Table of contents

“Evolution is the best problem solver”

Abstract

The complexity of the proteome outnumbers that of the transcriptome by orders of magnitude, due to the diverse post-translational modifications (PTMs) of proteins. PTMs (like acetylation, phosphorylation) have a central role in physiological and pathological intracellular processes. Therefore, they represent a very rich source of targets of biological and therapeutic interest, that, however, is not deeply exploited. As a matter of fact, the functional study and validation of individual PTM targets present formidable challenges and can only be indirect.

Intracellular antibodies (intrabodies) represent, in principle, a class of molecules that could be exploited to selectively target PTMs. Recent results provided a robust selection platform to isolate intrabodies against *native* PTM-proteins: the Post-translational Intracellular Silencing Antibodies (PISA) technology. PISA permits in vivo isolation of the required intrabody from a library of antibody domains using i) yeast two hybrid system (Y2H) and ii) Tethered Catalysis, by which it is possible to create two-hybrid system baits that present a constitutive post-translational modification.

As first proof of concept, PISA yielded a specific intrabody (scFv-58F) that selectively binds acetylated lysine 9 of histone H3 (H3K9ac), which was subsequently used to bind the Acetylated Histone H3 in cell, and to obtain a functional *in cell* interference of the acetylated h3 Histone pool. In this thesis, the scFv-58F was exploited to achieve a PTM-directed interference, testing the hypothesis that targeting H3K9ac yields more specific effects on gene expression than inhibiting the corresponding modifying enzyme (HAT) that installs that PTM. The results collected here prospect the H3K9ac-specific intrabody as the founder of a new class of molecules to directly target histone PTMs, inverting the paradigm from inhibiting the enzymes to acting on the PTM mark.

Furthermore, in this thesis work, an improved version of the PISA platform, referred to as PISA 2.0, was also developed. The technology in its 1.0 form represents a significant breakthrough in fact, but tethered catalysis-based selection may present some limitations. To name but one the impossibility of *a priori* predicting the site of the single PTMs. Accordingly, here I implemented the PISA platform, exploiting the Genetic Code Expansion (GCE) methodology that allows to genetically encode the post translational (PT-) modified amino acid in a desired position in the bait protein. Specifically, new Y2H strains were generated by integration of different orthogonal aminoacyl tRNA synthetase/tRNA (aaRS/tRNA) pairs to incorporate PT-modified amino acids (namely Acetyl-Lysine and Phospho-serine) in a specific residue (TAG

codon) of bait proteins of interest, stably expressed and ready to be screened against binding domain libraries. Eventually, since all the collected results highlighted poor incorporation efficiency in *S.cerevisiae*, I took advantage from the well stated bacterial-two-hybrid (B2H) system as a intrabody selection platform and the wide range of functional and orthogonal aaRS/tRNA pairs available in *E.coli* to set up the new 2.0 version of PISA in this model organism.

In conclusion, my PhD work is part of a wider project aimed at addressing the complexity of the proteome by the selection of intracellularly working binders of functional epitopes. In particular my independent work major advances have been reached in i) in demonstrating the specific interfering activity of one of our anti PTM-intrabodies (namely the scFv58) compared to currently used indirect methods to study PTMs and notably ii) in improving the current PISA platform by exploiting Expanded Genetic Code to genetically encode the desired target PTM, in order to facilitate, accelerate and automatize the isolation of anti-individual PTM intrabodies, specifically anti-Acetyl-Lysine binders.

1.Introduction

1 Introduction

1.1 The Biological Issue: The complexity of the Proteome

The complexity of the proteome outnumbers the complexity of the transcriptome by orders of magnitude, due to the extensive and diverse post-translational modifications (PTMs) of proteins [1] **[Figure 1.1]**. It is nature's escape from genetic imprisonment: PTMs, in fact, leads to a combinatorial explosion in the number of potential molecular states providing the foundation for sophisticated forms of cellular information processing and so the emergence of organismal complexity.

As the name implies, PTMs are covalent modifications that occur after DNA has been transcribed into RNA and translated into proteins by adding a modifying group, such as acetyl, phosphoryl, glycosyl and methyl, to one or more amino acids [2].

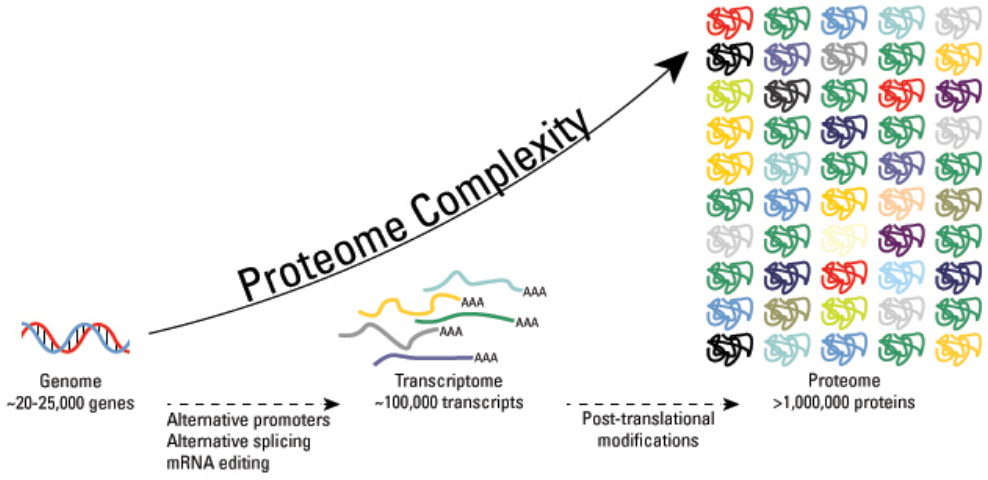


Figure 1.1. Post-translational modifications are key mechanisms to increase proteomic diversity. While the genome comprises 20,000 to 25,000 genes, the proteome is estimated to encompass over 1 million proteins. Changes at the transcriptional and mRNA levels increase the size of the transcriptome relative to the genome, and the myriad of different post-translational modifications exponentially increases the complexity of the proteome relative to both the transcriptome and genome.

Proteins are the central intermediaries between genotype and phenotype and realize resulting phenotype functions as a set of interactions mostly mediated by PTMs. PTMs are therefore both conditional edges of the cellular protein network (interactome) [3] and crucial signals to define a reversible physiological or pathological state [4].

The combinatorial insertion of different PTMs on Histones, described as "Histone code"[5], provides a remarkable example of how PTM can fine-tuned very different functional effects [6]. With respect to the type of modification and position on the tail, the "code" changes the biological meaning, and thus the downstream functional consequences. For instance, H3 histone acetylation on lysine 9 is related to increased transcriptional activity [7,8], and abnormal levels of acetylation (epimutations) alter cell function, possibly leading to malignant cell transformation [9].
Given their central role in intracellular and intercellular pathways, interest in protein PTMs especially acetylation and phosphorylation (the main PTMs according to the dbPTM database 2021) is very high [10–13]. However, while the analysis of these PTMs by proteomic techniques has made substantial progress [14], the functional study and validation of these targets present formidable challenges: what is hidden behind the intricacies of the PTM-mediated interactions still remains unclear [1].

1.2 The need: Targeting the Post-Translational-Modifications

Despite the huge interest and necessity, there is a gap between the richness of the proteome targets in cell physiology (and pathology) [10] and the development of new generation tools and drugs selectively targeting PTMs in vivo[4,9,13]. Functional studies of proteins in the context of intracellular protein networks are, in fact, a daunting task, requiring a PTM-specific loss of function. Currently, commonly used target validation tools such as genome editing (first of all CRISPR-Cas9) [15,16], gene Knock-Out (KO) [17], RNA interference (RNAi) [18], are not be able to distinguish between "targeting a defined protein PTM" and "targeting a Post Translationally Modified protein", thereby interfering with all version of a given protein (hub) simultaneously **[Figure 1.2]**. Typically, PTMs are studied indirectly, by using small molecules targeting the modifying enzyme, such as, for instance, kinases, acetyltransferase, or deacetylases [19,20] **[Figure 1.3]**. However, the information gathered with these inhibitors is biased by the pleiotropic effects of the targeted enzymes and limits their uses for target validation and therapeutic development [4,21]. Being only able to provide chemicals binding to the catalytic pocket of the enzyme, rather than to the PTM itself (and so, the branching point, or node) these chemicals show poor specificity.

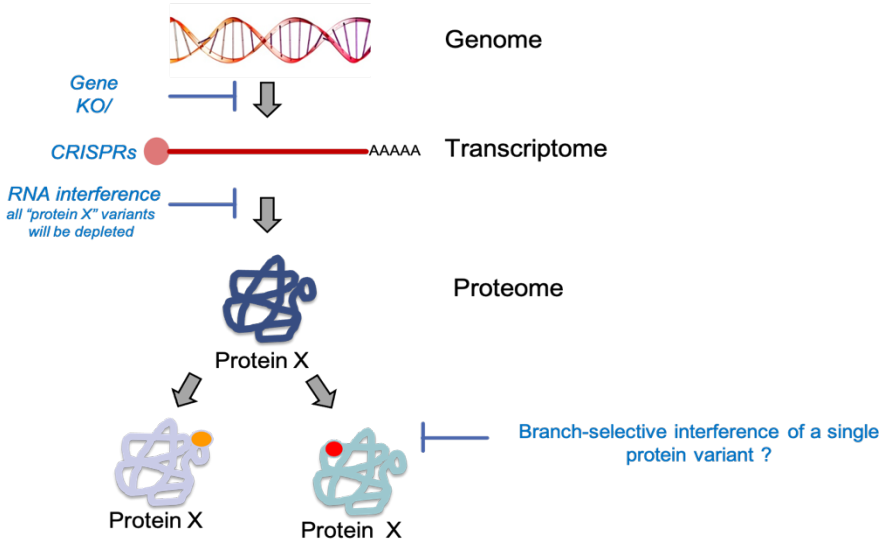


Figure 1.2.Current state-of-the-art methods to study cell function do not allow PTM-specific interference.

Besides, their action cannot be confined to organelles or subcellular compartments and there is no general selection strategy for them. Recent works have shown how to design screening schemes to discover small molecules specifically inhibiting PTM-modifying enzymes [22–24]. Nevertheless, none of these technologies can interfere precisely with single residues to study the effects of a single-modified PTM amino acid on cell function. Because of these difficulties, it is no surprising that none of the FDA-approved drugs targets a PTM directly [9].

The need for a specific PTM-interfering tool is well exemplified by acetylation, which plays a fundamental role in the epigenetic regulation of gene expression, a field of translational interest [7,8]. Histone acetylation is controlled by two enzymatic classes, the histone acetyl transferases (HATs) and the histone deacetylases (HDACs) and it promotes transcription by providing binding sites for bromodomain-containing transcription factors [bromodomain and extra terminal motif (BET) proteins][6]. Current state-of-the-art methods to study cellular mechanisms regulated by acetylation use broad inhibitors of HATs [25], HDACs [26] and BETs [27–29]; each of which modifies or interacts with several distinct proteins at different amino acid positions [21,30]. Besides histones, a variety of non-histone substrates have been shown to be acetylated by HATs, thus the HATs are now generally categorized as lysine acetyltransferases [30]. Inhibiting these enzymes will cause simultaneous loss of all of their acetylated downstream targets, thus greatly limiting the specificity and selectivity of the intervention **[Figure 1.3a-e].**

In this context, an ideal class of reagents with a virtually unlimited chemical repertoire of specificities, including PTMs, could be represented by antibodies. In particular intracellularly expressed antibody domains (called *Intrabodies*) [31], could be used for effective PTM selective

protein interference [32] **[Figure 1.3f]**. In the following chapters, I will discuss recent advances showing how intrabodies against PTM epitopes in the context of their *native* protein can be selected and how can be exploited to achieve PTM- selective interference in mammalian cells.

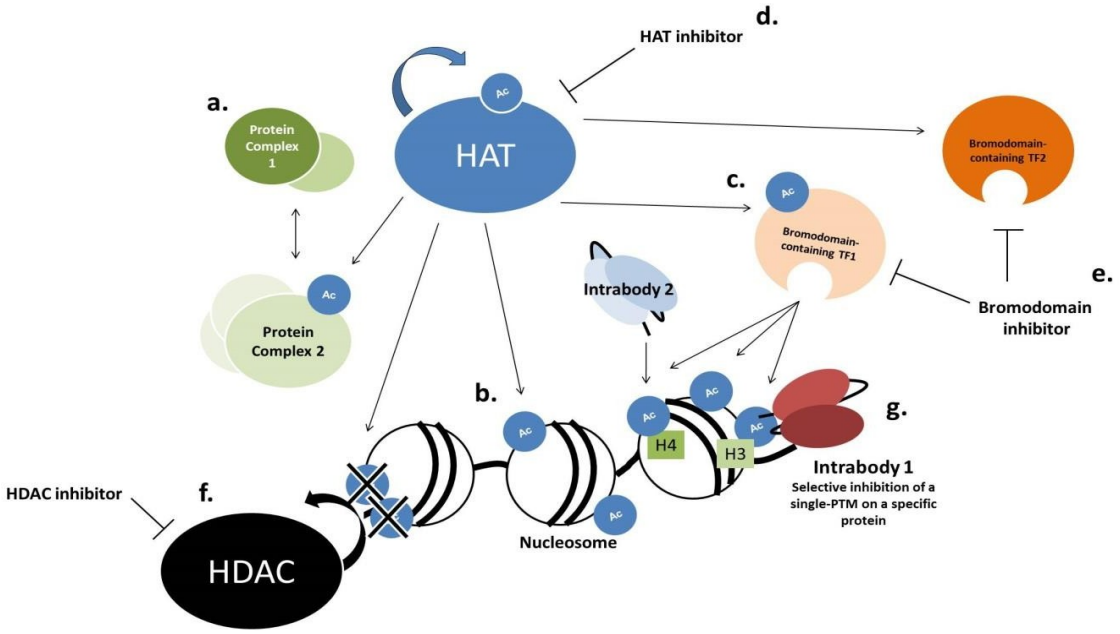


Figure 1.3. Inhibition of protein acetylation at different levels using existing chemicals and putative anti- PTM intracellular antibody domains. HAT enzymes, often associated to other proteins in large complexes, are able to acetylate a vast multitude of intracellular targets (a), including Histones (b) and Bromodomain-containing Transcription Factors (TFs) (c). Using a HAT inhibitor to study protein acetylation results therefore in the simultaneous inhibition of many downstream proteins (d). Bromodomains are acetyl-lysine binding domains present in several TFs and HATs. Inhibiting these domains means interfering with different proteins involved in very diverse processes (e). The same problems are valid for Histone Deacetylase enzymes (HDACs), which are able to remove acetyl-groups from multiple targets (f). An intrabody-based approach would be a valuable strategy to achieve a very selective targeting and inhibition of a desired PTM-modified proteins, including Histones and TFs that are involved in chromatin remodeling and transcription. The high specificity of intracellular antibodies would even permit to distinguish between different acetylated variants of the same protein target (g).

1.3 The proposed solution: Targeting the PTMs with Intrabodies

1.3.1 Antibody structure

A brief overview onto the basic antibody (Ab) structure is given since our knowledge of how the Ab structure relates to their function was exploited to create the antibody tools (intrabodies), used in this work.

Antibodies are glycoproteins produced and secreted by B-cells, which can bind biomolecules (antigenic determinants) of various nature with high specificity [33]. Immunoglobulins have a four-chain structure as their basic unit **[Figure 1.4].**

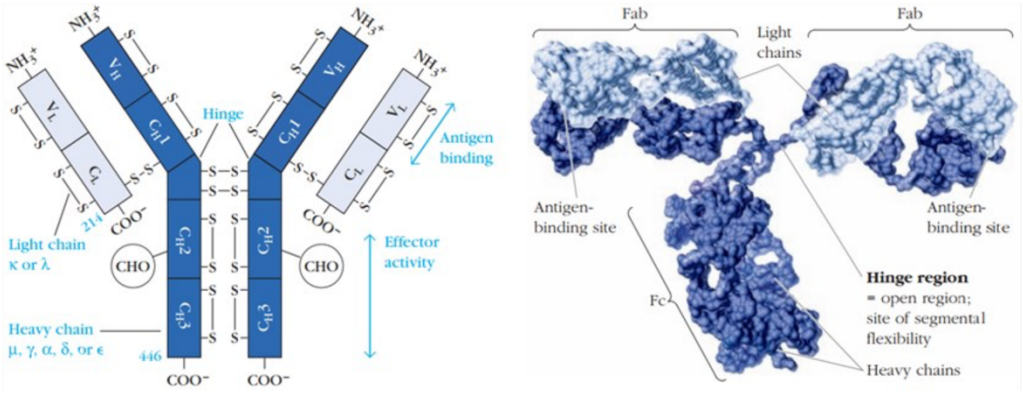


Figure 1.4. **General structure of an Antibody.** Each antibody has at least two Antigen Binding regions (Fab) and an Effector domain (Fc). The classical "Y" shape is also due to an unstructured hinge. Antibodies are composed by two heavy chains and two light chains associating together. Of particular importance to PISA technology and other recombinant antibody selection methods are Variable regions (VH and VL), for they are the minimum required domains to bind a specific antigen. CDRs=Complementarity Determining Regions. (Adapted from Owen et al.2013)

They are composed of two identical light chains (23kD) and two identical heavy chains (50-70kD). Each chain contains variable (V) and constant (C) regions (because of the amino acid diversity and conservation among different antibodies respectively) that fold in globular domains and harbor intrachain disulfide bonds, which underlie their stability. The chains are linked by interchain disulfide bonds [34].The light chain shows one constant (CL) and one variable (CV) region, while the heavy chain has a variable domain as well, and at least three constant regions (CH1, 2, 3). As illustrated in **Figure 1.4**, Variable regions (VH and VL) [35] are

responsible for the binding specificity to the antigenic determinant (epitope). Thus, the minimal antigen binding part of the antibody molecule is the combination of VH and VL, also called Fv region. Finally, Fab fragment (Fragment antigen binding) is a monovalent region of an antibody that can be obtained by proteolytic cleavage followed by a mild reduction. It is composed of one constant and one variable domain of each of the heavy and the light chains. Comparisons of the amino acid sequences of the variable regions of immunoglobulins show that most of the variability is found in three internal regions called "Complementarity Determining Regions" (CDRs)[1] or "Hypervariable Regions" (HVR) [36], as shown in Figure 5 **[Figure 1.5].** The amino acid residues between the CDRs in the Variable region are called the framework regions (FRs). FRs contain universally conserved Cysteines responsible for intrachain disulphide bonds that are particularly important for the stability and the correct folding of the molecule[34,37]. The internal space between FRs is filled with CDR1,2 and 3, structured loops that form the antigen binding site. The antibody modularity of such antibody structure is the key factor which paved the way for the development of an unsurpassed tool for research and the most rapidly growing class of approved biopharmaceutical drugs[38].

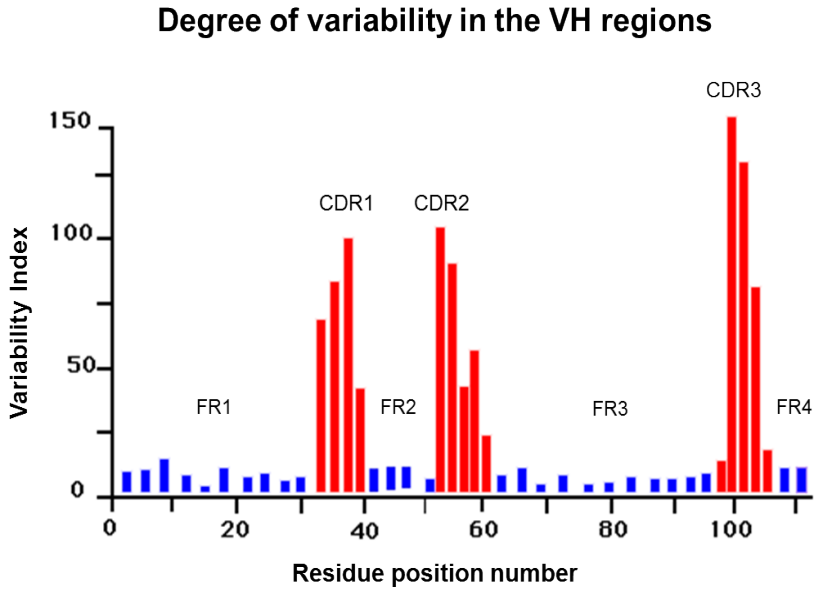


Figure 1.5. Diversity of Complementarity Determining Regions. Most of the sequence diversity reside in three defined regions called Complementarity Determining regions or Hypervariable Regions. These hypervariable regions are separated by less variable regions named as Frameworks. CDR3 has a higher index of variability if compared to CDR1 or 2. Adapted from: http://pathmicro.med.sc.edu University of South Carolina. FR= Framework Regions. Variability Index = Number of different amino acids found at that point/Frequency of the most common amino acid at that position. Thus, the index can range from 1, for a completely invariant position, to a maximum of 400 for a position in which all 20 amino acids are different. (Data from Natural Toxins Research Center at Texas A&M University – Kingsville)

[1] Within the variable *domain*, CDR1 and CDR2 are found in the variable (V) *region* of a polypeptide chain, and CDR3 includes some of V, all of diversity (D, heavy chains only) and joining (J) regions. Being the region these segments come together, the CDR3 is the most variable.

1.3.2 The Immune System in vitro: Recombinant Antibodies and Antibody Libraries

Our body is able to produce a virtually unlimited repertoire of antibodies (more than 10^12) against every kind of antigen [33]. Hence, thanks to their high affinity and specificity, antibodies have established themselves as the mainstay in research and biomedical applications [38]. Most importantly, immunoglobulins demonstrated to be excellent therapeuticals for the treatment of different diseases. In 1986, the Food and Drug Administration (FDA) approved the first therapeutic monoclonal antibody [39] for the prevention of transplant rejection. By the end of 2019, a total of 79 antibody-based drugs were approved to treat a range of autoimmune conditions, infectious diseases and cancers [40] . Many more potentially exciting therapies are also in the pipeline — including against HIV, Ebola and COVID19 [41–43].

This immeasurable contribution to research and medicine is rooted in the discovery of the hybridoma monoclonal antibodies by Milstein and Köhler [44], which started the so-called "antibody revolution" and spurred the rise of recombinant antibodies. The availability of antibody variable regions is sufficient to reconstitute high affinity recombinant antibodies, surpassing in this way the difficulties related to cloning and expression of whole immunoglobulin genes and allowing for versatility in effector functions and formats. Gene technology later revolutionized this potential as antibody genes could now be altered to order. In particular, the DNA recombinant technologies have allowed the reformatting of antibodies in different smaller formats based on Fv for their intracellular expression **[Figure 1.6]**:

Single chain variable fragment (scFv): the scFv format is formed of a VH and VL domain connected by a flexible linker, which is most commonly glycine- and serine-rich with dispersed hydrophilic residues [45]. This linker is long enough to allow the VH and VL domains, on the same polypeptide chain, to pair. It is believed that the linker must span a ~3.5 nm distance between the VL and VH without disrupting the formation of the antigen binding site [46]. In the last decades many scFv therapeutics have entered clinical development [47].

Single domain antibodies (also called Nanobodies) are composed either by the VH or VL domain of an Immunoglobulin. They are small (15KDa) and are able to bind target antigens with as high affinity as has been observed for traditional antibodies [36,48].

Single domain antibodies in fact form long finger-like extensions that can potentially target cryptic epitopes that are difficult for intact antibodies. As of 2022, 1 nanobody therapeutic has been approved by the FDA (for the treatment of thrombotic thrombocytopenic purpura) and multiple are being tested in clinical trials for the treatment of a range of diseases including cancers, rheumatoid arthritis and RSV infection [50]. In addition to their promise as

therapeutics, nanobodies have multiple further applications, including in life sciences research and as diagnostics (e.g., medical imaging for cancer diagnosis) [50]. scFv and single domain antibodies are successfully used both as in vitro probes and as intracellular antibodies in cells, and have become the standard format for the creation of in vitro diverse antibody repertoire DNA libraries. The source of these antibodies can be either the naïve repertoire or a "biased" repertoire from animal immunization, or a synthetic designed repertoire. In particular one home-made mouse naïve scFv library[51] and one home-made *llama glaba* heavy chain antibody (VHH) naïve library [52] in the Single Pot Libraries of INTrabodies (SPLINT) format [53,54] have been used in my thesis project.

