |             |             |             |
EcTyrRS BL21(DE3) CGTACCGATG GTTGAGATGG AAAAGGGCGC AGACCTGATG CAGGCACTGG TCGATTCTGA ACTGCAACCT TCCCGTGGTC 1120

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 1120
                        1.140         1.160         1.180         1.200
                        |             |             |             |
EcTyrRS BL21(DE3) AGGCACGTAA AACTATCGCC TCCAATGCCA TCACCATTAA CGGTGAAAAA CAGTCCGATC CTGAATACTT CTTTAAAGAA 1200

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... .......... 1200
                        1.220         1.240         1.260
                        |             |             |
EcTyrRS BL21(DE3) GAAGAGCGTC TGTTTGGTCG TTTTACCTTA CTGCGTCGCG GTAAAAAGAA TTACTGTCTG ATTTGCTGGA AATAA 1275

    EcTyrRS_PstI  .......... .......... .......... .......... .......... .......... .......... ..... 1275
```

S13. Amino acid sequence NbBCII10.

```
                                          20                      40
                                          |                       |
NbBCII10_His6 translation frame +1  MAQVQLVESGGGSVQAGGSLRLSCTASGGSEYSYSTFSLGWFRQAPGQEREAVAAI
                                      60                      80                      100
                                      |                       |                       |
NbBCII10_His6 translation frame +1  ASMGGLTYYADSVKGRFTISRDNAKNTVTLQMNNLKPEDTAIYYCAAVRGYFMRLP
                                      120                     140
                                      |                       |
NbBCII10_His6 translation frame +1  SSHNFRYWGQGTQVTVSSRGRHHHHHH*
```

Dankwoord

Tien jaar geleden begon mijn avontuur aan de UHasselt. Na mijn Bachelor Chemie heb ik een kort uitstapje gemaakt naar Leuven voor mijn Master, maar dan ben ik uiteindelijk toch teruggekeerd naar Hasselt voor mijn doctoraat. Daarom wil ik in de eerste plaats mijn promotor Prof. Dr. Wanda Guedens en mijn co-promotoren Prof. Dr. Peter Adriaensens en Dr. Erik Steen Redeker bedanken voor het vertrouwen en de kans die ik gekregen heb om mijn project zelf vorm te geven. Achteraf bekeken was het een zeer ambitieus project dat de nodige frustraties met zich meebracht, maar ik heb er geen moment spijt van gehad. Dat heeft ook te maken met mijn toffe collega's Brecht, Tien, Rebekka, Gunter en de rest van de groep Chemie (en daarbuiten). Ik wil jullie allemaal bedanken voor de babbels en grapjes die het mogelijk maakten om de frustraties, die soms gepaard gingen met het onderzoek, snel te vergeten. Ik wil ook mijn familie: Boccia, Bomma, Ludo, Antonello en Lieve, maar in het bijzonder mijn ouders bedanken. Ma en Pa, zonder de steun en motivatie die jullie me van kindsbeen af gegeven hebben zou ik hier vandaag zeker niet staan. Dus dit doctoraat is ook een beetje jullie doctoraat. Nu wordt het tijd dat ik even overschakel naar het Spaans om mijn prachtige echtgenote Kathia te bedanken. Kathia, has sido el mayor apoyo en los últimos años y has hecho que la etapa final de este Doctorado sea mucho más fácil para mí. Quién me iba a decir que ir a la recepción de Año Nuevo de IMO sería la mejor decisión que jamás haya tomado. Ahora los dos terminamos nuestros Doctorados y ¡no puedo esperar compartir la próxima aventura contigo! ¡A. Saint-Exupéry tenía razón!