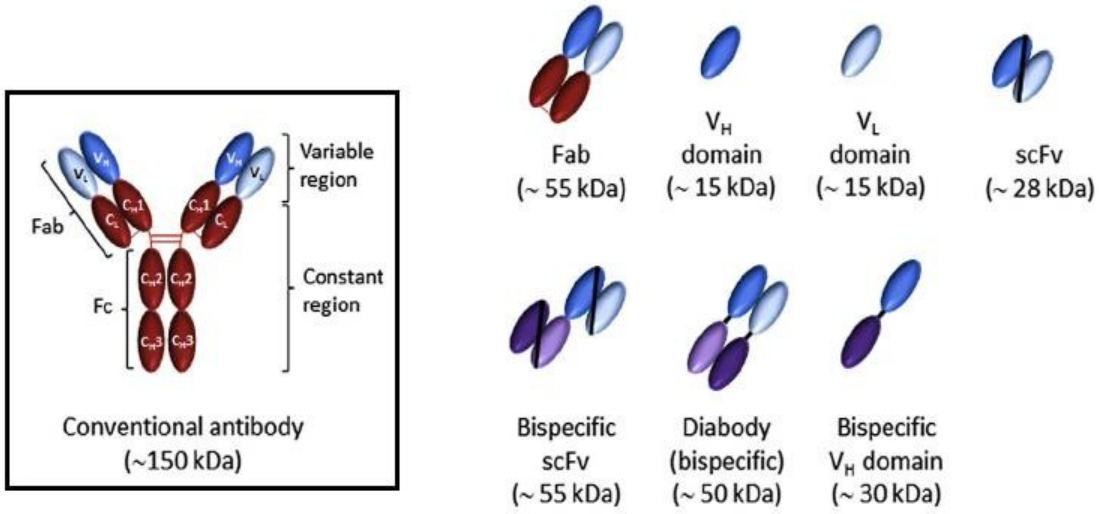


Figure 1.6. Recombinant antibody domains. Fab= Fragment antigen binding (CH1-VH + CL-VL); VH= Variable Heavy region; VL=Variable Light region; scFv= single chain Fragment variable (VH-linker-VL); Bispecific scFv (two scFvs of different specificities linked together in a single chain); Diabody= bispecific recombinant antibody in which VH-VL pairing of the same specificity is achieved using different chains; Bispecific single (e.g. VH) domain= structurally corresponds to just one chain of the diabody. Adapted from: Marcotte et Hammarström – "Passive immunization: towards magic bullets". Mucosal Immunology (2015).

These antibody domains are input of the so-called display technologies [22,55], which allow in vitro selection of antibody domains against a desired antigen, usually employed in form of peptide in solid phase. Despite the existence of several selection platforms, all of them share the property of associating a particular paratope specificity to a precise DNA sequence (phenotype to genotype connection), greatly facilitating the downstream manipulation of the isolated antibody binding site. Since display technologies are performed exploiting microorganisms or in vitro translation systems, the turnaround time for antibody generation is

reduced, and the potential for high- throughput generation of binders is greater[56]. In contrast to animal immunization, where there is little control over the nature of antibodies produced, manipulation of selection conditions can be carried out in vitro, for example, by including competitors for the direct selection of targets or epitopes of interest. Undoubtedly, the most popular display platform is Phage Display[55,57]. In this technology, antibody fragments are displayed on filamentous phage M13 by fusing them to the pIII or pVIII coat proteins, thanks to a phagemid vector and a helper phage, required for phage packaging and release respectively [58]. Antibody libraries with diversity up to ~ 10^{11} are cloned in the phagemid vector [58]. Phage particles exposing the antibody fragments are incubated with tar- get proteins, which can be immobilized to the surface of a microtiter plate well or an affinity column [59]. Direct coating of target proteins can be a problem for particular proteins or peptides that are affected by immobilization-associated issue, so the target antigens are often subjected to a biotinylation process allowing biotinylated proteins to be indirectly coated on plastic previously saturated with streptavidin, or bound to magnetic streptavidin-coated microbeads[22,60]. The latter approach is particularly popular since it guarantees the phageantigen interaction to occur in solution, preservation of the target structure and optimal epitope exposure [61,62].

After removal of aspecifically-bound phages, the best target-binding phages are eluted and used to infect bacteria, in order to enrich them in the phage pool. This process, called "panning", is repeated several times to isolate the best target specific antibodies **[Figure 1.7]**. Of course, this system has been used to identify or engineer 14 marketed therapeutic antibodies[63]- and many others undergoing clinical trials [64], nonetheless there are a few limitations that become apparent when targets with PTMs are considered. The ease of the panning selection with phage display is, in fact, offset by disadvantages with respect to the native PTM-oriented application. Pitfalls include i) the need to use unstructured linear peptides with chemically synthesized PTMs instead of native PTM proteins[65], and ii) generating antibody binders that are not adequate for their use in cells. Thus, the conformation of a linear unstructured PTM-containing peptide used as antigen will be different from that of the corresponding PTM-modified sequence in the context of the native protein Therefore, these anti-PTM antibodies will not always recognize the native protein. Moreover, selecting antibody domains from phage display libraries that fold in the periplasm of *Escherichia coli* cells does not guarantee that the antibody will function as a neutralizing intracellular antibody within the protein network in the cell, because of the different redox conditions [62,63].

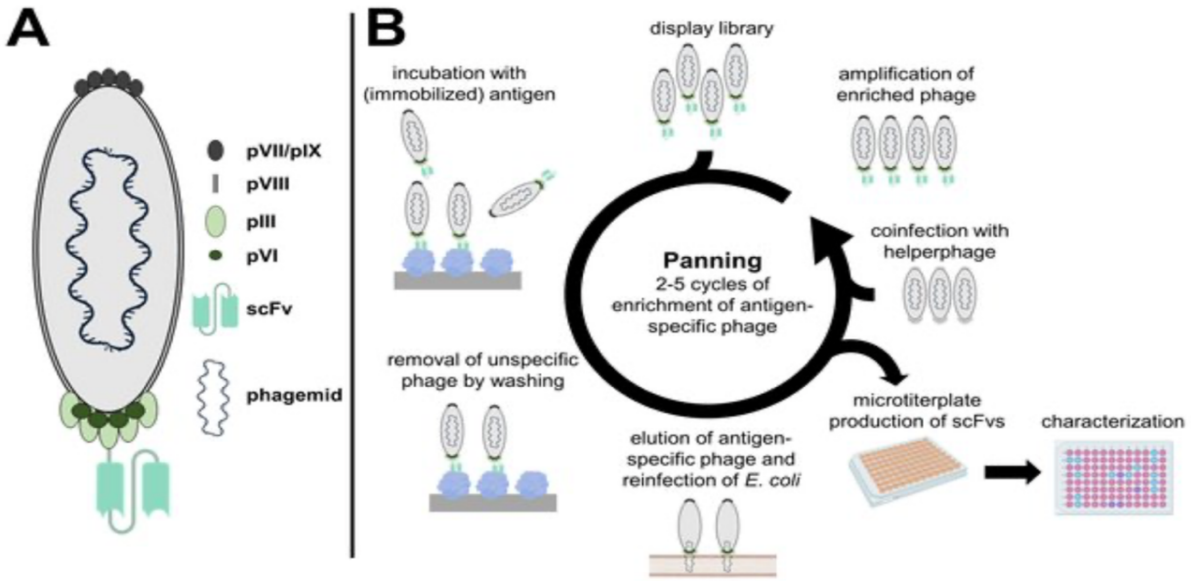
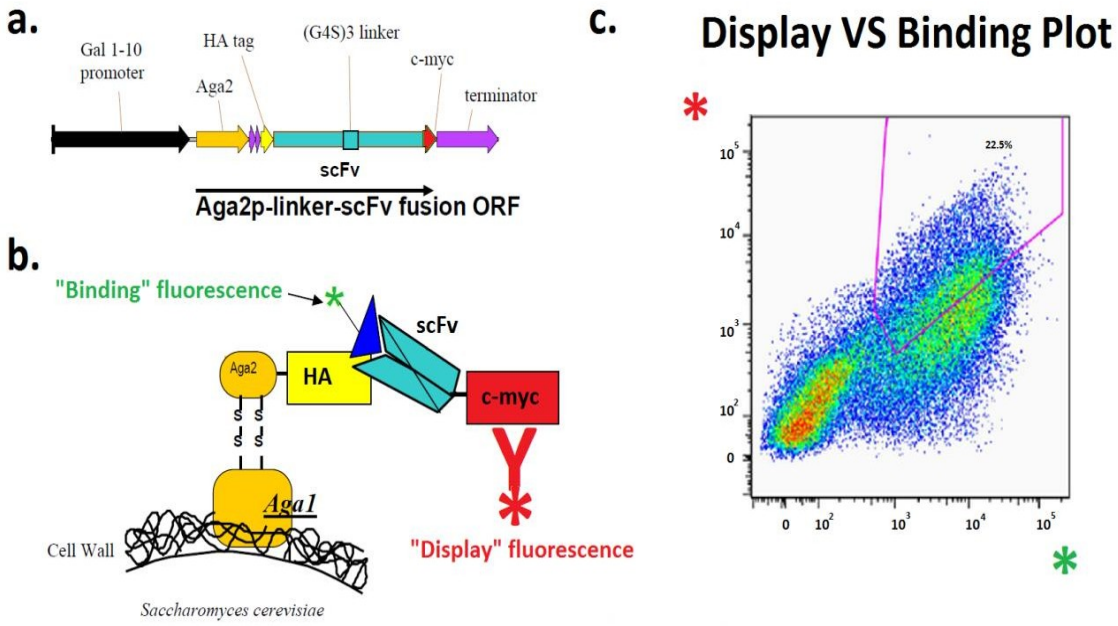


Figure 1.7. Schematic of Phage Display panning. (right panel) Phages displaying antibody clones from a diverse library are produced in *E.coli* and incubated in 96-well plate coated with the target antigen. Removal of the residual phage suspension and washing allows only specific anti-target phage-antibodies to remain bound to the antigen. Selected phages are afterfwards eluted and amplified in bacteria to allow other panning cycles. Early cycles will have a significant proportion of contaminant phages, but after 4-5 pannings the anti-target population will increase until reaching the almost totality of the enriched sub-library. After these cycles, phage-containing bacteria can be plated to amplify single colonies (and thus produce monoclonal phages) to be sequenced and tested for specific activity to the target. Nowadays, Phage Display sublibraries are sequenced at each cycle with Next Generation Sequencing to monitor specific clone enrichment and to easily get binding site information (mainly CDR3 sequence). **(Left panel)** Schematic representation of a M13 phage, the most commonly phage used in phage display. Note how the most represented protein is pVIII (capsid protein), while pIII is only poorly expressed and localised at the tip of the phage.

Another popular display technology is yeast surface display or yeast display [66] in which antibody domains such as scFvs or Fabs are displayed on *S.cerevisiae* in fusion with a surface protein. Unlike Phage Display, Yeast Display allows expression of antibody domains exploiting a eukaryotic system, which is able to confer to the proteins complex post- translational modifications such as glycosylation, possibly mimicking the mammalian environment. Another advantage in comparison to Phage Display is represented by the selection process, obtained by FACS **[Figure 1.8]**, that provides an immediate information about a wide range of binding affinities. With phage display, instead, binding strength can only be controlled by a threshold of stringency defined by the washing solutions used. Drawbacks of yeast display are instead a more limited number of diverse displayed items of the library (around 10^7-10^8) and significant polyclonality of antibody clones, due respectively to lower yeast transformation efficiency (even though some methods allow transformation up to 10^{10} cells) [67] and its ability

to accept more than one plasmid per transformation [68,69]. Description of the procedure can be found in **Figure 1.8.**



Lastly, worth of mention is surely the Ribosome Display [70]. In Antibody ribosome display, libraries of antibody domains, usually scFvs, are transcribed and then translated directly in vitro. The DNA library is genetically fused to a spacer sequence lacking a stop codon before its end. The lack of a stop codon prevents release factors from binding and triggering the disassembly of the translational complex. So, this spacer sequence stays attached to the peptidyl tRNA and occupies the ribosomal tunnel, allowing the antibody domain to protrude out of the ribosome and fold. What results is a complex of mRNA, ribosome, and antibody, which provides the physical linkage between genotype and phenotype. As valid for phage display, the ribosome library is panned to surface-bound target, and the complexes that bind well are immobilized. Subsequent elution of the non-specific binders is achieved via high salt concentrations, chelating agents, or mobile ligands. After several panning cycles, the antibody domains mRNA can then be reverse transcribed into cDNA and isolated [**Figure 1.9**]. Advantages of Ribosome Display over other cell-based in vitro display technologies consist of: (i) much greater number of diverse library clones, which is only dependent on ribosome functionality in the mixture; (ii) usage of error-prone PCR at every cycle as a tool for generating further diversity.

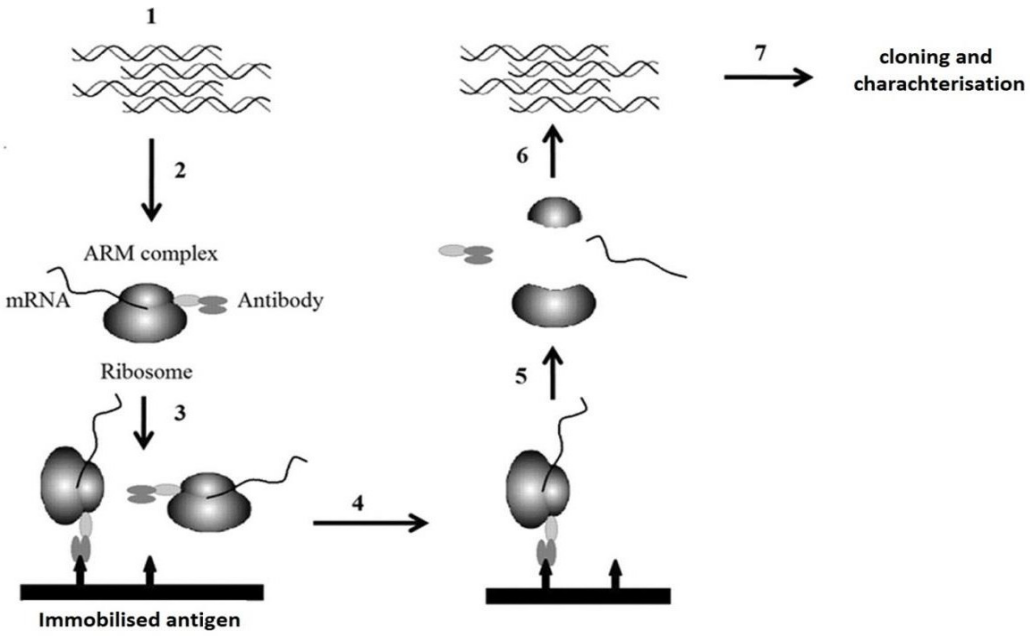


Figure 1.8. Scheme of the steps involved in *in vitro* antibody ribosome display. Step 1: The scFv antibody library (in this example) DNA template is prepared flanked with a T7 promoter sequence. **Step 2**: the antibody library DNA is transcribed to mRNA and translated to form an antibody-ribosome-mRNA complex (ARM complex). **Step 3**: the mixture is incubated with immobilized target antigen. **Step 4**: the unbound components are removed by washing. **Step 5**: the retained complexes are released. **Step 6**: the mRNA is reverse transcribed and amplified by PCR. **Step 7**: the PCR products are cloned in expression vectors and sequenced.

1.3.3 Ectopic Expression of Antibody Domains for Intracellular Immunization: using antibodies as genes

Even though in ordinary laboratory life we are used to thinking of antibodies merely as protein probes, antibody binding domains can be used as genes, in fact, and can be ectopically expressed in living cells to perturb levels of a particular intracellular antigen, a vision first proposed by Antonino Cattaneo and Michael S. Neuberger in 1987 [67]. As anticipated, these tools are called *Intracellular Antibodies* or *Intrabodies,* when expressed in different subcellular compartments (see below) but along the same lines also secreted antibodies can be ectopically expressed as genes for interference purposes [71–73]. The latter concept is now being considered for the delivery of therapeutic antibodies [74].

After this demonstration that even antibodies of the IgM secretory form (the most complex) can be efficiently assembled and secreted in non-lymphoid cells such as neurons, in the absence of supposedly obligatory lymphoid-specific quality control molecules and processes [75], the following step forward was to ask if it was possible to redirect antibodies to cytosol or to any other subcellular compartment [68], since Immunoglobulins are normally membrane proteins or secreted proteins and they are doomed to be part of the so-called "secretory pathway". While the polypeptide chain exits the ribosome, the hydrophobic N-terminal sequence (called the secretory leader) can be held by SRP, producing a translation blockage. The whole complex is then co-translationally inserted into the endoplasmic reticulum (ER) lumen, where antibody chains are allowed to fold correctly, to form intra- and inter-chain disulphide bonds, to be glycosylated and assembled. Before being secreted, antibodies have to pass a severe "quality control" to check their correct folding [76]. It was therefore absolutely not obvious that antibodies could be expressed intracellularly, away from the secretory pathway. What has been shown to be necessary and sufficient to target an intracellular antibody to a cell compartment different from the secretory pathway is to remove the so-called signal sequence (or leader sequence for secretion) and equip the antibody coding sequence with subcellular targeting sequences that are defined by being dominant and autonomous [33]. Intrabodies could be therefore equipped with targeting sequences to be redirected from a secretory destiny to an intracellular one by exploiting what cell biology of protein trafficking is teaching us. For instance, for a cytoplasmic expression the secretory leader can be just removed or replaced with a cytoplasmic consensus or a Nuclear Localization Signal may be inserted to target the protein to the nucleus.

The first demonstration of antibody relocation and subcellular targeting in mammalian cells is due to Cattaneo, Biocca and Neuberger in 1990, who described successful cytoplasmic and

nuclear expression of antibody chains [77]. This investigation led off to the concept of Intracellular Immunization with Intrabodies as a new strategy for functional studies in living cells by protein silencing. The first protein silencing with intrabodies targeted the p21/RAS protein [78,79] showing the generality of the approach to different biological systems. After this first demonstration of effective phenotypic knock out with intrabodies, a huge number of intracellular antibodies were reported to successfully interfere with the corresponding protein in cells [32,56,80]. For a precise targeting within the secretory pathway, the leader sequence for secretion needs to be retained and additional targeting sequences, such as the ER retaining C-terminal SEKDEL sequence are included [31].

Of course, as we will see, playing with signal peptides is often not sufficient to a protein that naturally folds into an "oxidative", chaperon-rich environment as the ER, to keep the same tertiary structure (and thus its function) also into different sub- compartments. The main reason behind this problem is represented by the so-called "folding problem". By ectopically expressing a protein into a reducing environment such as the cytoplasm, correct disulfide bond formation (especially intrachain bonds) and proline residue isomerisation by PPI enzyme cannot occur, so the native folding of the antibody may be hindered. Consequently, the cytoplasmic expression of an antibody domain is somewhat unpredictable. Intrachain disulfide bonds are one of the hallmarks of the immunoglobulin fold architecture, with almost the totality of known human and mouse germ line V regions containing a conserved pair of Cys residues (less than 0.4% of VH domains lack at least one of them) [81]. One of the reasons for this conservation appears to be the relatively low intrinsic stability of many V domains, which requires additional stabilization by the disulfide. However, it can be surmised that, in the vast natural antibody repertoire, domains which are intrinsically more stable may tolerate the removal or the absence of the disulfide bond, and conversely, that those that do tolerate disulfide bond removal may be intrinsically more stable than average [81].

One example is represented by ABPC48[82], a levan-binding antibody, which was the first natural antibody characterized in which the CysH92 residue is missing in favor of a tyrosine. Surprisingly, the unpaired cysteine residue appears to be exposed to the solvent, in contrast to the deeply buried disulfide bond of ordinary variable domains. This implies a particular conformation in which the VH is stabilized by the interaction of the TyrH92 residue with the CysH22. Another comparative study between the cytoplasmic expression in mammalian cells of well characterized scFv and camelid VHHs (selected from antibody libraries based on similar scaffold) have outlined the physico-chemical determinants that correlate with enhanced intracellular solubility and stability (intracellular solubility and stability are strictly related): the main factor determining antibody folding is the formation of an intrachain disulphide bond between Cys33 of CDR1 and Cys109 of CDR3 which exposes the CDR3 region for the appropriate recognition of the antigen, in addition to give stability to the structure. On the

contrary, llama antibody does not have this bond, but displays a hydrophobic core that packs Phe37 with Trp100 of CDR3, ensuring intrinsic stability[83]. Soluble expression in the cytoplasm is influenced by CDRs content and by the overall charge and hydropathicity of the intracellular antibody domain sequences[83,84].
From these few examples, it appears that variable domains have a range of folding stabilities and that overall stability of the fold is contributed by many critical residues or combinations of residues in the framework regions [85,86]. The absence of the cysteine disulfide bonds will be therefore tolerated or not, according to the overall structure and stability of each individual variable domain. The issue is very important and two different strategies have been followed to improve stability and solubility of antibody fragments under conditions of intracellular expression: knowledge-based [87] and selection-based approaches. In the "intracellular expression of natural antibody" scenario, therefore, one could either try to engineer, in a case-by-case process, existing antibody clones targeting a particular intracellular antigen (knowledge-based approach) or try to exploit the natural tolerance that a subset of variable domains show (intracellular selection). Rational design of "well-behaved" intracellular antibodies includes the construction of synthetic CDR diversity onto Variable region frameworks derived from well-behaved intrabodies [88–90]. Notwithstanding the successful examples of rational design, it is difficult to identify a priori -in a general way- those amino acids replacements that are needed to introduce stabilization. Also, the resulting diversity is more limited than that of natural V-region repertoires and the randomized CDRs might re-introduce destabilizing sequences. Therefore, the question arises as to how to possibly access the naturally most stable antibody domains from the natural antibody repertoire domains? Is it possible to design an antibody-antigen intracellular selection strategy?

Motivated by this question, Visintin et al. demonstrated that in the naïve repertoire of mouse immunoglobulins there is a significant subset of intrabodies to target virtually any antigen, and that scFv cDNA libraries derived from these non-immunised animals (SPLINT libraries) [53] can be directly selected inside yeast cells against a desired antigen to overcome the non-predictability of expression of antibodies [91,92]. This technology is called Intracellular Antibody Capture Technology and is described in the next paragraph.

1.4 Intracellular antibody selection methods

As extensively described in the previous chapter, the interior of the cell poses significant challenges regarding antibody folding, assembly, and functionality and will often result in reduced efficacy of intrabodies. It is clear that the ex vivo selections do not guarantee the identification of antibodies suitable for intrabody applications[81,93]. Therefore, it would be preferable to develop an efficient, fast, and easy to handle intracellular selection system if the final application requires intrabodies. A number of variants of intracellular selection system exist: the yeast two hybrid (Y2H) selection[94]; the intracellular antigen capture-Y2H based technology (IACT) [95], the bacterial two hybrid[96,97] and the protein fragment complementation assay (PCA)[98,99] or bimolecular fluorescence complementation assay (BiFC)[100]. Due to their relevance to the subject of this thesis, the two-hybrid (2H) techniques are described here in more detail. Both the B2H and Y2H platforms were used to set up the technology presented in my work.

The two hybrid systems are one of the most popular, cost effective and scalable *in vivo* genetic approaches [101]. They are based on the principle of restoring protein activity upon non-covalent reconstitution of split protein fragments. One fragment of a modular protein is fused to a protein X, and the other fragment is fused to a protein Y. The resulting fusion proteins may be referred to as bait and prey. If the proteins X and Y interact, the modular protein is reconstituted, regains its activity, and its activity is detected through a reporter gene. Although both the bacterial two hybrid (B2H) and yeast two hybrid (Y2H) work on similar principles, there are fundamental differences and both have distinct advantages and limitations. Both methods can be used on a small scale or as high-throughput screening approaches. On one hand, Y2H system has been extensively tested and sometimes is preferred over B2H systems, given that users have created more resources such as a variety of vectors and other modifications[102,103]. On the other hand, the B2H method seems a better fit than the Y2H system to screen membrane associated proteins. In the following chapter, I provide detailed protocols for the yeast (Y2H and IACT) and bacterial two hybrid systems exploited in this project[101].

1.4.1 The Intracellular Antibody-Y2H based Capture Technology

The IACT [92] is an application of the two-hybrid system based technology [94], to the in vivo selection of functional intrabodies against a wide range of targets, using yeast. Basically, IACT allows the isolation of functional intrabodies selected from a pool of genes (library), which are all able to tolerate the lack of disulphide bonds and efficiently recognize the intracellular desired antigen at the same time.

It reproduces the clonal selection process that selects B-cells in vivo based on the interaction of the expressed antibody with an antigen [33]. Like the B-cell constitutes a physical link between the antibody protein and its gene, this system is made in a way that such a physical link allows researchers to select the genes based on the interaction phenotype of the encoded antibodies [104]. In IACT, L40 yeast strain is used. L40 is not able to grow in absence of histidine (it is auxotrophic for histidine) because of the lack of HIS3 gene, a fundamental enzyme for the production of histidine. Like in the two-hybrid system, L40 yeast is co-transfected with two different vectors. As shown in **Figure 1.10**, "vector 1", the so-called "prey", bears one of the putative intrabody clones of the antibody domain library that are fused to the VP16 activation domain. "Vector 2", instead, represents "the bait", which is composed by the DNA binding domain (DBD) of LexA (a bacterial protein, exogenous to yeast genome) attached at the N-terminus of the protein antigen to be screened. LexA-fusion protein will be able to bind a series of contiguous LexA operons that flank the HIS3 and LACZ genes. Thus, a positive interaction between bait and prey will determine the expression of His3 and LacZ enzymes, causing a particular yeast cell to form a colony on histidine-lacking medium. On the contrary, cells in which no antigen-intrabody interaction occurs, will die when plated on the same medium. Colonies grown onto His-deficient media are also tested for beta-galactosidase activity in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (abbreviated in X-Gal) to exclude false positives. Clones that show both good growth and a blue color are called "Double Positives' and taken into account for further binding confirmations and characterization of the bound area (epitope mapping). Finally, intrabodies are usually also validated in mammalian cells and with an in vitro binding assay to prove their affinity and specificity for the target in other assays.

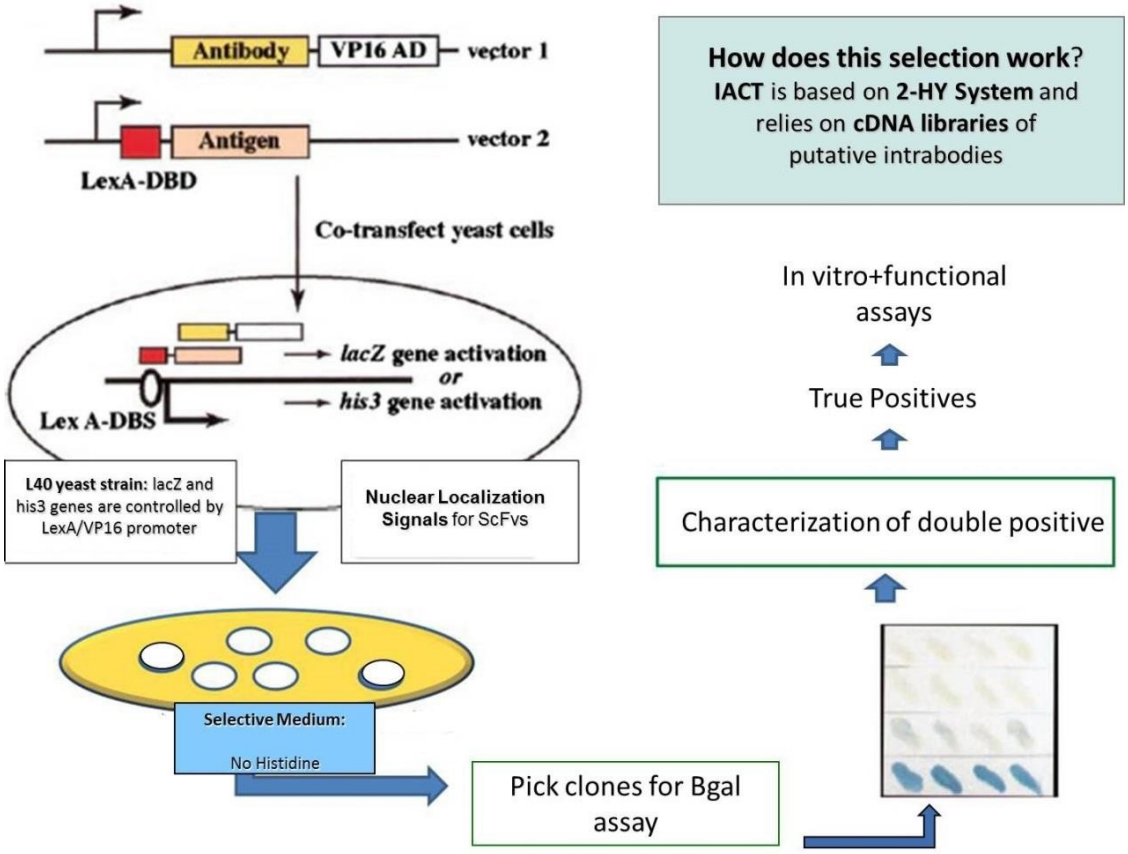


Figure 1.9. **The IAC Technology.** A library of putative intrabodies is screened against a bait (target protein) in the L40 yeast strain. If the intrabody (the prey) is able to interact with the bait, His3 and LacZ genes will be transcribed, allowing yeast cells to survive onto a selective medium lacking Histidine. Moreover, these cells should also turn blue if treated with X-Gal. Double positives are then further characterized by a secondary screening to confirm their true positivity. These intrabodies will be also used for in vitro and in cell assays to confirm interaction and functionality (Adapted from Visintin et al. PNAS 1999).

1.5 The Post Translational Intracellular Silencing Antibodies Technology (PISA)

Even though the IAC technology has proven to be a valuable strategy to isolate intrabodies against a very large number of different protein antigens, things become more challenging if the epitope to be targeted is a Post-Translational Modification (PTM), which is the main objectives of the PISA project. Such a screening would require a constitutive PTM on the bait used for the selection. Indeed, PTMs are very labile epitopes, they result from a balance between the activity of two antagonist enzymes (e.g., Histone Acetyl-Transferase VS Histone deacetylase). Moreover, bait modification by endogenous enzymes might not be sufficiently high to guarantee a stable modification on every copy of the recombinant protein.

The PISA selection, a novel strategy to achieve intracellular selection of antibody domains against PTM-proteins (e.g., acetylated proteins), was developed in our lab to face this no-trivial challenge. In the PISA technology embodiment, the Intracellular Antibody Capture Technology (IACT) [91,95] and Tethered Catalysis [105] were combined together. As described above, IACT allows isolation of conformation-sensitive intrabodies directly in the intracellular environment, from a library of antibody domains, with no need of protein manipulation. Tethered catalysis, the other component of PISA, allows the production of fusion baits consisting of the protein sequence to be modified and targeted, genetically fused to the modifying enzyme that normally acetylates (or phosphorylates) the native protein in vivo. Use of these PTM chimeric proteins as baits allows surpassing the need for purification of protein targets, which would be extremely difficult for native PTM proteins. Indeed, antibody display platforms rely on peptides in order to isolate anti-PTM binders[106–108].
Basically, PISA screening **[Figure 1.11]** works by co-transforming L40 yeast strain with bait/prey couples in which bait is represented by a tethered catalysis construct, whilst prey is one of the putative intrabody clones from the SPLINT cDNA library[53]. Baits are thus chimeric LexA-fusion proteins in which the target antigen is fused to a modifying enzyme like HAT, which modifies the target of interest. The positive bait-prey interaction activates transcription of HIS3 and LacZ genes and allows yeast clone survival on selective media. HIs3+/LacZ+ clones are subjected to fingerprint analysis. The selected clones undergo plasmid DNA extraction from yeast cells and a transformation in bacterial cells in order to isolate individual antibodies in case of yeast polyclonality. A secondary screening against the original baits, and a counter-screening against truncated and unrelated baits, which are unable to harbor any PTM, are performed to identify true positive intrabodies that bind to the PT-Modified epitope

exclusively. After the targeted epitope is assessed, tagged versions of the intrabodies are expressed in bacteria and purified. Biochemical and biological assays can be performed to test the affinity, specificity and sensitivity of the selected anti PTM-intrabodies. PISA technology can be intrinsically extended to different PTMs such as acetylation and phosphorylation [109], and it is not limited to scFv domains; libraries of alternative binding domains, such as Nanobodies [110] or DARPins [111], can be easily adapted to this selection format.

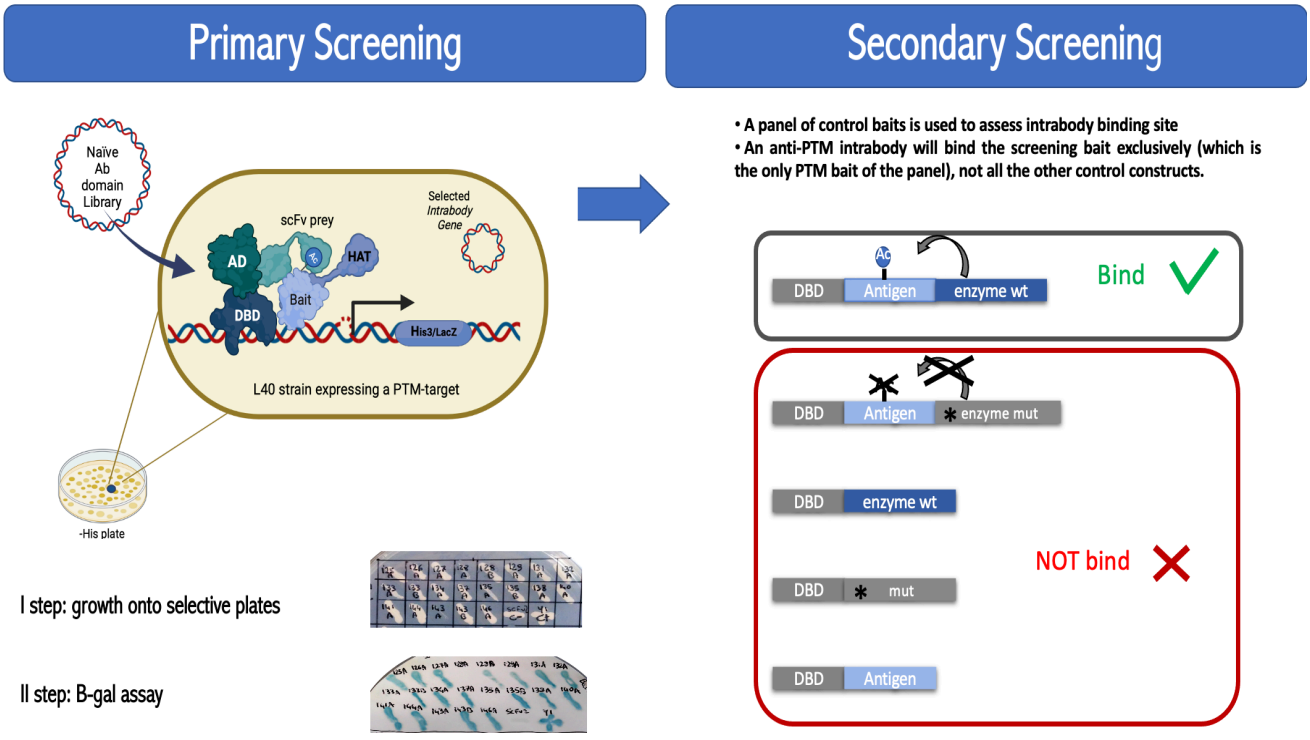


Figure 1.10. Schematic of the PISA screening: how to select intrabodies against PTMs? A SPLINT naïve library (Single Pot Library of intrabodies) of either ScFvs or VH domains is co-transformed with a tethered catalysis LexA-bait in L40 2-HY yeast strain. In tethered catalysis, the LexA-Antigen bait is fused to the enzyme that naturally modifies it in vivo to achieve a constitutive, in cis PTM. In our case, we used HAT enzymes (Histone Acetyl Transferase) to induce acetylation on the targets inside the cell. Only in case an intrabody binds to any part of the PTM-bait, yeast plated on Histidine-lacking medium will grow, producing a colony to be tested also for beta-gal activity. His+/LacZ+ plasmid clones are then extracted and co-transformed again with the screening bait to confirm interaction, as well as with a panel of control baits to map the bound epitope (Secondary screening). Intrabodies specific for acetylated antigen will be only able to bind the original screening bait. Unlike other antibody-based technologies, PISA requires no manipulation of the protein antigen, which is folded and modified as a native protein.

1.5.1 The PISA's Intrabodies in Action

As the first proof of concepts, Chirichella et al challenged PISA against acetylated proteins[32]. They chose HIV1 integrase and yeast H3 histone (fused to their corresponding P300 and Gcn5 acetyltransferases modifying enzymes) as test baits. The selections yielded two specific intrabodies: an anti HIV-1 acetylated integrase (Ac-IN) and anti-acetyl-K9 Histone H3 (H3K9Ac), which were subsequently used to obtain functional interference in vivo[32]. Remarkably, for the first time, the authors have demonstrated biological downstream effects following PTM-selective interference in living cells with intrabodies.

As for the integrase bait, after evaluating the specific binding of anti-AcIN in vitro **[Figure 1.12]**, they demonstrated acetylation selective functional interactions and interference in cells: the early interception of HIV-1 integrase by the intrabody in the cytoplasm significantly inhibits infectivity.

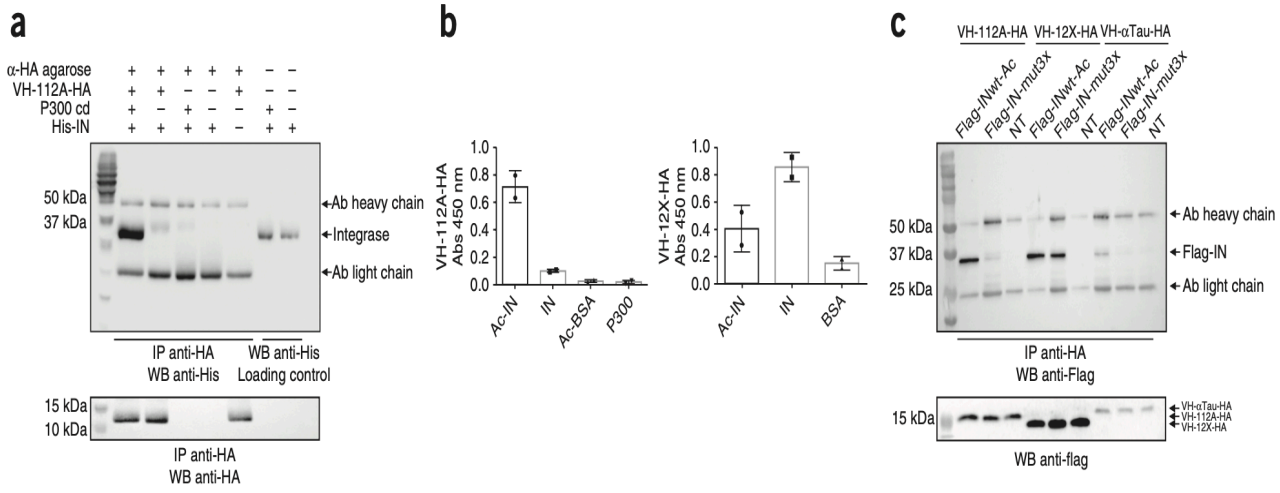


Figure 1.11. Biochemical validation of anti-acetylated integrase intrabody VH-112A. (**a**) Coimmunoprecipitation (co-IP) of HA-tagged VH-112A with His-tagged acetyl-integrase versus integrase (both expressed in Escherichia coli), by anti-HA agarose, followed by anti-His (top) and HA (bottom) blotting. (**b**) ELISA showing binding of VH-112A-HA and VH-12X-HA with acetyl-His-integrase, coated in solid phase. BSA, acetylated BSA, and P300 enzymes are negative controls. Data are shown as individual points and means ± s.d. N = 2. (**c**) Co-IP of HA-tagged antibody domains with FLAG-tagged wildtype (wt) or 3× mutated (K264R, K266R, K273R) integrase in HeLa cells (Chirichella et al 2017. Nat Methods).

To probe the generality of the PISA platform, a second acetylation-dependent intrabody was selected against acetyl-K9 Histone H3 (scFv-58F), in its native form or when it is in chromatin. The histone and H3 Lysine specificity of scFv-58F was determined in vitro and in yeast cell where the intrabody was shown to specifically and directly bind the H3K9ac epitope. Accordingly, they showed that scFv-58F, intracellularly expressed (IE-scFv58F) in the yeast

nucleus, specifically binds and immunoprecipitation native H3AcK9. Moreover, IE-scFv58F reduced acetylated histone availability in chromatin **[Figure 1.13].**

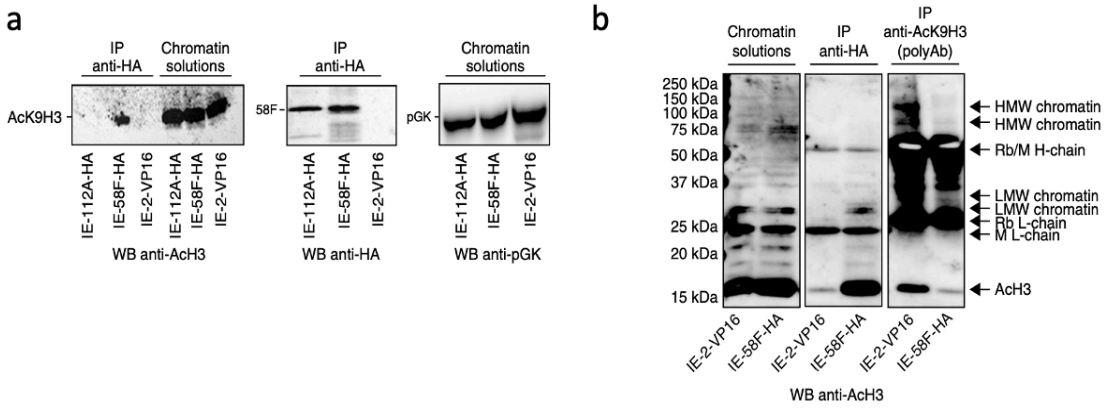


Figure 1.12. Biochemical characterization and validation of anti-acetylated H3 histone intrabody ScFv-58F. (a) IP with anti-HA agarose of chromatin solutions (soluble extracts) from yeast cells intracellularly expressing ScFv-112A-HA, ScFv-58F-HA, or ScFv-2-VP16. Membrane blotted for AcK9H3 and HA. **(b)** IP of chromatin solutions from yeast cells intracellularly expressing either the unrelated ScFv-#2-VP16 or ScFv-58F-HA, blotted for AcK9H3. Left, chromatin ladder from soluble yeast extracts used in the IP experiment in c. Middle, IP of chromatin solution using anti-HA. Right, IP of chromatin solutions from both yeast cell samples with a polyclonal anti-AcK9H3 antibody. HMW, high molecular weight; LMW, low molecular weight; Rb, rabbit; M, mouse; H-chain, heavy chain (IP antibody); L-chain, light chain (IP antibody); AcH3, acetyl-K9-H3 histone; IE, intracellularly expressed. The three panels in b derive from the same gel (Chirichella et al 2017. Nat Methods).

Notably, the scFv-58F was proved to exert functional consequences on transcription in yeast cells. Microarray analysis revealed that approximately 100 mRNAs were selectively regulated in a statistically significant way by scFv-58F, with a greater number of downregulated genes (this higher number of downregulated genes showed the silencing effect of the scFv-58F). Since H3 acetylation regulates transcription, masking the AcK9H3 site in cells might avoid many bromodomain-containing proteins to bind their target, and hence regulate gene expression. In conclusion, the authors strongly demonstrated that the intracellular interference of scFv-58F with AcK9H3 has significant functional consequences on the transcriptome, providing the first evidence of such interference mediated by single-PTM inhibition in living cells **[Figure 1.14].** They also envisaged that acetylation specific intrabodies could be used to selectively target only the acetylated protein pool or tracking acetylated chromatin *in vivo*.

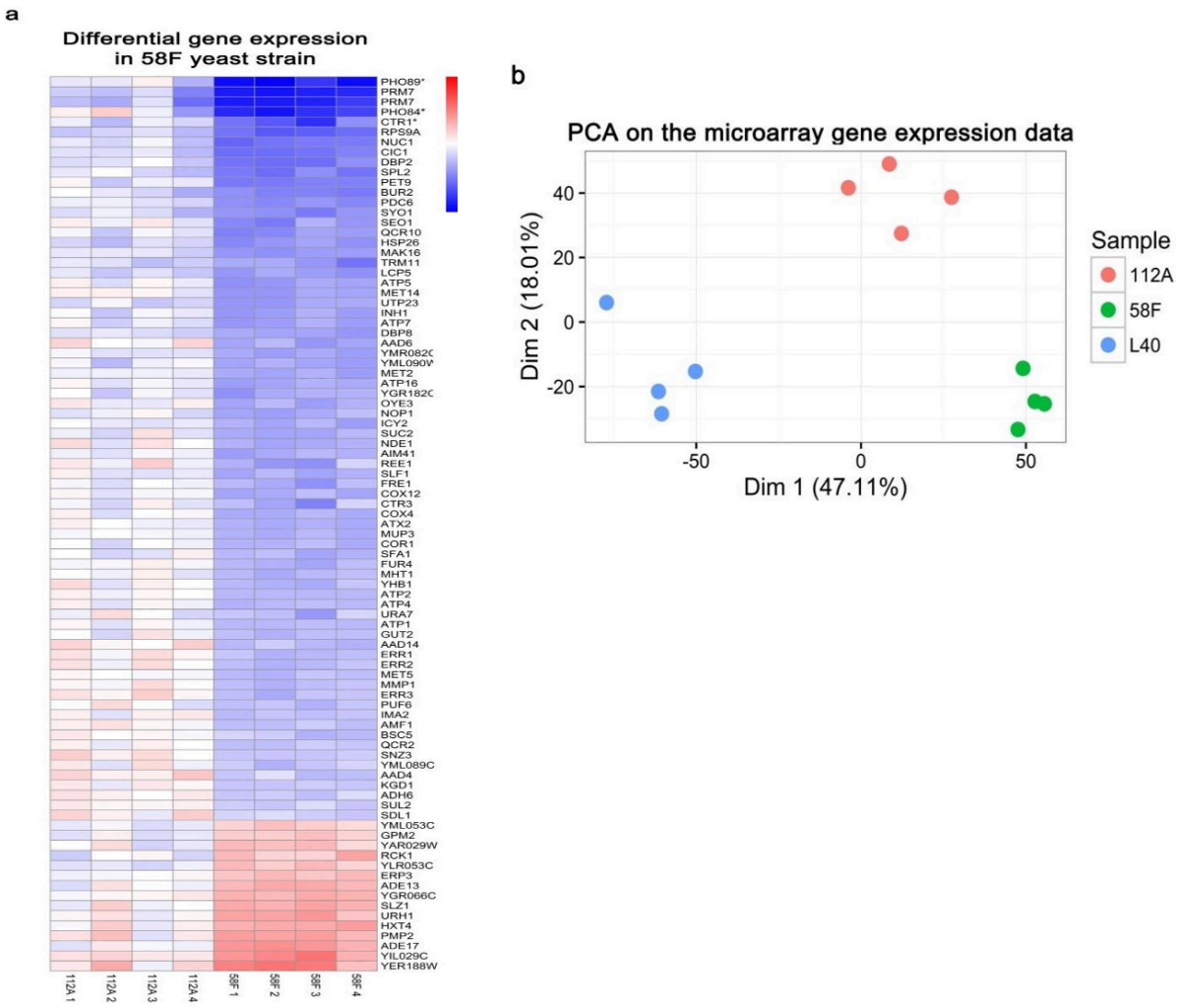


Figure 1.13. Functional validation of anti-acetylated H3 histone intrabody ScFv-58F. Selective interference with acetylated H3 histone by the ScFv-58F intrabody regulates gene expression in yeast cells. (**a**) Heatmap of the significant differentially expressed genes between scFv112A (anti-acIN) and scFv58F (anti-H3AcK9) samples. Gene expression levels were filtered out to retain those showing a significant over- or under- expression (pAdj<0.05). Data were filtered according to the following conditions: i) pAdj<0.05 in the 58F-112A comparison; ii) L40-112A pAdj > 0.05; iii) L40-58F pAdj < 0.05. The values were normalized to the L40 sample levels. Four biological replicates for each sample were used. The list was further reduced applying a fold change threshold of 1.5. The statistical analysis was performed with R limma package and Benjamini correction was applied to the pValue (pAdj). In the graph, the marked (*) entries represent genes with a known association to histone acetylation. (**b**) Principal Component Analysis on the microarray gene expression data. Each sample is composed of four independent biological replicates. Dim1 and Dim2 represent the first two principal components, the proportion of variance (POV) held by these components is reported in brackets as percentage (Adapted from Chirichella et al 2017. Nat Methods).

Importantly, intrabodies targeting Histone PTMs, such as the scFv58F, represent ideal candidates to dissect some Histone code information and to shed light on the functional understanding of chromatin modifications [5,6]. As a matter of fact, they differ from the commonly-used class of HAT or BET inhibitors targeting acetyltransferases ("epigenetic

writers") and bromodomains ("epigenetic readers")[108] respectively, because they do not target epigenetic modifiers. Rather, they target epigenetic marks (e.g., AcH3K9 or other modifications), allowing more focused biological questions to be addressed. Accordingly, it will be of interest to compare the predicted superior specificity of inhibiting PTM marks with respect to the current approaches, that target instead the modifying enzymes responsible for installing the PTM [51]. During my thesis work, I have contributed to a collaborative project in the lab aimed at addressing these questions [112]. We experimentally tested this prediction, by asking whether the use of the anti-H3K9ac intracellular antibody scFv-58F produces a more defined and specific modulation of gene expression than that induced by the inhibition of the chromatin modifying enzymes (epigenetic writers) that install that acetylation.

1.6 Beyond Tethered Catalysis: PISA & Expanded Genetic Code

As anticipated, in the present embodiment, PISA technology relies on tethered catalysis[105]: the yeast-two-hybrid (Y2H) host cells express a bait genetically fused with the catalytic domain of a modifying-enzymes to be constitutively PT-modified in cis. This approach turned out rather general and flexible, and its extension to different PTM-inducing enzymes, including kinases, has been successfully established in the lab. However, tethered catalysis PISA screening has some limitations providing only a partial solution to the possibility of dissecting the contribution of each individual PTMs.

- First, large bait dimensions might be subjected to extensive degradation and difficult migration into the nucleus;
- Moreover, a reduction of transformation efficiency is plausible;
- Additionally, this embodiment hampers the impossibility of a priori designing sites of single PTMs: chimeric baits can be, in fact, in principle modified at all the possible sites;
- Last but not least a limitation is rooted in the fact that many modifying enzymes are unknown, curbing the researcher's imagination;

A natural extension of the technology would be to perform PISA selections in 2H platform cells where the expanded genetic code technologies [113] permit site-specific genetically encoded of several PT-modified amino acids in baits for intrabody selection **[Figure 1.15].** Genetic code expansion (GCE) technologies[113–115] (described in the next paragraph) allow the addition of non-canonical amino acids (ncAAs)[2], by reprogramming the genetic code of various organisms. Coupling tRNA and aminoacyl-tRNA synthetase (tRNA/aaRS) pairs [116], obtained by molecular evolution, allows the incorporation of ncAAs into proteins in a living cell[116–122]. In thesis project, I purpose a novel PISA protocol (the PISA 2.0 version), in which the two-hybrid host cells will be genetically modified to host an orthogonal, *in vitro-evolved* synthetase able to recognize a PT-modified amino acid like acetyl-lysine [123–125], via its corresponding tRNA, which bears a CUA anticodon and a bait mutagenized to insert the amber stop codon UAG at the desired position. During translation, the tRNA will thus recognize the amber stop

[2] Amino acids beyond the canonical 20 plus pyrrolysine and selenocysteine. ncAAs are not naturally co-translationally incorporated into proteins, are normally synthetized chemically in the laboratory and they can have a myriad of different side chain structures and backbone configurations. The abbreviations ncAAs and UAAs (unnatural amino acids) are used as synonyms in this thesis work.

codon UAG, and the synthetase will interpret it as "acetyl-lysine" instead of a STOP signal [125].

In conclusion, Genetic Code expansion technology would allow the creation of baits carrying a single genetically encoded PTM, stably expressed and ready to be screened against a binding antibody domain library. Implementation of this method bears several advantages. It will allow a higher grade of selectivity to isolate intrabodies against a single PTM in a given protein and it will also eliminate the need for large fusion baits (the actual protein target) decreasing the possible nonspecific interactions. Besides, once a system for the co-translational insertion of a specific ncAA is implemented, it will be applicable to every protein. This is also a great advantage over the current PISA platform, which needs cloning of a specific enzyme for each new bait protein, even if the chemical nature of the PTM is the same. Finally, a plethora of proteins carrying the amber stop codon at the desired PTM residue position can be designed a priori; and given the growing list of UAAs [115,118,119] , the potentiality of the technology can be limited only by imagination.

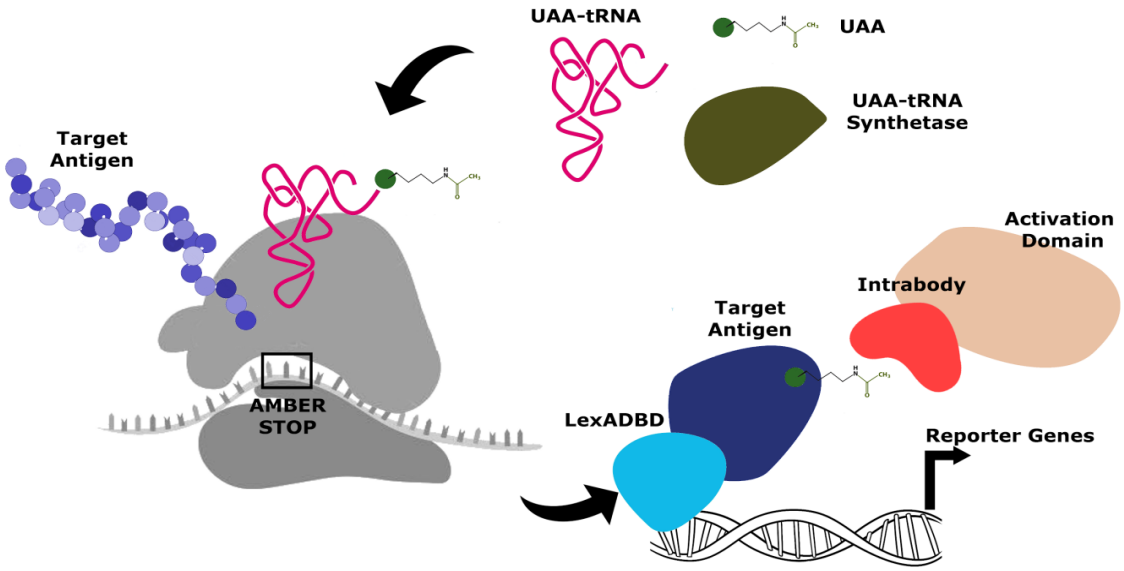


Figure 1.14. Illustration of the PISA 2.0 Platform: the application of the Genetic Code Expansion method to PISA. The upgrade of the PISA technology foresees the substitution of tethered catalysis with expanded genetic code technologies to encode the PTM genetically in the target protein. After a successful selection of a tRNA /AA-tRNA synthetase matching set able to work in yeast, the pair will be integrated in the S. cerevisiae L40 screening strain as an amber codon reallocation. The new strain will then enable to perform PISA screening to select intrabodies against PTMs genetically encoded in the target antigen through the use of intragenic amber stops. UAA = unnatural amino acid

1.7 Learning from Nature to Expand the Genetic Code

1.7.1 Genetic Code Expansion: history and state of the art

Protein synthesis is a fundamental process that involves the transcription of the genetic information contained in the DNA sequence into an mRNA molecule that will be used as a template for the condensation of amino acids into a polypeptide chain. Such a process is guided by the macromolecular complex of the ribosome[2,126]. Amino acids are delivered to the ribosome linked to an adaptor, a tRNA molecule, capable of base pairing with the mRNA codons, thus coupling the alphabet of DNA with the one of proteins. The fundamental role to link amino acids with the correct tRNAs, generating the decoding scheme, is provided by a class of enzymes known as aminoacyl-tRNA synthetases. Codons are composed of three nucleotides. Since DNA, as well as RNA, has four possible base pairs, there are a total of $4^3 = 64$ possible codons. Sixty-four triplet codons are used to encode the 20 canonical amino acids, and the initiation and termination of protein synthesis [2]. Accordingly, with the notable exceptions of the atypical amino acids selenocysteine (in all domains of life)[127] and pyrrolysine (in some methanogenic archaea and bacteria)[128,129], all proteins are synthesized from this limited set of building blocks: the 20 canonical amino acids[130,131]. Though a number of arguments have been put forth to explain the nature and number of amino acids in the code, it is clear that proteins require many additional chemistries to carry out their natural functions[132]. Thus, while a 20 amino acid code is sufficient for life, it may by no means be ideal.

The discovery of natural non-standard genetic codes in various organisms (including Archaea [128,133], bacteria [134], ciliata [135] and mitochondria of many higher eukaryotes [122,136,137]) inspired the development of the Genetic Code Expansion (GCE) field[114–116,118,120,138–140].

Recently, synthetic biologists explored if the genetic code could be expanded by introducing non canonical amino acids into proteins in vivo through the engineering of the translational apparatus of the cells [116,121,122,136,138,140,141]. Genetic code expansion now enables the site-specific, co-translational incorporation of hundreds of diverse ncAAs into proteins synthesized in cells and animals [117,120–122,142]. This approach empowered us to address important questions that cannot be tackled by classical approaches [113,118,119,140,142–144]. For example, it allows the genetic encoding of post-translationally modified versions of natural amino acids in a desired residue of a protein[124,145]. This methodology provides both a powerful approach to overcome the limitation of chemical and biosynthetic strategies[3] [118]

[3] Both chemical and biosynthetic strategies have been developed to incorporate ncAAs into proteins. The former are simple and straightforward but are often limited by the homogeneity and/or size of the protein that can be synthesized. For example, chemical modification of amino acid side chains can lead to nonselective and non- quantitative derivatization. Moreover, only a limited number of

that have been used so far to incorporate ncAAs into proteins and turned particularly useful to implement our PISA 2.0 platform.

Even though a new wave of research in the field is driving strategies for the wholesale reprogramming of the genetic code and cellular translation [113,144], here I will provide key background on GCE experiments based on the incorporation of one type of ncAA at a time in response to an amber codon (the so called "amber suppression methodology" or "UAA methodology") [115] **[Figure 1.16]**, being the approach used in this work.

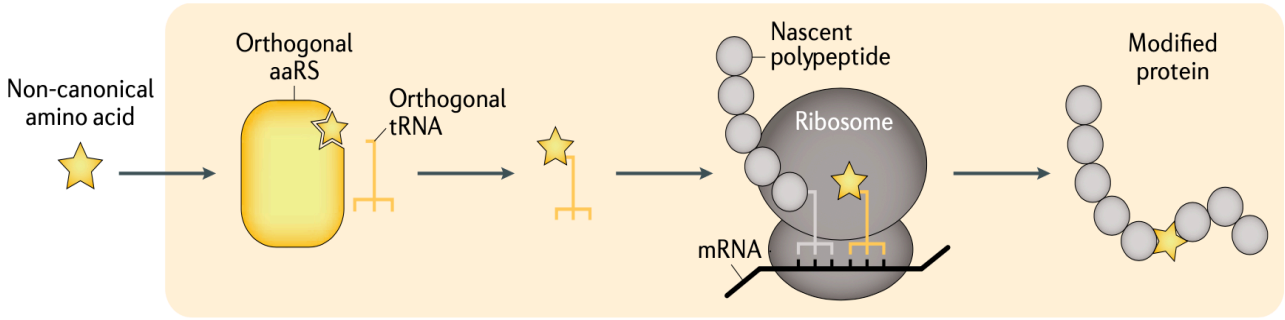


Figure 1.15. Schematic of Genetic Code Expansion Methodology. To incorporate an unnatural amino acid (star) into a protein, the amino acid is normally added to the cell growth medium and is then taken up by the cell. A cell (orange box) expresses an 'orthogonal' aminoacyl-tRNA synthetase (aaRS) which specifically aminoacylated an orthogonal amber suppressor tRNA with the unnatural amino acid. The aminoacylated orthogonal tRNA enters the ribosome, where it reads an amber codon in the mRNA; this codon was introduced into the DNA that encodes the transcript for the target protein. Protein synthesis creates a modified protein bearing the non-canonical amino acid at the encoded position. (Adapted from De la Torre et al. Nature Review, 2021)

residues can be chemically modified with exogenous agents. Solid-phase peptide synthesis allows a large number of modifications to be made to protein structures but is generally limited to peptides and smaller proteins owing to the decreased yield and purity associated with the synthesis of larger proteins. The recent development of chemical and intein-mediated peptide ligation allows the semisynthesis of larger proteins, but substitutions are confined largely to the N or C terminus.

1.8 The Genetic Code Expansion for *site-specific* incorporation of PTM-AAs in a designed residue.

The co-translational introduction of a ncAA at a defined site in a protein directly in a living organism can be achieve with a tRNA/synthetase pair that does not interfere with the endogenous translational apparatus (namely, that is orthogonal), and that can link the ncAA to the tRNA. Each orthogonal aaRS (o-aaRS) aminoacylates its cognate orthogonal tRNA, but minimally aminoacylates the other tRNAs in the host organism; similarly, the orthogonal tRNA is aminoacylated by its cognate synthetase, but minimally aminoacylated by the aaRSs of the host. To ensure that the ncAA is incorporated uniquely at the site specified by its codon, the tRNA must be engineered to recognised orthogonal codon not assigned to a canonical amino acid, while maintaining selective aminoacylation by the o-synthetase [116,117,138,139]. A unique codon is in fact needed to direct the expression of the ncAA without compromising the existing genetic code[146]. The amber stop codon TAG is the least used among the three stop codons in E.coli and yeast, rarely terminates essential genes, and is efficiently translated by amber suppressor tRNA in vivo and in vitro[147,148]. Hence, the use of TAG to encode novel ncAAs is expected not to significantly perturb the growth of a host organism used for the PISA technology[142,149].

A two-step procedure is now well established to isolate orthogonal synthetase–tRNA pairs, with the vast majority of efforts taking place in *E.coli*[117,122] **[Figure 1.17 a-b]** . In the first step, a synthetase–tRNA pair is selected from an organism that is evolutionarily divergent from the host of interest[150]. When expressed in the host, the o-synthetase and o-tRNA should not cross-react with the host synthetases and tRNAs because their sequences and structures would be diverged thanks to natural evolutionary processes. As a matter of fact, to ensure accurate genetic code interpretation, aaRSs select their cognate tRNAs through a defined set of nucleotides (known as identity elements), located in the anticodon loop and the acceptor stem of the tRNA[151]. tRNA identity elements are not fully conserved throughout evolution for many aaRS/tRNA pairs, which prevents cross-reaction between aaRSs and tRNAs from a different species[151,152]. Furthermore, specific interactions with the anticodon are characteristically absent in some aaRSs, such as pyrrolysyl-tRNA synthetase (PylRS)[48], making their corresponding tRNAs well-suited for codon reassignment [151,153]. However, the pairs do, in general, still recognize a natural amino acid in the host cell; so, a second step is required to alter the specificity of the synthetase enzyme in order to exclusively recognize an unnatural amino acid and not natural amino acids[140]. This can be achieved by a two-step genetic selection in which large libraries of mutations in the synthetase active site are created and then synthetase variants that uniquely use the unnatural amino acid are selected in

positive/negative screening protocols[140,154] **[Figure 1.17b]**. At the end, this approach creates an orthogonal aaRS/tRNA$_{CUA}$ pair that decodes the amber codon (TAG), also called amber suppression pair.

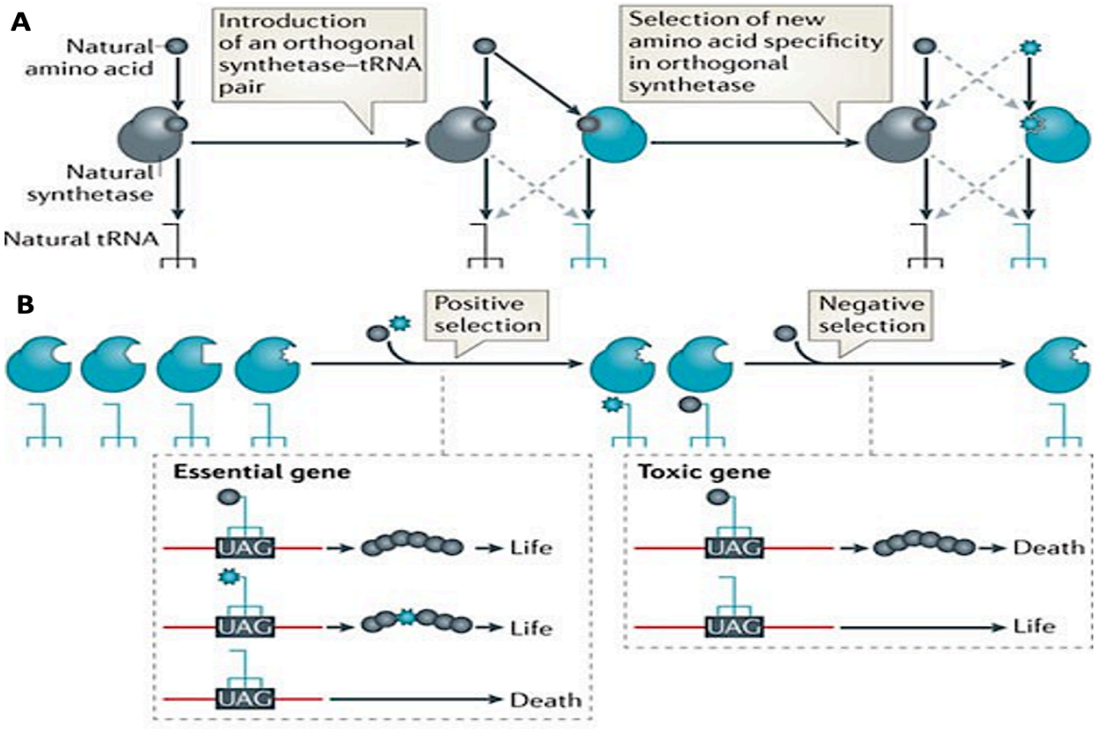


Figure 1.16. Double Sieve Selection to isolate o-aaRS/tRNA pair. A. Orthogonal synthetase tRNA pairs are generated in two steps: import of a heterologous aminoacyl-tRNA synthetase–tRNA pair into a host containing a set of natural synthetases that use natural amino acids; and the subsequent selection of a mutated active site in the orthogonal synthetase that recognizes an unnatural amino acid. **B.** To generate a synthetase with this altered specificity, a large library of active-site variants of the synthetase is subject to positive selection for activity with either natural or unnatural amino acids, by virtue of their ability to suppress an introduced stop codon and so allow complete translation of a gene that is essential for survival. The synthetases that use natural amino acids are subsequently removed by a negative-selection step, in which they use natural amino acids to suppress a stop codon introduced in a toxic gene, which leads to cell death.

The active site of the synthetases could be further engineered (when necessary), often through bioinformatics analyses combined with directed evolution protocols[4] [155], to selectively use newly desired ncAAs or to select variants with higher catalytic efficiency and

[4] **Directed molecular evolution** (**DE**) is a method used in protein engineering that mimics the process of natural selection to steer proteins or nucleic acids toward a user-defined goal. It consists of subjecting a gene to iterative rounds of mutagenesis (creating a library of variants), selection (expressing those variants and isolating members with the desired function) and amplification (generating a template for the next round). It can be performed *in vivo* (in living organisms), or *in vitro* (in cells or free in solution). In this manuscript however, I will use molecular evolution to indicate the *in vitro* molecular evolution.

specificity[119,121]. In particular, many ncAA-charging aaRS variants have been evolved using conventional methods of mutagenesis (site-saturation mutagenesis in the aminoacylation active site) followed by selection or screening techniques [156]. Nonetheless, innovative methods have been recently applied to the evolution of o-aaRSs: again in *E.coli* **[Figure 1.18a-c]** [150]. The use of phage-assisted continuous evolution **[Figure 1.18b]** [157], multiplex automated genome engineering [158] **[Figure 1.18a]** and several flow cytometry-based strategies **[Figure 1.18c]** [159] have resulted in a number of aaRS suitable for use with genetic code manipulation. Of course ,tRNA engineering can also improve the efficiency of an o-aaRS/tRNA pair by optimizing orthogonality of the tRNA[145], its binding to the o-aaRS or elongation factor Tu[160], or the decoding strength of the targeted codon[161]. In contrast to *E.coli*, efforts to engineer OTSs in *S.cerevisiae* or mammalian cells to date have use only a narrow range of approaches [162]. Finally, another prerequisite is that the unnatural amino acid must be efficiently transported into the cytoplasm when added to the growth medium, or biosynthesized by the host, and unmodified by endogenous metabolic enzymes [142]. As a matter of fact, ncAAs structurally close to canonical amino acids (i.e.: acetyl-Lysine) may enter cells through endogenous amino acid transporters when added to the growth media, yet those deviating significantly or highly charged (e.g.: phosphorylated ncAAs) may not. In the latter case, the host organism can be engineered by either enhancing endogenous metabolic pathways, or by introducing exogenous pathways[163,164]. In this thesis, ncAAs bearing the most important PTM (PTM-AA) were used[11]. The majority of them were simply added to the growing media, yet for some others either overexpression of bottleneck enzymes or deletion of competing pathways were also required[145,165]. In conclusion, by adding the PTM-AA to the host cells of PISA, providing the genes for the aaRS/tRNA and for the protein target of interest containing an amber stop codon at the desired position, PTM-AAs can be now -translationally and site-specifically incorporated.

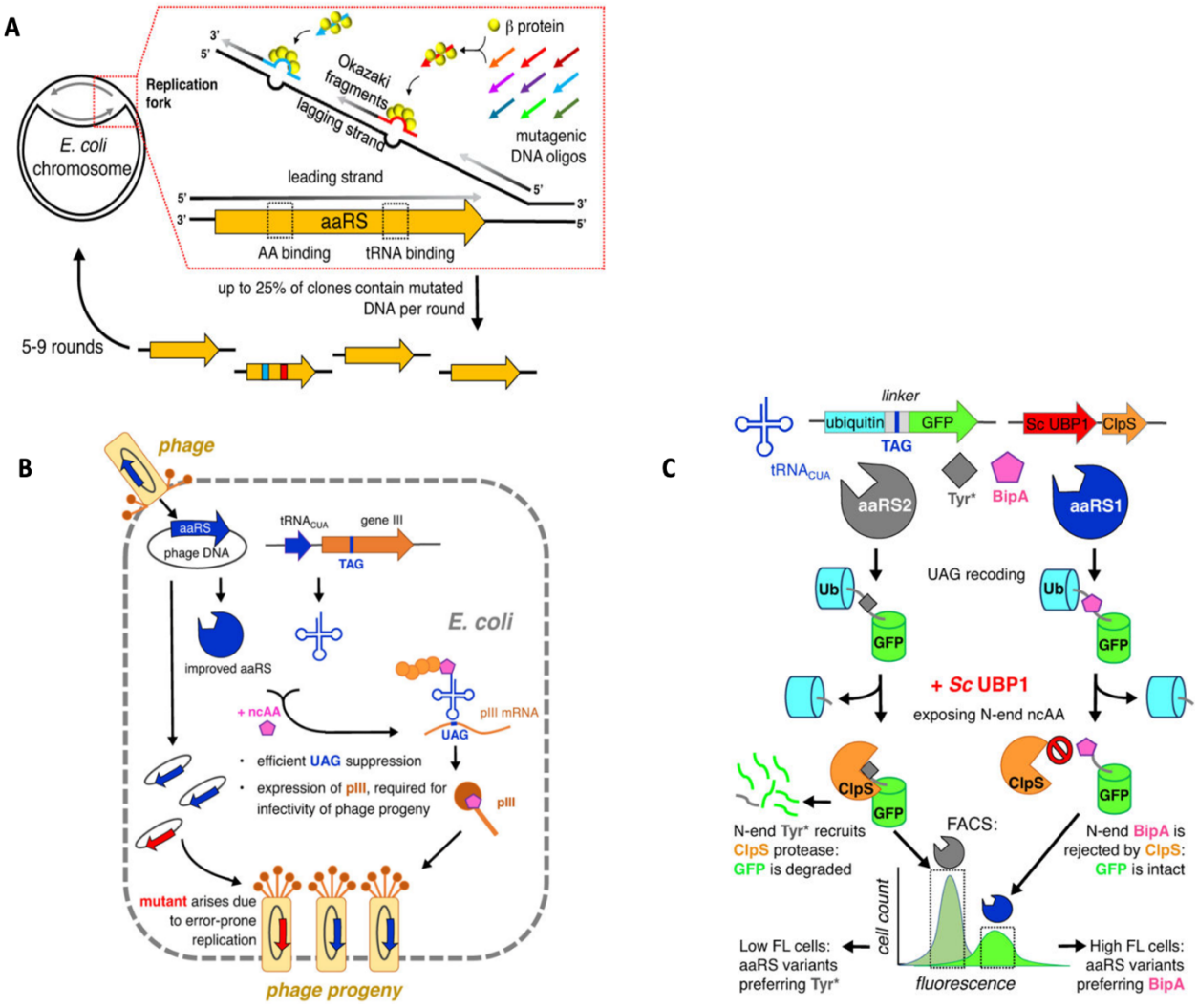


Figure 1.17. **Advanced methodology for aaRS evolution. (A) Multiplex-automated genome engineering (MAGE)** uses a modified λ-Red allelic replacement system to introduce changes in a targeted chromosomal gene (e.g. orthogonal aaRS) by integrating mutagenic oligos. Short regions flanking the randomized sequence of the oligo direct β protein to introduce the primer into the replication fork. After two rounds of duplication, only one of the DNA copies from the depicted replication fork will incorporate mutagenic oligos, resulting in the production of no more than 25% of recombinant clones for every round of MAGE. The desired level of diversity is achieved by repeating MAGE cycles, increasing the size of the oligo library, and the number of targeted sites. **(B) Phage-assisted continuous evolution (PACE)** employs a bacteriophage that lacks the gene for protein III (pIII), which is required for infectivity. A copy of the pIII gene, with at least one TAG codon, is encoded in a plasmid in E. coli. Active aaRS mutants aminoacylate the suppressor tRNA leading to pIII expression and propagation of bacteriophage. **(c) Post-translational proofreading (PTP)** exploits the N-end rule for protein degradation in E. coli. Green fluorescent protein (GFP) is fused to ubiquitin (Ub), separated by a TAG stop codon. Expression of the full-length reporter is the result of aminoacylation of a suppressor tRNA by orthogonal aaRS variants. Ubiquitin cleavase protein 1 (UBP1) from S. cerevisiae cleaves Ub from the reporter, exposing the N-terminus residue of GFP. ClpS surveys the N-terminus of GFP and applies the N-end rule to facilitate degradation of GFP if an undesired AA is incorporated. In contrast, if the desired ncAA is installed at the N-terminus, GFP accumulates and cells with high fluorescence emission are sorted and collected. The system is adaptable as ClpS can be engineered to discriminate against specific ncAAs. (Adapted from Vargas-Rodriguez et al. Curr Opin Chem Biol. 2018.)

1.8.1 Orthogonal aaRS/tRNA pairs used in this work

Several orthogonal translation systems (OTS)[113] have been discovered and developed so far in both prokaryotic[141] and eukaryotic hosts[122,137,162] and the ones used in this work will be described in the following chapter **[Table 1.1].**

Despite the vast majority of effort in developing new OTS took place in *E.coli* [116,141], investigators were also able to artificially alter the genetic code in eukaryotic as well[117,120,137]. Noteworthy, the general strategy described above was also applied to evolve o-aaRSs/tRNA directly in *S.cerevisiae* [125,143,149,162], considering its critical role as a model biological organism. These aaRSs evolved in yeast may also be transferable to other eukaryotic albeit with potential changes in activity [136,142].

Two-thirds of all ncAAs incorporated to date are mediated by the wild-type or engineered *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase *(Mj)*TyrRS/$tRNA^{Tyr}$ pair [150,166,167] or the *Methanosarcina* pyrrolysyl PylRS/$tRNA^{Pyl}$ pair due to their robust orthogonality and the tolerance of anticodon mutations in their tRNAs[152,153]. The *M. jannaschii* TyrRS/$tRNA^{Tyr}$ pair is used in bacteria[141,167], yet this pair is not suitable for eukaryotic applications because identity elements overlap with eukaryotic TyrRS/$tRNA^{Tyr}$[117]. In contrast, the Methanosarcina pyrrolysyl-tRNA synthetase/ tRNA pairs are orthogonal to a much wider range of hosts, (*E. coli*[123,124,168], *Salmonella sp.*[169], *S. cerevisiae*[149], cultured mammalian cells[170], *Arabidopsis thaliana*[168], and *Mus musculus*[171]) including the PISA 2.0 selection host organisms (*S.cerevisiae*[149] and *E.coli*[123,124,168]). In particular I took advantage from the *Methanosarcina barkeri*-derived N6-Acetyllysyl-tRNA synthetase/$tRNA_{CUA}^{Pyl}$ (*Mb*AcKRS3/$tRNA_{CUA}^{Pyl}$) variant[123,124,149], evolved to direct the site-specific incorporation of the crucial Nε-acetyl-L-lysine in *E.coli*[123] and yeast[149] and the *M.mazei* derivative N6-Acetyllysyl-tRNA synthetase/ tRNA (*Mm*AcKRS1/$tRNA_{CUA}^{Pyl}$)[172]. The achievements obtained with these *Methanosarcina* pairs motivated the exploration of PylRSs from other organisms: Chin and coworkers discovered, in fact, a new class of single domain PylRS/ $tRNA^{Pyl}$ pairs from *Methanomassiliicoccales*[173,174] that were shown to be very active and orthogonal in *Escherichia coli* [175]. The aminoacyl-tRNA synthetases of these pairs lack the N-terminal domain[174], so they are smaller in size and more soluble than the *Methanosarcina* PylRSs but share a homologous active site with them; this facilitates the transplantation of mutations discovered with existing PylRS systems into the new PylRS systems to reprogram their substrate specificity[162]. Hence, I decided to include these new OTSs in the study to verify if they indeed increased the ncAAs incorporation also in yeast. I reasoned that, if these synthetases were found to display a significant activity in our model organism, they could be evolved to recognize the ncAA of our interest, using methods previously used [116,120,143].

Especially I employed the Methanomassiliicoccales evolved variants (referred as Me-HisRS-1/ $tRNA_{CUA}^{Pyl\ 13C}$, Me-HisRS-2/ $tRNA_{CUA}^{Pyl\ 6C10}$, Me-HisRS-3/ $tRNA_{CUA}^{Pyl6B03}$), that were used to direct the incorporation of 3-*methyl*-L-*histidine* (Me-His) [unpublished, gifted by Dr.J.Chin Lab) and led to yields of a ncAA-containing protein that were several times higher than that obtained previously.
Another effective OTS turned out to be the evolved derivative of a *Methanococcus maripaludis* phosphoseryl-tRNA synthetase (*Mmp*SepRS/*Mm*$tRNA_{CUA}^{Sep}$) used to incorporate O-Phospho-L-Serine (pSer) and its non-hydrolysable analogue into proteins in E. coli[145,176] and in mammalian cells [136,145]. Accordingly, I did test the engineered *Mmp*SepRS/$tRNA_{CUA}^{Sep}$ activity also in yeast, being phosphoserine of great biological interest.

I am going to briefly describe the development of the SepRS system since it is a perfect example of how Mother Nature has come up with unique solutions to tackle biological problems and how researchers can look at those solutions to help and inspire their work. Phosphoserine is a metabolic precursor to serine, and in certain methanogenic archaea pSer is used in a two-step pathway to direct the incorporation of cysteine into proteins[177]. In the first step of this pathway an unusual aminoacyl-tRNA synthetase (SepRS) aminoacylates phosphoserine on to $tRNA_{GCA}^{Cys}$, to create pSer $tRNA_{GCA}^{Cys}$. A second enzyme then converts p-Ser $tRNA_{GCA}^{Cys}$ to Cys $tRNA_{GCA}^{Cys}$. Suitably, the aminoacylation of $tRNA_{GCA}$ with phosphoserine in the first step raised the possibility of altering the SepRS/$tRNA_{GCA}$ pair to create a SepRS/tRNA(Tran)$_{CUA}$ pair (in which the CUA anticodon is transplanted in place of the native GCA anticodon) for the site-specific incorporation of phosphoserine into proteins [178]. However, since SepRS recognizes the anticodon of $tRNA_{GCA}$, the SepRS/$tRNA_{CUA}$ pair, in which the GCA anticodon was simply substituted by CUA, was a very inefficient amber suppressor [176]. By evolving the sequence surrounding the mutated anticodon, and then the anticodon recognition region of SepRS, scientists were able to select a new orthogonal pair the SepRSv1.0/$tRNA_{CUA}^{v1.0}$, referred to herein as SepRS/$tRNA_{CUA}^{Sep}$ that drives the insertion of p-Serine in response to amber codons 10 times more efficiently than previous pairs [136]. Last, the functional amber suppression system *E. coli tyrosyl*-tRNA synthetase *Ec*TyrRS/ $tRNA_{CUA}^{Tyr}$ [117,179,180]and the wt *Mm*PylRS/$tRNA_{CUA}^{Pyl}$ [123]were provided by Dr. Chin Laboratory and used as positive control for the functional assays in yeast and in bacteria respectively.

Table 1.1. Orthogonal aminoacyl-tRNA synthetase variants applied for Genetic Code Expansion in this work.

Enzyme	Organism	Amino acid	Description[a]	Source
(Mm)SepRS	*Methanococcus maripaludis*	**phospho-Ser**		D.T. Rogerson et al.
PylRS	*Methanosarcina mazei*	**Pyrrol-Lysine**	L301/L305/Y306/L309/N346	S. Hancock et al.
(Mb)AcKRS3	*M.barkeri*	**AcK**	Mutation from (Mb)PylS L266M/L270I/Y271F/L274A/C313F	Umheara et al.
(Mm)AcKRS	*M.mazei*	**AcK**	Mutation from (Mm)PylS L301V/L305E/Y306/ L309A /M382S	This work
(Mb)AcKRS3^{IPYE}	*M.barkeri*	**AcK**	Mutation from (Mb)AcKRS/V31I/T56P/H62Y/A100E	Bryson et al.
(Ma)AcKRS-1	*Methanomethylophilus* alvus	**AcK**	Mutation from (Ma)aaRS L266M/L270I/Y271F/L274A/C313F	This work
(Ma)AcKRS-2	M.alvus	**AcK**	Mutation from (Ma)aaRSL301V/L305E/Y306F/L309A/C348F/M382S	This work
(Ml)MetHisRS	*Methanomassiliicocales* luminyensis	**Met-His**	Mutation from (Ml)PylRS L266M/L270I/Y271F/L274A/C313F	unpublished Dr.Chin lab
(MaRum)MetHisRS	*M. archaeon RumEn M1*	**Met-His**	Mutation from (Ml)PylRS L266M/L270I/Y271F/L274A/C313F	unpublished Dr.Chin lab
(MaBin)MetHISRS	*M. archaeon PtaU1.Bin030*	**Met-His**	Mutation from (Ml)PylRS L266M/L270I/Y271F/L274A/C313F	unpublished Dr.Chin lab

[a] Sequence comparison of o-aaRS variants compared to their WT counterpart.

1.8.2 Methods to Increase the ncAA Incorporation Efficiency

The ncAA incorporation efficiency can approach 75% of the wt protein, but in less favorable cases, efficiencies may be as low as 10% [181,182]. Low incorporation efficiency results in reduced yield of the full length protein containing the unnatural amino acid and in an increase of truncated protein products terminated at the introduced amber codon, thus creating difficulties with employing genetically encoded unnatural amino acids (especially in yeast and mammalian cells) [183]. Several interrelated factors may affect the efficiency of ncAA incorporation in response to an amber codon, including the sequence context of an amber codon and its position in a gene [184], the efficiency of tRNA decoding and peptide bond formation on the ribosome with the amino acids that precede and follow the ncAA in the nascent chain [185–187], and the effects of competition with release factors that terminate protein synthesis in response to stop codons[113]. Many of these confounding effects are not easily separated, and therefore their contributions to ncAA incorporation efficiencies have not been systematically studied or addressed.

However, in order to boost the amber suppression system, research has been reported in which the copy number of orthogonal synthetase/tRNA genes[114,188], as well as the promoters that drive the expression of these components and the ratios in which they are expressed were altered[188–190] **[Figure 1.19]**. In some cases, efforts to increase orthogonal synthetase/tRNA copy number can improve incorporation efficiency at a target codon[191] but may also cause a global increase in read-through of genomically encoded (legitimate) termination codons and, therefore, mis-synthesis of some essential proteins [116,192,193]. Another reason amber suppression is less efficient is that amber-suppressor tRNAs compete with release factors (RFs)[5] [114,194]. Efforts have been focused on strategies to remove the competition with RFs **[Figure 1.19]**. In *E. coli*, an orthogonal ribosome was evolved to no longer efficiently recognize RF1[192]. This approach enabled efficient incorporation of multiple ncAAs into proteins in response to multiple amber codons on an orthogonal message, without enhancing readthrough of stop codons on endogenous messages [192]. RF1 has also been deleted from *E. coli*, facilitating multisite ncAA incorporation[195,196]. Recent work has created mutants of eRF1 that do not efficiently recognize the amber stop codon, and therefore split the degeneracy of stop codon recognition in mammalian cells [197].

[5] In *E. coli*, release factor 1 (RF1) recognizes TAG (amber stop) and RF2 recognizes TGA (opal stop), while both release factors recognize TAA (ochre stop); in eukaryotic cells a single release factor, eRF1, recognizes all three stop codons.

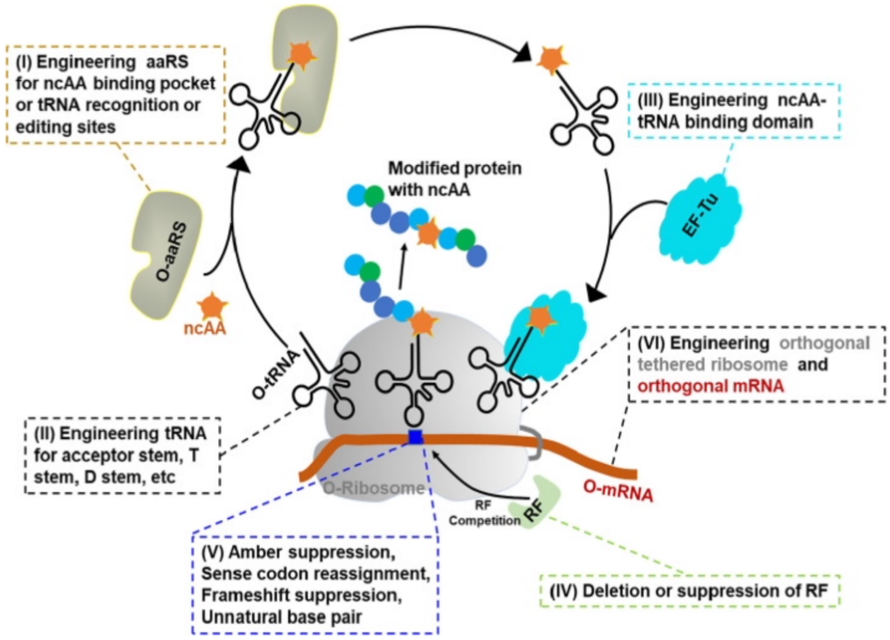


Figure 1.18. **Strategies for efficient ncAA incorporation by comprehensive optimization of OTS.** ncAA incorporation efficiency could be enhanced by means of engineering the orthogonal aaRS (o-aaRS) for ncAA binding pocket, tRNA recognition or editing sites to catalyze ncAA-tRNA formation (I); engineering tRNA elements (acceptor stem, T stem, etc.) (II) and ncAA-tRNA binding domain of EF-Tu for efficiently transport of ncAA-tRNA to the ribosome (III); deletion or suppression of release factor (RF) to avoid competing with the amber suppressor tRNA for the UAG codon (IV). To simultaneously incorporate distinct ncAAs into one protein, blank codons have been generated by reassignment of sense or nonsense codons, introduction of quadruplet codons and unnatural nucleotides(V). An orthogonal ribosome was developed by recognizing a new mRNA (O-mRNA) (VI) and evolved to selectively decode a set of quadruplet codons, enabling the multi-site incorporation of distinct ncAAs.

Extending genetic code expansion beyond bacterial cells and into eukaryotic hosts have furtherly required several new challenges to be addressed[183]. Two important factors for ncAA incorporation in eukaryotes (especially in yeast, the host organism of PISA) are the proper expression of the orthogonal procaryotic tRNAs and the stability of the target mRNA[183,198,199]. Regarding the o-tRNAs, transcription and processing are different in prokaryotes and eukaryotes. A major distinction is that *E. coli* tRNAs are transcribed through promoters upstream of the tRNA gene, whereas eukaryotic tRNAs are transcribed through promoter elements within the tRNA known as the A- and B-box [200]. The A- and B-box identity elements are conserved among eukaryotic tRNAs but are absent in many *E. coli* tRNAs [201,202]. Creating the consensus A- and B-box sequences in *E. coli* tRNAs through mutation could cripple the tRNA [203]. In addition, bacterial tRNA genes encode full tRNA sequences, whereas eukaryotic tRNAs have the 3′-CCA trinucleotide enzymatically added after transcription [203]. Therefore, transplanting bacterial tRNA directly into the tRNA gene cassette in yeast does not generate functional tRNA. Namely, a general method to express

bacterial tRNAs in yeast involves placing an external Pol III promoter containing the consensus eukaryotic A- and B-box sequences upstream of the target bacterial tRNA gene[204] . The 3′-CCA trinucleotide of the tRNA is excluded, and the tRNA(−CCA) gene is followed by a 3′-flanking sequence of an endogenous yeast tRNA[203]. A primary RNA consisting of the promoter and the tRNA is transcribed and the promoter is cleaved post-transcriptionally to yield the mature tRNA [67]. Two yeast Pol III promoters, the RPR1 promoter and the SNR52 promoter, have been shown to efficiently drive the expression of *Escherichia coli* tRNAs in yeast. Alternative method using the yeast $tRNA^{Arg}$ fused upstream of the target tRNA [68] has also been developed. Yeast infact possesses an unusual bicistronic $tDNA_{UCU}^{Arg}$ -$tDNA_{GUC}^{Asp}$ gene in which the two mature tRNAs are generated from a single precursor RNA. The $tDNA_{UCU}^{Arg}$ in fact provides the A and B-box sequences required to transcribe tDNAs inserted in place of $tDNA_{GUC}^{Asp}$. Regarding the stability of the target mRNA, eukaryotic cells have an mRNA surveillance mechanism, nonsense-mediated mRNA decay (NMD), to identify mRNAs containing premature stop codons and target the mRNA for rapid degradation [205]. The amber stop codon is thus far the most frequently used codon to encode unnatural amino acids, but NMD could shorten the target mRNA lifetime, resulting in a lower protein yield. Inactivation of NMD would preserve the stability of the UAG-containing mRNA and thus enhance the incorporation efficiency. An NMD-deficient yeast strain was generated by knocking out the UPF1 gene, an essential component for NMD. The unnatural amino acid incorporation efficiency was indeed increased more than 2-fold in the UPF1Δ strain compared to the wild-type yeast [198,199].

The NMD pathway is active in all eukaryotes and the core factors are highly conserved. However, these similarities notwithstanding, the respective mechanisms of NMD in yeast are different compared to ones of higher eukaryotes. Since it is relevant to this thesis, a brief description of a recently proposed model for NMD in yeast will be provided [and detailed in the **Figure 1.20**] [206,207]. The process begins with the recognition of premature termination codons (PTC) in mRNA transcripts. These PTCs are typically located upstream of the last exon-exon junction, a feature that distinguishes them from normal stop codons. The NMD machinery identifies the presence of these PTCs and marks the mRNA for degradation [206,207]. In yeast, PTC definition can occur independently of a downstream exon boundary, and EJC protein components are either predominantly absent or not required for NMD [208]. Therefore, other determinants must present the downstream signal required for NMD substrate recognition. One possibility is that a stop codon is thought to be recognized as premature in yeast when it precedes a downstream element, or DSE [209]. However, there is also evidence that the distance between the polyA tail and the location of the stop codon may be a factor in premature stop codon recognition [210]. Subsequently, when a PTC is present in an mRNA, the Upf proteins

are recruited to form a complex on the mRNA. This assembly occurs during translation and involves interactions between Upf1, Upf2, and Upf3. Upf1's ATPase and helicase activities are crucial for its role in remodeling the ribonucleoprotein complex.
This complex triggers the recruitment of additional factors, including the decapping enzyme Dcp1/Dcp2, the 5'→3' exonuclease Xrn1, and the cytoplasmic exosome complex, which together degrade the mRNA from both ends. The predominant mode of NMD seems to use deadenylation-independent decapping and Xrn1-mediated 5′-to-3′ decay [205,211]

Wrapping up, despite exciting progress towards engineering cells for encoded ncAAs, many challenges remain [193,212], particularly in yeast. I reckon that further improvement of ncAA incorporation would be important for effective applications, especially in yeast and mammalian cells. In my work I in fact tried to leverage diverse solutions and to follow all the state-of-the art methods to tackle the challenges in reprogramming the PISA 2.0 yeast strain.

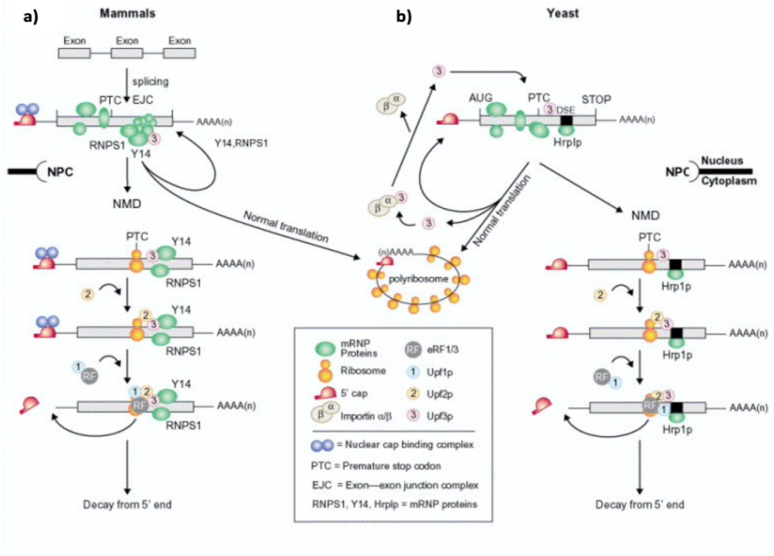


Figure 1.19. Comparison of NMD pathways in mammals and the yeast S. cerevisiae. b) Model to explain how the NMD pathway occurs in S. cerevisiae. NMD occurs during cytoplasmic RNP remodeling. An mRNA is exported to the cytoplasm as an RNP complex with nuclear RNA-binding proteins attached, such as Hrp1. During the initial rounds of translation, the attached nuclear proteins are displaced by the ribosomes and complete transition to a cytoplasmic RNP is achieved. A premature termination codon prevents RNP remodeling and activates NMD. The translating ribosome pauses at a premature termination codon and signals to the eRF1-eRF3 complex to bind to the A site. The Upf1p becomes associated with the eRF1-eRF3 termination/pre-surveillance complex during the termination process. After hydrolysis of the peptidyl-tRNA bond, eRF1 dissociates from the ribosome. Dissociation of eRF1 allows Upf2p (or Upf3p) to bind the eRF3-Upf1p complex. (3) Rearrangement of the complex: Upf2p (or Upf3p) joins the complex and displaces eRF3 to form the mature post-termination surveillance complex. Because the ribosome failed to displace Hrp1 (or other yet unidentified DSE marker factors) from the DSE, the surveillance complex can recognize the DSE marker complex as a signal that RNP remodeling is incomplete. Subsequently, the aberrant transcript is rapidly decapped and the body of the mRNA is degraded by a 5′→3′ exoribonuclease.

1.9 Objectives and Summary of my PhD Thesis Project

My PhD thesis is part of a wider project which aims at addressing the complexity of the Post Translational Proteome by the selection of intracellularly working binders of functional epitopes, being PTMs or oligomeric states of a native protein (the PISA project). My PhD work was mainly involved i) in demonstrating the specific interfering activity of one of our anti PTM-intrabodies (namely the scFv58) compared to currently used indirect methods to study PTMs (the scFv-58 project) and notably ii) in improving the current PISA platform by exploiting Genetic Code Expansion (GCE) genetically encode the desired target PTM, in order to facilitate, accelerate and automatize the isolation of anti-individual PTM intrabodies, specifically anti-Acetyl-Lysine binders (the PISA 2.0 version project in yeast and *E.coli*).

1.9.1 The intrabody scFv-58F Project: the pioneer of a new class of "Antibody Chromatin Readers" molecules

The scFv-58F intrabody selected through P.I.S.A technology was shown to specifically silence a post-translationally modified version of the histone H3, through the direct binding of the lysine 9 residue (K9) in the nucleus of mammalian cells [32].

The aim of this part of the project was to characterize the silencing activity of this intrabody, through the side-by-side comparison between *its direct H3K9Ac-binding* and *the indirect* and broader approach of histone modifier enzyme inhibition. This objective was pursued in order to provide a direct demonstration, in a real-world application, of the theoretical prediction that the direct competition with the PTM epitope on a target protein yields more specific effects than inhibiting the enzyme responsible for installing that same PTM[51]. This prediction is reasonable, given the pleiotropy of PTM installing enzymes, each of which has manifold substrates, but needs to be formally tested.
As anticipated above, developing compounds that can selectively target PTM-proteins, as opposed to inhibiting the modifying enzymes, represents a 'holy grail' of drug development in different fields, including epigenetic therapeutics [112]. Modulation of the chromatin processes involving histone PTMs is mediated by chromatin modifying enzymes[20]. Due to their activity

in physiology and pathology, several chemical compounds have been developed to inhibit the function of these proteins[25–27,29]; some of which are undergoing clinical trials [25,26]. However, the pleiotropy of these classes of proteins represents an intrinsic weakness of epigenetic drugs [21]. Each such enzyme modifies multiple residues on different proteins[30]. In order to study the effects of *a single epigenetic PTM on cell function* [16,213] and to obtain a more focused therapeutic outcome, it would be important to interfere precisely with single acetylated residue.

Thus, the proposed strategy of directly targeting single epigenetic marks with PTM-specific intrabodies is predicted to achieve a superior specificity and selectivity with respect to the current approaches that target instead the epigenetic modifying enzymes responsible for installing the PTM **[Figure 1.20]**. However, since a direct comparison between the two strategies was lacking, the objective was to experimentally test this hypothesis and prediction **[Figure 1.21].**

- First, using Next Generation RNA sequencing we characterized the transcriptomic consequences of intrabody scFv-58F expression in yeast cells, in comparison to yeast cells exposed to two broadly used HAT inhibitors: CPTH2, a synthetic inhibitor of the Gcn5 HAT, responsible for H3K9 acetylation, and Curcumin, a natural compound, reported to inhibit p300/CBP HAT activity.
- After that, the differential gene expression analysis was performed allowing for the identification of a group involved in specific biological function as target for the transcriptional regulation by H3K9. These results were also independently confirmed throughout an exhaustive comparison with recent findings in literature.

The general implication of this comparative study is that the use of anti-PTM intrabodies for a PTM-specific interference in living cells will provide a new level of precision and specificity in the description of epigenetic, offering new tools for research. We envisioned the scFv-58F as the pioneer of a new class of *Antibody Chromatin Reader molecules*.

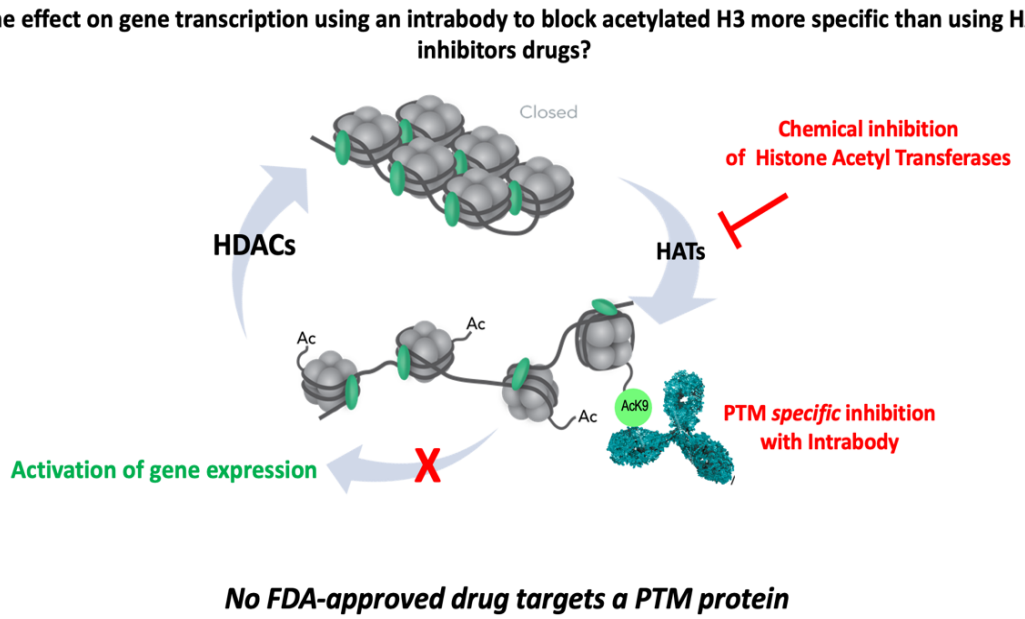


Figure 1.20. The scFv-58F Project. Inhibition of protein acetylation at different levels using existing chemicals vs anti-PTM intracellular antibody domains. The scFv-58F Project aims at testing whether the silencing effect on gene transcription using an intrabody to block acetylated H3K9ac would be more specific than using HAT inhibitors drugs. Using a HAT inhibitor to study protein acetylation results, in fact, in the simultaneous inhibition of many downstream proteins. An intrabody-based approach would be a valuable strategy to achieve a very selective targeting and inhibition of a desired PTM-modified proteins, including Histones and TFs that are involved in chromatin remodeling and transcription. The high specificity of intracellular antibodies would even permit to distinguish between different acetylated variants of the same protein target.

1.9.2 The PISA 2.0 project: Expanded Genetic Code technologies to genetically encode individual PTMs in the protein target.

Targeting the post translational proteome implies the possibility of dissecting the individual contribution of each individual PTM[1,11]. *Tethered catalysis-based* PISA screening can provide only a limited solution to this problem, because i) the PTM-installing enzyme must be unambiguously known, ii) it must work in a tethered catalysis cis-fusion format, iii) chimeric baits are in principle modified at all (many of) the possible sites [105]. By exploiting the *Genetic Code Expansion* (GCE) technology, the systematic isolation of anti-individual-PTMs intrabodies could be greatly accelerated: *GCE in fact would in principle allow to genetically encode the PTM-AA in a desired position in the protein target.* This would provide the ultimate solution to hugely accelerate the selection of intrabodies against a target PTM.

The second part of my work was therefore aimed at implementing the current PISA platform exploiting the Genetic Code Expansion technology [113,115] (instead of tethered catalysis), developing *the PISA 2.0 version* **[Figure 1.21].** Notably, GCE technology provides the orthogonal aminoacyl-tRNA synthetase/tRNA (tRNA/aaRS) pairs to incorporate PT-modified unnatural amino acids into proteins, when an amber codon is encountered in their sequence during translation, in living cells (including *Saccharomyces cerevisiae or E.coli*) [116,124,138,149,156]. So, adapting this GCE methodology to the PISA selection system, several strains can be created by integration of different aaRS/tRNA pairs in the yeast genome. Each of them thus will insert a particular PT-modified amino acid (i.e Acetyl-Lysine, and Phospho-serine) in proteins of interest, stably expressed and ready to be screened against our antibody domain libraries **[Figure 1.22].**

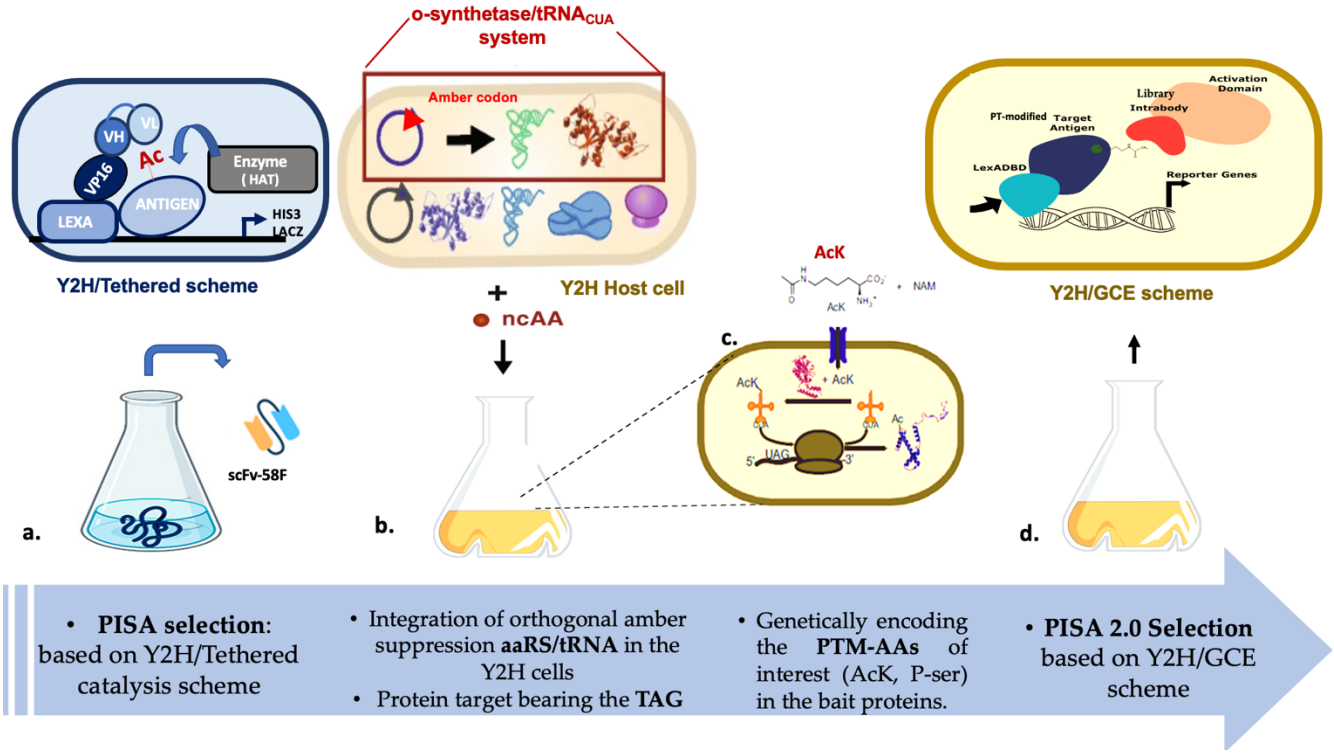


Figure 1.21. Scheme of The PISA 2.0 Project. (a) An upgrade of the PISA technology (referred as The PISA 2.0 version) foresees the substitution of tethered catalysis with expanded genetic code technologies to encode the PTM genetically in the target protein. The PISA 2.0 Project involves: ***(b)*** the creation of yeast stable lines stably expressing orthogonal synthetase/tRNA pairs evolved to incorporate the PTM-AA of interest (namely Phospho-serine and Acetyl-Lysine) in response to the amber codon in the target proteins; **(c)** the set-up of the protocol for installing PTM-UAAs in the PISA 2.0 baits, carrying an amber stop codon in correspondence with those amino acids commonly modified by modifying enzymes; and **(d)** performing the PISA 2.0 selection in the given two-hybrid "acetylator" or "phosphorylator" strain able to produce the bait harboring the PTM in a desired position ready to be used to select intrabody in the standard PISA protocols.
Legend. Y2H= Yeast Two Hybrid; aaRS/tRNA= aminoacyl-tRNA synthetase and tRNA pair; ncAA= non canonical amino acid; PTM-AA: Post-Translationally Modified Amino Acid acids. AcK = Acetyl-Lysine; GCE= Genetic Code Expansion.

- Initially, I provided a proof of concept of the method adapting the GCE methodology to the yeast-two hybrid platform ***(The PISA 2.0 Project in yeast).*** First, an acetylator yeast line was produced via the integration of the *Methanosarcina barkeri* aaRS/tRNA pair (*Mb*AcKRS3/tRNA)[149]. This pair enables the incorporation of N-acetyl-Lysine (AcK) into the amber stop codon (TAG) in proteins.

- After having evaluated that all the components required for genetically encoding AcK were stably integrated and produced in the cell, the functionality and orthogonality of the AcKRS/tRNA pair was validated *by encoding the ncAA into the bait protein of interest.*

- Additionally, before moving onto the PISA 2.0 main selection, I focused my effort on systematically stepwise *enhancing the amber suppression system efficiency* in order to increase the AcK incorporation in the given acetylator yeast [183,198,199]. Accordingly, I attempted new methods i) to improve the expression efficiency of the orthogonal prokaryotic tRNA by using special Pol III promoters and ii) to improve the target mRNA stability (containing the UAG) in yeast via the inactivation of the nonsense-mediated mRNA decay (NMD) pathway [205].

- Remarkably, I also sought to enhance the catalytic activity of the AcKRS: a mutant library of the synthetase was made using the error prone PCR method. However, this experiment suggested a suboptimal activity of the AcKRS/tRNA pair at least in our strain.
- Accordingly, other aaRS/tRNA couples were analyzed, in order to find the best working pair to implement PISA 2.0. Noteworthy, the aaRS/tRNA to encode Phospho-serine was also studied. For this comparative study a highly sensitive fluorescence assay and a cutting-edge approach called tRNA extension (tREX) were employed [214].

At the end of this step-by-step systematic analysis, the collected results highlighted a poor incorporation efficiency in *S.cerevisiae.* This part of the work provides nonetheless the most systematic and robust analysis of the various steps involved in GCE methods in yeast cells. Furthermore, the strains derived turned out as a platform that can be exploited by researchers in the field.
As a way forward, towards the development of a PISA 2.0 (GCE-based) selection approach, I decided to move from the yeast to the *E.coli* system, taking advantage of well an established bacterial-two-hybrid (B2H) system [215,216] as a intrabody selection platform. The *E.coli* system would allow exploiting the wide range of functional and orthogonal aaRS/tRNA pairs available, to set up the new 2.0 version of our PISA **(The PISA 2.0 Project *in E.coli*).** In this embodiment, the intrabody selection will take place in the bacterial two- hybrid strain in which any orthogonal tRNA/aminoacyl-tRNA synthetase pair can be readily inserted to genetically encode PTMs in the protein bait of interest. Accordingly, during the last part of my thesis work, I accomplished the preliminary steps to adapt the PISA 2.0 project to bacteria.

- Initially, I performed a trial B2H selection against a protein target of interest to test the system in our laboratory setting. Some adjustments to ameliorate the original protocol were also made.
- At the same time, I implemented two orthogonal tRNA/synthetase pairs - one for the insertion of N-ε-Acetyl-L-Lysine [106] and the other for the insertion of O-Phospho-L-Serine [146] - in this bacterial two hybrid system. The orthogonality of the AcKRS/tRNA pair was evaluated using a fluorescence assay, the OTS to encode phospho-serine has not been tested yet since further engineering is required.

In conclusion, regarding the overall project, a given "acetylator" or "phosphorylator" bacterial two-hybrid strains will be available for use with standard PISA protocols. This will allow to accelerate and automatized the number of selections performed, generating a pipeline of anti-PTM binding domains for target validation, functional studies and potentially therapeutic purposes.

2.The scFv-58F Project

2 The scFv-58F Project[6]:

Post-translationally modified proteins represent a huge and untapped source of biologically important and disease-relevant targets, but their systematic biological elucidation and validation for therapeutic purposes is hampered by the lack of specific tools and experimental strategies. In particular, histone PTMs (mainly Lysine acetylation) play a pivotal role in the epigenetic regulation of gene expression [5,6], mediated by chromatin modifying enzymes [217]. Due to their relevant activity in physiology and pathology, several chemical compounds have been developed to inhibit the function of these proteins, also in a therapeutic perspective [20,218,219]. However, the pleiotropy and multispecificity of these classes of proteins represent an intrinsic weakness of current epigenetic drugs [21] **[Figure 2.1]**.

Ideally, rather than inhibiting the PTM-installing enzymes (e.g a Histone Acetyl-T), that act on several different targets, one would need to target a single PTM site on a given protein per se [51]. In our lab, we have recently shown the feasibility of an experimental strategy (the PISA Technology) of directly targeting single epigenetic marks with PTM-specific intrabodies [32]. This is predicted to achieve a superior specificity and selectivity with respect to the currently used inhibitors, but a direct comparison between the two strategies is lacking. In this part of my thesis project, we tested this prediction by exploiting the specific anti-H3K9ac scFv-58F, recently selected by PISA technology and shown to induce a functional in-cell interference of the acetylated H3, by selectively binding to acetylated histone H3 in cells [32,112]. For the first time, we compare the effects on gene transcription induced by the specific anti-H3K9ac scFv-58F to those achieved by to two well characterized Histone Acetyl-Transferases inhibitors (HATis), Curcumin [220,221] and CPTH2 [222,223].

[6] Most of the experiments and results featured in the following chapters were published on August 2022 on the journal "Int. *J. Mol. Sci.* "(Lisi et al. 2022). Some parts of the Results and Material & Methods are less detailed compared to the one of "the PISA 2.0 Project" because they referred to preliminary results done by the lab. The data presented in this study are openly available in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number **GSE194577** (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE195547).

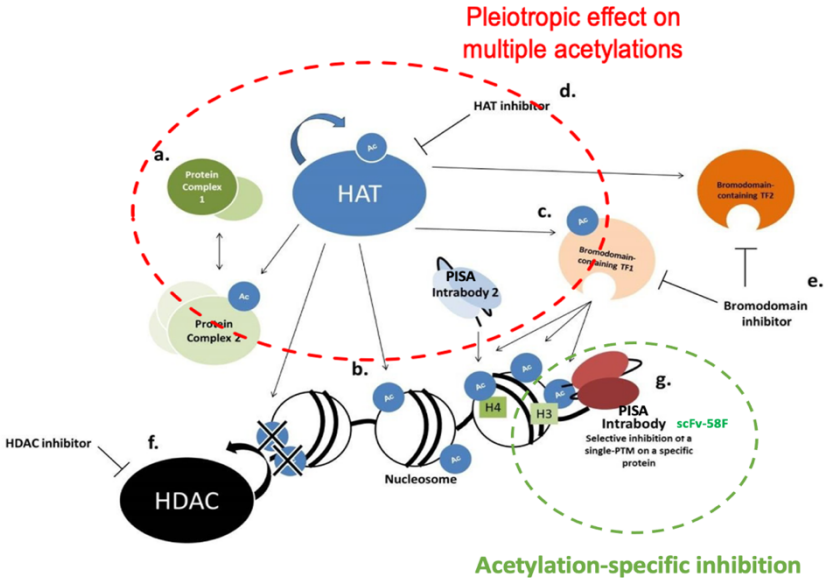


Figure 2.1. The scFv-58F Project. Inhibition of protein acetylation at different levels using existing chemicals vs anti-PTM intracellular antibody domains. The scFv-58F Project aims at testing whether the silencing effect on gene transcription using an intrabody to block acetylated H3K9ac would be more specific than using HAT inhibitors drugs. HAT enzymes are able to acetylate a vast multitude of intracellular targets (a), including Histones (b) and Bromodomain-containing Transcription Factors (TFs)(c). Using a HAT inhibitor to study protein acetylation results therefore in the simultaneous inhibition of many downstream proteins (d). Inhibiting these domains means interfering with different proteins involved in very diverse processes (e). The same problems are valid for Histone Deacetylase enzymes (HDACs), which are able to remove acetyl-groups from multiple targets (f). An intrabody-based approach would be a valuable strategy to achieve a very selective targeting and inhibition of a desired PTM-modified proteins. The high specificity of intracellular antibodies would even permit to distinguish between different acetylated variants of the same protein target (g).

2.1 The scFv-58F Project: Material and Methods

2.1.1 Plasmids and constructs

The coding sequence of the scFv-58F intrabody [32] with the HA tag at its C-terminal was cloned in the yeast expression vector pLinker220 (pL220)[95] .An empty pL220 plasmid with the HA tag was prepared as control plasmid as follows. Two complementary primers with the HA sequence with the *BssHII* and *BamHI* sites at the 5'and 3'-terminal, respectively (Forw: 5' AGCCGAGCGCGCATTACCCTTATGATGTGCCAGATTATGCTTGAGGATCCCCGGG 3'; Rev: 5'-CCCGGGGATCCTCAAGCATAATCTGGCACATCATAAGGGTAA-TGCGCGCTCGGCT- 3') were *in vitro* annealed (95°C for 5 minutes, followed by a temperature gradient from 95°C to 25°C with a 2°C decrease every minute). The annealed product was *BssHII* and *BamHI* digested and cloned in pL220 plasmid. A second control plasmid was obtained by cloning the coding sequence of an unrelated intrabody (named scFv-645 anti *Mus musculus* Neuroligin 2) with the HA tag fused at its C-terminal, in pL220.
The scFv-58F was also subcloned in the mammalian cell expression Vector pEF1α-IRES-ZsGreen1 (Clontech 631976) inserting a *BssHII* restriction site at the 5' prime end and an HA tag at the 3' prime end (for cytoplasmic expression (58Fcyto) or three nuclear localization signals (NLS) before the HA tag at the 3' prime end for nuclear expression (58FNLS). The scFv-645 was also cloned in the same vector with the three nuclear localization signals (NLS) before the HA tag at the 3' prime.

2.1.2 Yeast strains

All the experiments have been performed using the *S. cerevisiae* L40 yeast strain characterized by this genotype: mat-a hisΔ200trp1-901 leu2-3,112 ade2 LYS2::(lexAop)4-HIS3 URA3::(lexAop)8-lacZ Gal4 [224].
The pL220 plasmids carrying the scFv58F-HA, the scFv645-HA or the HA tag only, were transformed in L40 yeast strain following Lithium Acetate transformation protocol [225](cfr. The PISA 2.0 Project in yeast).

2.1.3 Drug Treatments: *Trypan blue* cell viability test

Curcumin (Sigma #78246) and CPTH2 (Sigma #C9873) were solubilized in DMSO and used as HAT inhibitors. Cell viability was assessed in pL220-HA yeast strain by growing the cell in

liquid media with increasing concentration (0-800µM) of the two drugs for 4 or 8 hours by counting viable cells with *trypan blue* dye exclusion test. The *trypan blue* dye exclusion test allows counting viable cells with the Burker chamber at the optical microscope **[Figure 2.2].** Generally, 20-50 µl of growth cultures were used and after proper dilution for feasible count, the final dilution has to be 1:1 with *trypan blue* dye. In order to estimate the number of total cells for each concentration the following formula has been applied:

$$Num.\ cells = \frac{TOT\ count}{Num.\ counted\ quadrants}\ x\ 10000\ x\ Diluition\ factor$$

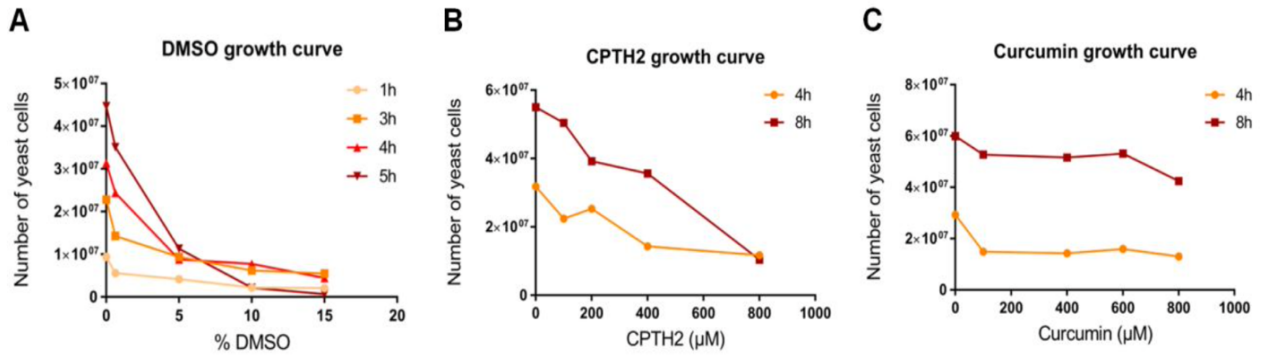


Figure 2.2. Yeast strain pL220-HA growth curves in the presence of different concentrations of DMSO, CPTH2 or Curcumin. For DMSO growth curve (A), the effect on yeast viability was assessed by measuring the absorbance at 600 nm at different time points. The number of yeast cells was estimated considering the following conversion factor: OD (600 nm)=1.0 corresponding to $1.4x10^7$ cells. For CPTH2 (B) and Curcumin (C) growth curves, viability was assessed via trypan blue staining after 4 and 8h of growth

2.1.4 Drug treatment: Yeast sample preparation

Curcumin was purchased from Sigma-Aldrich (#78246), purity ≥99.5%. Storage temperature of 2-8°C.
Its molecular weight is 368,38 g/mol and 19,5 mg were used for the preparation of 1 ml stock solution 50 mM in 100% DMSO.
Starting from the stock solution previously prepared, several concentrations were tested in order to evaluate the solubility of Curcumin and the range between 100 and 800 µM shows a good solubility in DMSO.

Purchased from Sigma-Aldrich (#C9873) purity ≥98%. Storage temperature of 2-8°C.
The molecular weight of CPTH2 is 298,1 g/mol and 1 ml stock solution 17,1 mM in DMSO has been prepared.

From this stock solution, several dilutions were prepared in order to assess the solubility of CPTH2 and the tested concentrations of 800, 400, 200 and 100 μM show good solubility in DMSO.

Five experimental conditions have been considered: two drug-treatment conditions (Curcumin or CPTH2), two intrabody-expressing conditions (scFv-58F or scFv-645) and one control condition (yeast strain carrying the empty vector). For each condition, three biological replicates have been considered. All the yeast cells have been treated with the same concentration of 0.4% DMSO. Treatment with Curcumin (200 μM), CPTH2 (400 μM) or control DMSO (0.4%) was performed using the pL220-HA yeast strain, incubated in selective liquid medium for 1 hour in shaker (250 rpm) at 30°C. After the treatment in shaker a volume of 1,5 ml (in double for each sample) was collected and centrifuged and cells used for either RNA or protein extraction.

2.1.5 Next Generation Sequencing Study

RNA extraction. An aliquot of the same samples, of the five experimental conditions described above, was used for total RNA isolation. Briefly, after the one hour of growth, the yeast cultures were centrifuged (5 min at 3000 rpm) and the RNA extraction was performed using the Yeast Ribopure® kit (Ambion, Invitrogen 500 #AM1926), following manufacturer instructions. This kit combines disruption of yeast cells with Zirconia Beads, phenol extraction of the lysate, and glass-fiber filter purification of the RNA. The integrity of RNA was assessed by agarose electrophoresis (data not shown). The fifteen RNA samples have been delivered to the Leibniz Institute on Aging (FLI) 503 in Jena where library preparation and sequencing steps have been performed.

RNA sequencing. Sequencing of RNA samples was performed using Illumina's next-generation sequencing methodology [60]. In detail, total RNA was quantified using Agilent 2100 Bioanalyzer Instrument (Agilent RNA 6000 Pico). Libraries were prepared from 1000 ng of input material using TruSeq Stranded mRNA (manufacturer's instructions) and subsequently quantified and quality checked using Agilent 2100 Bioanalyzer Instrument (DNA 7500 kit). Libraries were pooled and sequenced in one lane of HiSeq 2500 System running in 51 cycle/single-end/high output mode. Sequence information was converted to FASTQ format using bcl2fastq v1.8.4

RNA-seq data analysis. Data were released as FASTQ files and then processed using 513 Linux-based software tools. An initial quality check of the raw sequence data was performed using *FASTQC.* Trimming of the 3' adapters was performed with the *cutadapt* tool. Using the 3rd version (April 2011) *S. cerevisiae* genome as reference genome, the mapping procedure was

carried out with the *segemehl* software. The number of mapped reads within the coding regions of each gene was calculated with the *bamutils* software tool (NGSutils suite). The raw counts statistical analysis (i.e., PCA and hierarchical clustering), the differential expression analysis with the *DESeq2* package, as well as the subsequent statistical analysis (i.e., MA plots and heatmaps), were all performed in R. Gene Ontology enrichment analysis was performed using the ClusterProfiler package in R.

2.1.6 Functional assay

Retrotranscription and Real-time qPCR. After an initial treatment with DNAse (1h at 37°C), 400 ng of RNA for each yeast sample have been retrotranscribed using the AMV-RT (Promega #5101) and oligo-(dT) primers. The following TaqMan assays have been used: TUB1 (#Sc04175846_s1), URA7 (#Sc04099112_s1), MAK16 (#Sc04097538_s1), SPG1 526 (#Sc04127525_s1) and SNZ1 (#Sc04154021_s1).
Real-time PCR reactions were run in 96-well plates (Bio-Rad), using the Step-OnePlus real-time PCR system (Applied Biosystem) following this protocol: 2 min at 50°C, 10 min at 95°C (Taq-pol activation) followed by 40 cycles with 15 sec at 95°C (denaturation) and 1 min at 60°C (annealing and extension). Quantitative values for cDNA amplification were calculated from the threshold cycle number (Ct) obtained during the exponential growth of the PCR products. Threshold was set automatically by the Step onesoftware. Data were analyzed by the ΔΔCt methods using Tubulin (TUB1 gene) for yeast RNA samples or beta Actin (ACTB gene) to normalize the cDNA levels of the transcripts under investigation.

2.1.7 In vitro assay

TCA protein extraction protocol. The protein extraction on yeast samples treated with various drugs/DMSO concentrations was performed using the TCA protocol.
Briefly, 1.5 ml of cell cultures were pelleted and resuspended in 50 µl of lysis buffer (1.85 N NaOH, 7.4% β-mercaptoethanol, 1X Protease Inhibitors Cocktail (Roche #11836170001)) for 10 minutes at RT. 50 µl of cold 50% Tri-Chloro Acetic Acid (TCA) were added and incubated in ice for 10 minutes. Samples were centrifuged at 14000 rpm for 10 minutes at 4°C. The pellets were washed with 110 µl of cold 90% Acetone incubating at -20°C for 20 min. Samples were centrifuged at 14000 rpm for 10 min at 4°C, the pellets resuspended in 40 µl of Sol. A (0.5M Tris base, 5% SDS) briefly sonicated, and 40 µl of Sol B (75% glycerol, 1.92% DTT, 0.05%

bromophenol blue) were added. Samples were incubated at 95°C for 10 min, spun and 10 µl loaded on acrylamide gel to perform SDS-PAGE.

Western blot assay. For Western Blot analysis the following antibodies were used. Total H3 or acetylated H3 were detected using anti-Histone H3 Rabbit Polyclonal Antibody (1:1000; Abcam #1791) and anti-acetyl-Histone H3 Rabbit Polyclonal Antibody (1:10000; EMD Millipore #06-599), the HA tag using anti-HA (Roche # 11867423001). For loading control of yeast protein samples anti-PGK Mouse monoclonal (1:10000; Abcam #22C5D8) was used. As secondary antibodies, Anti-Rabbit-HRP (1:2000; Santa Cruz Biotechnologies #sc2004), anti-Mouse-HRP (1:5000; Santa Cruz Biotechnologies #sc-2005), anti-Rat-HRP (1:2000; Santa CruzBiotechnologies #sc-2006) were used. Chemiluminescence was acquired through a Chemidoc XRS instrument.

2.2 The scFv-58F Project: Results

2.2.1 The transcriptional effects of the anti-H3K9ac scFv-58F intrabody in comparison to those induced by HAT inhibitors in yeast cells

In this part of the thesis, I will report the results related to the scFv-58F Project in which we investigated the global gene expression changes driven by the intrabody anti-H3K9ac scFv-58F and compared them with those induced by the inhibition of the corresponding modifying enzymes. Since Histone acetylation on lysine 9 (K9) is known to strongly regulate chromatin and transcription [226] we chose site-specific histone H3K9 acetylation as a target for this comparison.

2.2.2 Choice of HAT inhibitors

To begin with, we started determining the transcriptomic response in yeast cells either expressing the scFv-58F or treated with HAT inhibitors [25] before moving onto the mammalian system. As extensively described [cfr. Chpt. 1.5] when expressed in cells, the anti-H3K9ac scFv-58F specifically and directly binds acetylated K9 on histone H3 [32]. For this comparative study, in order to obtain proper conditions of HAT inhibitions, two chemical HAT inhibitors (HATis) were chosen: CPTH2, a synthetic inhibitor of the Gcn5 HAT [222], responsible for H3K9 acetylation [223], and Curcumin, a natural compound, undergoing clinical trials [220], reported to inhibit p300/CBP[7] HAT activity, but for which also non-HAT targeted activities are reported [221]. The preliminary drug validation tests done by the lab revealed that both Curcumin and CPTH2 were solubilized in DMS0 0,4% (a concentration compatible with yeast cell survival) and that just after one hour of yeast growth, the inhibition of HATs with those compounds was achieved **[Figure 2.3 A-C].**

Three yeast stable strains available in the lab were used: the pL200-58F-HA strain, stably expressing the scFv-58F-HA embedded in the Y2H vector (pL220), the pL200-645-HA strain expressing an unrelated intrabody scFv-645-HA (an anti neuroligin-2 intrabody) and a pL220-HA strain, carrying the empty vector pL220. The latter represented the baseline condition in our experimental design. Also, the introduction of a second intrabody with different binding

[7] P300 and CBP are coactivator proteins which interact with numerous transcription factors and act to increase the expression of their target gene. p300 and CBP each contain a histone acetyltransferase domain and a bromodomain that binds acetylated lysines and a PHD finger motif with unknown function.

specificities in this study is due to the fact that just the presence of an antibody expressed in the intracellular environment may entail an effect on mRNA expression in yeast cells, that is independent from its binding ability. This could have an impact on the transcriptome analysis and thus, it is essential to subtract this effect from that due to the presence of the scFv-58F intrabody.

The inhibition of histone H3 acetylation was performed by treating the control yeast strain pL220-HA with either Curcumin (200 μM) or CPTH2 (400 μM) in liquid media for 1 hour**.** The baseline control was represented by pL220-HA cells treated with the same concentration of DMSO (0.4%) used to dissolve the drugs. Then the condition of HAT inhibition was confirmed principally considering the effects on the global pattern of histone acetylation in yeast protein samples by WB analysis **[Figure 2.3 C].**

In this series of western blot, we observed the strong effect of the two drugs on bulk histone H3 acetylation level (using an "anti-H3Ac Pan" primary antibody probe) after one hour of growth in comparison to the null effect detectable in the presence of DMSO only **[Figure 2.3].**

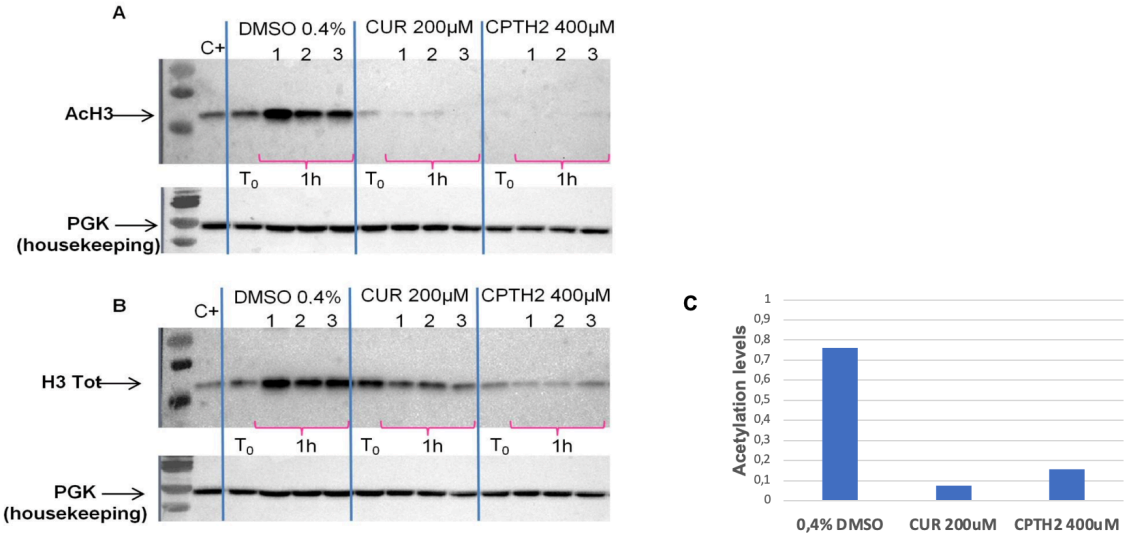


Figure 2.3. Western blot of H3 acetylation after HATi inhibition. A) Acetylated H3 (top panel) and the housekeeping Phosphoglycerate kinase (PGK) (bottom panel) and **B)** total H3 (top panel) and the housekeeping PGK (bottom panel) in protein extract from the yeast strain pL220-HA treated with Curcumin (200 μM), CPTH2 (400 μM) or DMSO (0.4%) for 1 hour at 30°C. Three biological replicates (1-2-3) are shown for each condition. PGK was used as loading control. **C) The histogram represents the histone acetylation level as a function of treatment type.** The acetylation level is expressed as the ratio between the densitometric quantification of αH3Ac/αH3 for each concentration of CPTH2, Curcumin or DMSO (each band was normalized on its correspondent housekeeping signal). For each treatment, the displayed value was obtained by averaging the three measurements. No confidence interval is displayed due to the error propagation of the normalization process. Data should be considered only as a general trend.

2.2.3 NGS data analysis:

For the RNA-sequencing (RNA-seq) analysis, total RNA was extracted from pL220-HA grown either in presence of Curcumin or CPTH2, as well as from pL200-HA, pL200-58F-HA or pL220-645-HA (the two intrabody-expressing strains), and the samples of poly-A RNA underwent sequencing on Illumina *HiSeq* platform. The released data were then processed using Linux-based software tools compared either the two HATi-treated or the two intrabody-expressing conditions against the DMSO-treated baseline control condition. In particular, differential gene expression (DEG) analysis was achieved using the DESeq2 package, the raw counts statistical analysis (Hierarchical Clustering and Principal Component Analysis (PCA)) was performed using R and the script employed will not be reported due to their length and complexity.

2.2.4 Principal Component Analysis (PCA) and Hierarchical Clustering

To make the data easy to explore and visualize we used Principal Component Analysis (PCA). From the first plot **[Figure 2.4A]** comprising all the samples, it stands out that Curcumin samples account for most of the variation, while the second principal component is represented by CPTH2 samples. Despite the large distance of Curcumin samples from the others which determines a crushing of the plot, the projections on the CPTH2-Curcumin straight line display that 645, 58F and DMSO samples are closer among them (with the second a little bit further away from DMSO). In order to better visualize the distances between the other components, Curcumin samples have been omitted in the second PCA plot **[Figure 2.4B].** As expected, among the remaining samples, the principal component of variation is accounted for by CPTH2 samples, while the second component is accounted for by DMSO samples. It is easier to visualize how the distance between 58F and DMSO samples is greater than that between the latter and 645 samples and how the 58F group appears to be in the direction of CPTH2 samples. Moreover, looking at these plots it emerges that the three biological replicates are robust among them in all five the experimental conditions.

We then used the Hierarchical Clustering statistical analysis to visualize the relationships among different samples which are clustered on the basis of the distance between them. Briefly, this graph can be read following the color-code from dark blue (for “exactly the same”, on the diagonal), to white (for “severely different”). As it can be seen, DMSO, 645 and 58F define a clear group, indicating again the proximity among these three treatments. However, as expected, 645 and 58F define a smaller sub-group comprised in the previous, highlighting a non-negligible distance from the DMSO treatment. On the other hand, the two drugs employed show a larger distance from the DMSO control treatment but also among them. Notably,

Curcumin appears to have very little in common with the other treatments as it is clustered in a different branch of the dendrogram **[Figure 2.4C].**

Altogether, Hierarchical Clustering and PCA of the RNA-seq data showed that the Curcumin and CPTH2 treated samples account for most of the variation **[Figure 2.4A and Figure 2.4B],** with the two intrabody-expressing conditions being closest to the DMSO-treated control samples **[Figure 2.4A-C].** Despite being close to DMSO-treated controls, the scFv-58F and scFv-645 samples formed two clearly separate clusters, indicating a different impact on global gene expression induced by the expression of the two intrabodies.

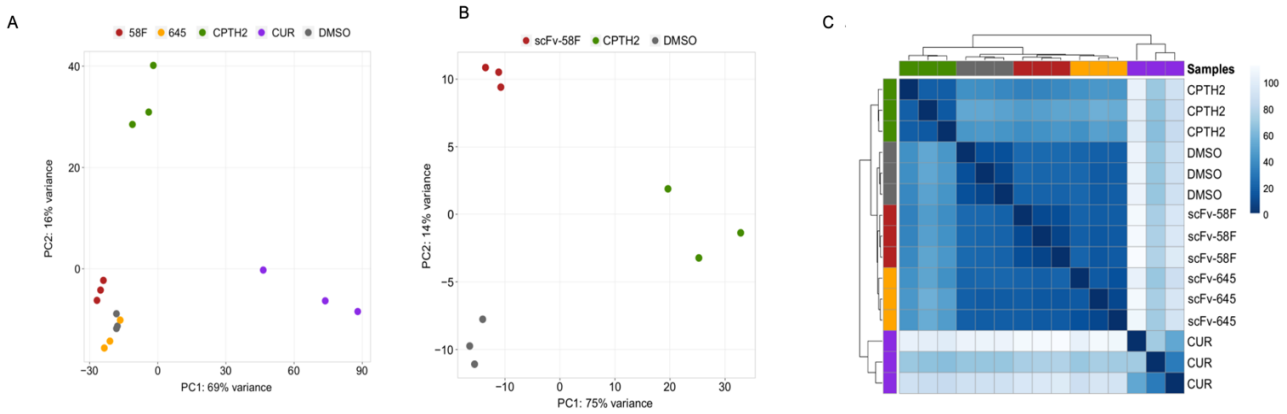


Figure 2.4. PCA plot and Hierarchical Clustering: CPTH2 and Curcumin treatments account for most of the variation in gene expression. (A) Principal Component Analysis (PCA) of the RNAseq data from three experimental conditions 58F, CPTH2 and DMSO (negative control). For clarity, Curcumin-treated samples (which account for most of the variation) and the scFv-645 samples were omitted in this PCA plot and are reported in **(B).** Among the remaining nine samples, the principal component of variation is accounted for by CPTH2 samples (75% of variance explained by first principal component, i.e., PC1). **(C)** Hierarchical clustering of RNAseq data from three biological replicates for each analyzed condition in yeast cells (CUR= curcumin-treated cells, CPTH2= CPTH2-treated cells, DMSO= DMSO-treated cells, 58F= cells expressing scFv-58F, 645= cells expressing scFv-645)

2.2.5 DESeq2 analysis

The results obtained with the DESeq2 analysis were then used to realize MA plots for each of the four "treatment-DMSO" couples. Comparing the plots **[Figure 2.5A-E],** it emerged that in the "Curcumin vs DMSO" plot **[Figure 2.5B]** the cloud of points was extremely large with several significantly differentially expressed genes and a similar, albeit less spread, situation stood out looking at the "CPTH2 vs DMSO" plot **[Figure 2.5A].** This trend changed in the two plots regarding the "intrabodies vs DMSO" conditions. Indeed, the cloud of points appeared to be drastically reduced and the "red spots" representing the differentially expressed genes were severely reduced **[Figure 2.5C-D],** particularly in the unrelated intrabody plot [**Figure 2.5D**]. These observations lead to two important considerations. On one hand, both Curcumin and CPTH2 treatments broadly affected gene expression in *S. cerevisiae* to a much greater extent, inducing deregulation of 3703 and 1963 genes (padj<0.05 and fold-change cutoff=1.5), respectively [**Figure 2.5A-B**]. This finding was in line with the expected pleiotropy of the HATi drugs. On the other hand, the scFv-58F intrabody induced less dramatic transcriptomic changes (378 DEGs with padj<0.05 and fold-change cutoff=1.5) **[Figure 2.5C]**. Despite the lower number of DEGs in the intrabody expressing strains for scFv-58F and for the negative control non PTM-targeting intrabody scFv-645 **[Figure 2.5D**], expression of scFv-58F still led to deregulation of a discrete set of genes. Notably, 73% (i.e., 276 out of 378) of the genes affected by scFv-58F expression were downregulated. This is in line with the selective interference of the intrabody with the H3K9ac modification, which usually marks active promoters. The set of DEGs in the scFv-58F-expressing cells provide, therefore, a transcriptional fingerprint of the effects of specifically blocking the H3K9ac epigenetic mark in yeast cells.

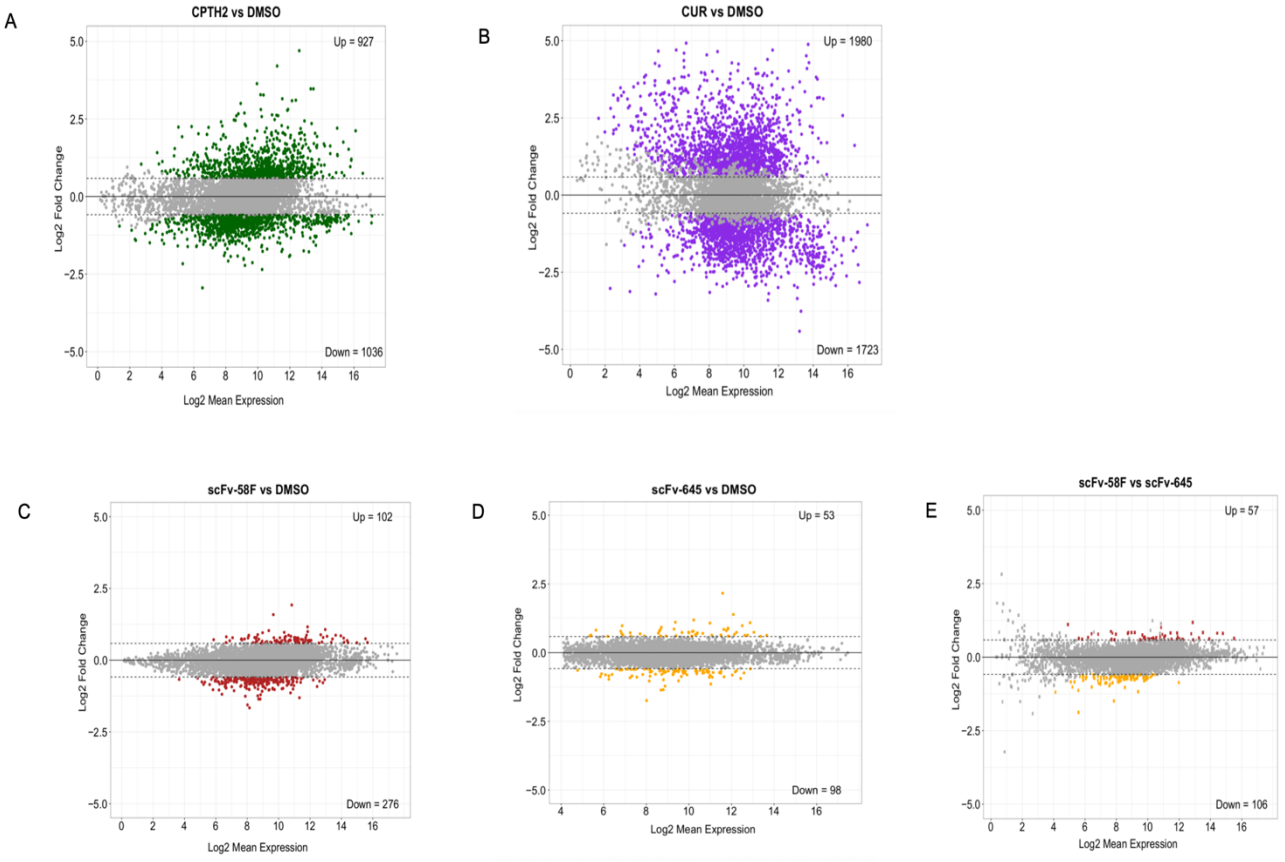


Figure 2.5. Both Curcumin and CPTH2 treatments broadly affected gene expression, while the scFv-58F intrabody induced less dramatic transcriptomic changes. (A-E) MA plot (i.e., log2 fold-changes on y-axis versus log average expression signal on x-axis) of the significantly differentially expressed genes (FDR=0.05), for CPTH2-treated cells vs the DMSO-treated control cells **(A)** and for Curcumin-treated cells vs the DMSO-treated control cells **(B)** MA plot representing the differentially expressed genes (FDR=0.05) for scFv-58F expressing cells **(C)** and yeast cells expressing the scFv-645 intrabody (**D)**, versus DMSO-treated cells. MA plot from the differential gene expression analysis between scFv-58F- and scFv-645-expressing yeast cells is also reported **(E)**. The number of significantly up-/down-regulated genes (padj<0.05 & FC cutoff=1.5) is reported for all comparisons.

The gene expression output data of DESeq2 were then visualized and analyzed using Heatmaps. We use the color code "red" for upregulated and "blue" for downregulated, in comparison to the DMSO control condition. To investigate the impact of the two HATi on gene expression, on the set of differentially expressed genes of "CPTH2 vs DMSO" a fold change cutoff of 2 was set and genes were selected considering the significance threshold of 0,05 (Padj<0,05; P value adjusted with the Benjamini correction). Furthermore, only those genes non-significant for 645 (Padj>0,05) were selected in order to account for non-specific gene expression changes due to the expression of an intracellular antibody in yeast cells. This set of differentially expressed genes was ordered by increasing log2(fold-change) in the heat-map shown in **Figure 2.6A.** This analysis revealed that the global gene expression pattern induced by CPTH2 and

Curcumin shows an overall similarity. Moreover, as displayed this set of genes (n=435) was only modestly affected in the scFv-58F condition.

Next, different statistical filters have been set in order to underline the presence of genes subsets specifically modulated by the scFv-58F intrabody. To this purpose we selected those genes that are significantly differentially expressed in the presence of the H3K9ac-targeting intrabody (padj<0.05 and fold-change cutoff=1.5) and subtracting those that are modulated by the non-specific control intrabody scFv-645. This leads to a list of 89 genes whose expression is specifically modulated by the scFv-58F intrabody. Notably, the overall expression pattern of the 89 genes defined in this way, was strikingly similar between the scFv-58F and CPTH2 conditions **[Figure 2.6B].** This result may be due to the fact that CPTH2 is an inhibitor selected through a yeast-based screening with the purpose to obtain a small-molecule specific for Gcn5 HAT. This HAT is the same used in the P.I.S.A screening Tethered catalysis bait for the selection of the scFv-58F and that is known to specifically acetylate H3 [32]. From the Venn diagram it also stands out that a large fraction (i.e., 62/89, ~70%) of the genes specifically modulated by scFv-58F was also modulated and differentially expressed after CPTH2 treatment **[Figure 2.6C]**. Furthermore, for this set of genes the expression pattern observed after Curcumin treatment appears to be reversed, if compared to scFv-58F and CPTH2 treatments. This is most likely due to the fact that not all actions of Curcumin on gene expression are ascribable to HAT inhibition, but also to some other known opposite effects on HDACs [227] and on other non-histone targets [228]. The expression of scFv-58F alters a subset of genes that constitute only one branch of the targets of the CPTH2 HATi drug, de facto pruning all the other undesired (i.e., non H3K9ac-dependent) transcriptional side-effects. Moreover, despite CPTH2 and Curcumin having a very similar overall effect on the cell, based on global transcriptional patterns, they differ in this particular subset modulated by the intrabody scFv-58F, namely the genes regulated by H3K9 acetylation, in which they demonstrated an opposite effect. This fact corroborates the idea that the effect of these HATi drug types is pleiotropic and modular and that we can directly affect a single branch of their targets with intracellular antibodies interfering with one specific PTM target.

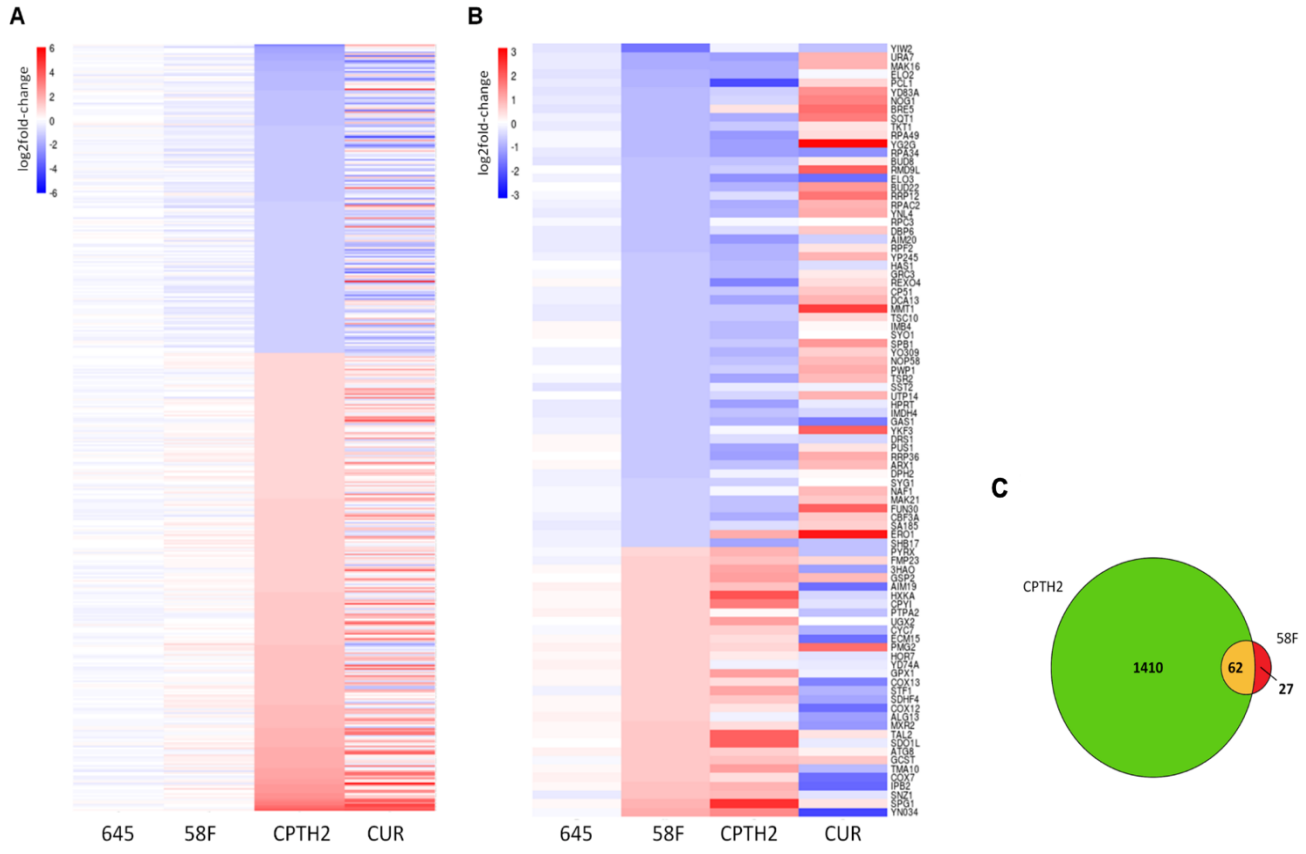


Figure 2.6. The scFv-58F modulates a small but largely overlapping subset of genes compared to CPTH2 treatment. (**A**) Heatmap of mRNAs (n=435) that are significantly modulated after CPTH2 treatment. mRNAs are sorted by log2FC in CPTH2 vs DMSO: expression patterns in the presence of CPTH2 and Curcumin show overall strong similarity. Statistical constraints: FC cut-off=2; Padj(CPTH2 vs DMSO)<0.05; Padj(scFv-645 vs DMSO)>0.05. (**B**) Heatmap of mRNAs (n=89) that are significantly modulated by scFv-58F. The genes modulated by the control intrabody scFv-645 were subtracted (see statistical constraints below). mRNAs are sorted by log2FC in scFv-58F vs DMSO: expression patterns in the presence of scFv-58F and CPTH2 treatment are similar, while the expression pattern is opposite with Curcumin treatment. Swiss-Prot Entry names were provided in the heatmap where primary gene names were not available. Statistical constraints: FC cut-off=1.5; Padj(scFv-58F vs DMSO) <0.05; Padj(scFv-58F vs scFv-645)<0.05; Padj(scFv-645 vs DMSO)>0.05. (**C**). Venn diagram showing the overlap between DEGs in scFv-58F and CPTH2 treatments vs DMSO (FC cutoff=1.5).

A subset of the genes most strongly modulated by scFv-58F was selected for validation by qPCR: we selected the two top upregulated (i.e., URA7, MAK16) and the two top downregulated genes (i.e., SPG1 and SNZ1) by scFv-58F expression **[Figure 2.7]**. Notably, qPCR data (solid columns) replicated the results obtained by NGS (striped columns) with a high grade of statistical significance. From these results we conclude that the transcriptional effect obtained by a direct interference with H3K9ac, using an H3K9ac-selective intrabody, is much more defined and restricted, when compared to that obtained with HATi treatment, which results instead in a much broader transcriptional outcome.

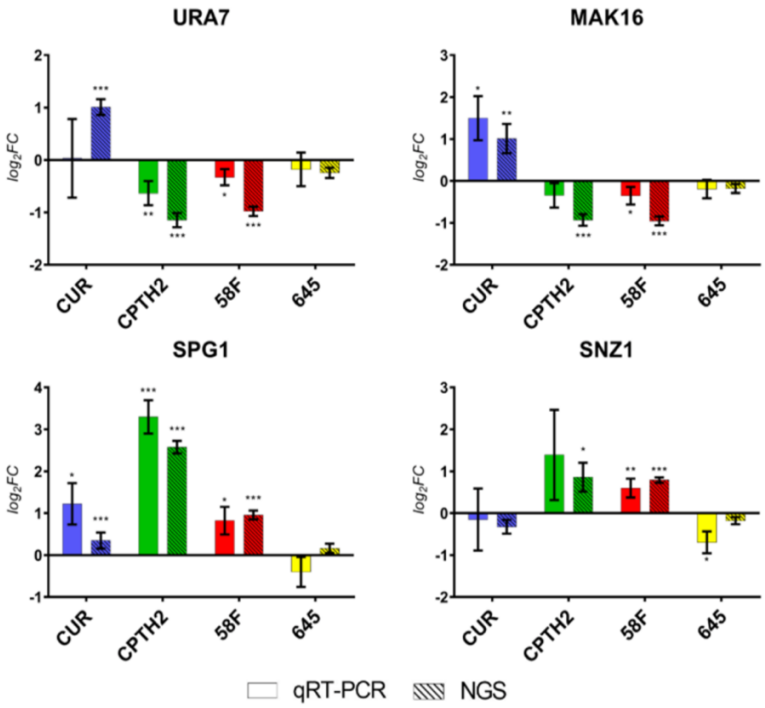


Figure 2.7. Validation through RT-qPCR of selected candidate mRNAs identified from RNAseq data. Log2FC +/- SD values of real-time PCR (solid colour) and RNAseq data (striped) for the four chosen differentially expressed genes in the presence of scFv-58F. Values for each treatment are representative of n=3 biological replicates. For real-time PCR data: ΔCt = Ct GENE – Ct TUB1. log2(FC) = -ΔΔCt normalized on tubulin (TUB1) and DMSO control samples. SD = $\sqrt{\boldsymbol{var}(\boldsymbol{\Delta Ct\ DMSO}) + \boldsymbol{var}(\boldsymbol{\Delta Ct}\mathbf{TREATMENT})}$, where "TREATMENT" represents each of the four different experimental conditions (CUR; CPTH2; scFv-58F; scFv-645). Student t-test (TREATMENT vs DMSO, two tails) was performed on ΔCt values. Calculated P values were adjusted with the Benjamini- Hochberg procedure; the same statistical correction was applied for RNAseq data.

2.2.6 Gene ontology enrichment analysis and gene by gene comparison

The 89 genes differentially expressed in the presence of the scFv-58F were submitted to Gene ontology (GO) enrichment analysis performed with the online software *DAVID Ontology*. In particular, these genes were analyzed providing as a background the list of genes not significantly changed in the scFv-645-expressing yeast cells.
The main GO terms obtained **[Figure 2.8]** were found to be nucleolus, ribosome biogenesis and rRNA processing, among others, showing a prominent modulation of genes involved in ribosome biology. This indicates that the H3K9 acetylation might be specifically connected with the collective transcriptional regulation of genes involved in ribosome biogenesis and function. Notably, a similar link between H3K9 post-translational modifications and ribosome function in yeast was previously reported in histone mutagenesis studies [229]or via RNA-seq and ChIP-seq analysis [230,231] .

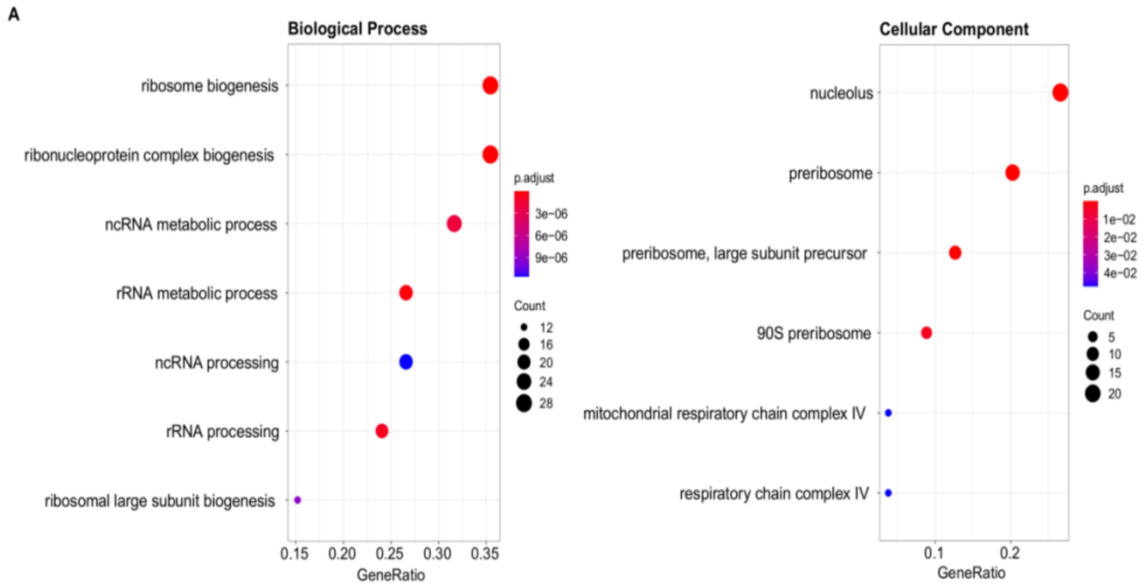


Figure 2.8. Gene ontology enrichment analysis: scFv-58F modulates genes involved in ribosome biogenesis and function. Gene ontology enrichment analysis of those genes significantly modulated in the presence of ScFv-58F while non-significant in the presence of ScFv-645. Bubble plots depict the main significant terms for the categories "Biological Process" (left) and "Cellular Component" (right).

Looking at the sorted list, it stands out that among these 89 genes differentially expressed the majority (65% of DEGs (n=58)) are downregulated. Theoretically, this is in line with the selective interference by the intrabody with the acetylated H3K9, a PTM known to mark

transcriptionally active promoters and hence generally associated with transcriptional activation [7]. Therefore, binding of the intrabody to H3K9ac may competitively hinder the access of other transcriptional regulators to this docking site on chromatin, thereby affecting transcription.

The following step was to investigate the single genes with the most different expression level into this list, in one direction or another. Among the main downregulated genes[8] **[Table 2.1]**, it is possible to observe that several genes like MAK16 and NOG1, but also RPA49, RPC82 and RPC19 are involved in different phases of the ribosome biogenesis, in line with the gene ontology output. It is also interesting to notice that, for the strongest downregulated gene YIW2, a putative N-acetyltransferase activity related to GNAT family HATs (like Gcn5), although not still completely characterized, has been reported (https://www.uniprot.org/uniprotkb/P40586/entry#function).

Table 2.1 Downregulated genes involved in GO significant terms in the presence of scFv-58F (with Log2FC expression respect to the control DMSO).

Downregulated genes	Function
YIR042C (YIW2)	Pupative N-acetyltransferase activity (GNAT dom.)
URA7	Catalyze ATP-dependent amination of UTP to CTP
MAK16	Constituent of 66S pre-ribosomal particles; required for maturation of 25S and 5.8S rRNAs
ELO2	Component of elongase II, which elongates 16-18 carbon fatty acyl-CoAs
PCL1	G1/S-specific cyclin partner of the cyclin-dependent kinase (CDK) PHO85
NOG1	Involved in the biogenesis of the 60S ribosomal subunit
TKL1 (TKT1)	Transketolase 1 (pentose phosphate pathway; needed for synthesis of aromatic amino acids)
RPA49	Component of RNA-pol I which synthesizes ribosomal RNA precursors (heterodimer with RPA34)
RPC82 (RPC3)	Specific core component of RNA-pol III which synthesizes small RNAs, such as 5S rRNA and tRNAs
RPC19 (RPAC2)	Common core component of RNA-pol I and III

[8] Ten genes among the downregulated group that emerge from the sorted heatmap are reported. Functions have been obtained by the Uniprot and SGD databases.

On the other hand, looking at the other side of the list regarding the upregulated genes[9] **[Table 2.2]**, it is difficult to identify a strong biological area common to a group of genes. However, several genes like SPG1, COX7 and TMA10 code for proteins with mitochondrial localization. From the comparison of the two lists of selected genes, it is interesting to notice that among the downregulated genes there is TKL1 that codes for the transketolase-1 involved in the pentose phosphate pathway, while among the upregulated the NQM1106 gene codes for a transaldolase that is also involved in the same pathway. The presence of two genes involved in the same metabolic pathway and inversely regulated deserves particular attention. However, it is important to remember that caution must be paid in considering significance changes in metabolic genes in NGS data analysis, because the first effects observed with every type of treatment are generally due to uncontrolled experimental factors acting on metabolism.

Table 2.2 Upregulated genes involved in GO significant terms in the presence of scFv-58F (with Log2FC expression respect to the control DMSO).

Upregulated genes	**Function**
YNR034-A (EGO4)	Uncharacterized
SPG1	"Stationary Phase Gene" (mitochondria)
SNZ1	Catalyzes the formation of pyridoxal 5'- phosphate (member of a stationary phase-induced gene family)
PBI2 (IPB2)	Cytosolic inhibitor of vacuolar proteinase B (component of the LMA1 complex, involved in the facilitation of vesicle fusion)
COX7	Subunit VII of cytochrome c oxidase (mithocondria)
TMA10	Associates with ribosomes; protein abundance increases in response to DNA replication stress (mithocondria)
NQM1 (TAL2)	Transaldolase (pentose phosphate pathway)

For the sake of completeness, also the DEGs impacted in the CPTH2, and Curcumin treated samples were submitted to GO enrichment analysis. While 58F modulated genes involved in ribosome biogenesis and function, CPTH2 mainly affected genes involved in oxidative stress and ion homeostasis **[Figure 2.9].** Interestingly, before applying the correction for multiple testing (FDR=0,05), the pathway ribosome biogenesis and rRNA processing resulted among the process in which the DEGs modulated by CPTH2 were involved (Padj=0,32) as those of scFv-

[9] Seven genes among the upregulated group that emerge from the sorted heatmap are reported. Functions have been obtained by the Uniprot and SGD databases

58F. Regarding the Curcumin-treated condition, choosing a FC cut off= 2 it still resulted in more that 2 500 DEGs (in line with the broad effects exerted by this drug), too many to perform GO enrichment analysis. Only choosing a FC cut off= 4, we selected a proper number of genes to correctly perform the analysis. Mainly, the Curcumin targets were involved in ribosomal function and sporulation **[Figure 2.9].** In yeast this morphogenetic event is associated with cells under stress condition and starvation, involving the alteration of the vegetative machinery.

In the end, the analysis turned out to be useful to furtherly support our hypothesis that the number of genes dysregulated by the scFv-58F is lower than those dysregulated by CPTH2 or curcumin, due to the specificity of the target.

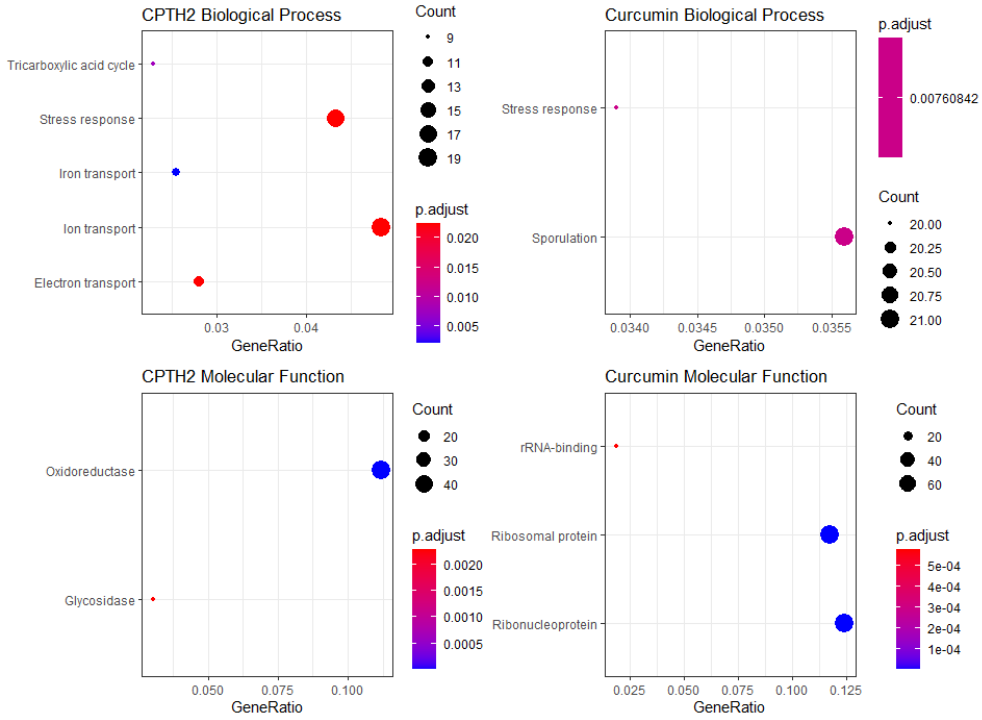


Figure 2.9. Gene ontology enrichment analysis. Gene ontology enrichment analysis of the genes significantly modulated by CPTH2 **(left)** FC cut-off=2; Padj(CPTH2 vs DMSO)<0.05; and Curcumin **(right)** FC cut-off=4; Padj(Curcumin vs DMSO)<0.05.

2.3 The scFv-58F Project Discussion

2.3.1 Summary of the main findings of the scFv-58F Project.

This thesis work has provided the first evidence that the functional effect on gene expression of the direct interference on a PTM-modified protein, achieved by an intrabody PTM-specific, is highly more selective than that obtained inhibiting the enzymes responsible for installing the post-translational modification itself. Indeed, the silencing effect on transcription obtained by the expression of the ScFv-58F intrabody able to specifically recognize H3K9Ac results in a remarkably more restricted and specific effect on gene modulation than that obtained inhibiting the enzymes responsible for H3 acetylation. In summary, this part of my thesis work led to the following achievements and biological findings:

- The first characterization of the transcriptomic consequences of the scFv-58F expression in yeast cells, in comparison to yeast cells exposed to two broadly used HAT inhibitors using Next Generation Sequencing.
- The identification of a group of genes involved in specific biological function as target for the transcriptional regulation by H3K9. Results also independently confirmed throughout an exhaustive comparison with recent findings in literature.
- The validation of *a PTM Chromatin Reading Antibody approach* as a novel strategy to achieve a new level of specificity in epigenetic research.

2.3.2 The transcriptomic effects of the anti-H3K9ac scFv-58F intrabody in comparison to those induced by HAT inhibitors

Several steps have been made in order to highlight the differences between these two approaches. Firstly, two HAT inhibitors have been validated to achieve proper conditions of histone acetylation inhibition in yeast, to be compared to the cells expressing the scFv-58F intrabody.

The choice of proper controls and experimental design have been crucial to highlight the differences obtained by the two methods of PTM perturbation, for the NGS data analysis. First of all, the yeast line with an empty pL220-HA plasmid offers the ideal genetic background for

the drugs and DMSO treatments. The yeast line carrying the unrelated scFv-645 intrabody provides instead a way to calibrate the metabolic and non-specific effects due to the expression of a generic intrabody into the yeast cells. Moreover, both "drugs treatment" and "intrabody-expressing" conditions had a common control (the same DMSO concentration) that helps to level out the differences between all the considered experimental conditions. On the contrary (and as expected), the so designed experiment allowed to emphasize that the presence of the H3K9Ac- selective intrabody perturbs the yeast cells at the transcriptional level in a highly specific manner. HATs inhibition by the two drugs affected the expression level of a multitude of genes, with Curcumin displaying a larger effect, in line with the literature. The HAT inhibitory activity of Curcumin is principally exerted on p300/CBP, even though it has pleiotropic [228], often contrasting effects. Indeed, Curcumin has also been reported to inhibit HDAC enzymes, even more potently than other HDAC inhibitors like valproic acid and sodium butyrate [232]. Moreover, Curcumin activities both in DNA methyltransferases inhibition [233] and in regulation of miRNA expression [234,235] have been reported. Such broad biological activities underlie the massive transcriptional dysregulation observed after Curcumin treatment.

On the other hand, CPTH2 is a synthetic small-molecule with a specific inhibitory activity on Gcn5 [222]. This lysine acetyltransferase enzyme is known to acetylate histone H3 lysine 9 [27,223], in addition to other non-histone targets such as the tumor suppressor p63, the c-myc oncoprotein, the NF-kB transcription factor [236] and the metabolic co-activator PGC-1alfa [237]. Accordingly, the NGS data analysis shows that CPTH2 yields a more limited effect on gene expression compared to Curcumin (but a significantly broader effect on gene expression than that observed in the yeast cells expressing the scFv-58F intrabody.) It is remarkable that the genes specifically modulated by scFv-58F expression, in great part (~70%) overlap with those modulated by CPTH2 treatment, despite representing a minor fraction of the latter (4,2%). Interestingly, inspection of the scFv-58F DEGs not overlapping with CPTH2 DEGs (27 genes) showed they encoded for proteins involved in processes such as rRNA/ncRNA processing and metabolism.

Altogether, these data had supported the prediction that targeting a specific PTM yields more specific results than inhibiting the enzyme that installs that particular PTM, reflecting the conceptual difference of the two approaches; while the enzyme inhibitors determine a broad cascade of downstream effects, due to the pleiotropic activity of the enzyme on different histone residues and on many non-histone substrates [30,238], the intrabody selectively acts on one particular histone post-translationally modified residue (H3K9ac). Thus, while CPTH2 or Curcumin act on broadly acting *writers*, the scFv-58F intrabody targets directly a single and specific *word*.

From the mechanistic point of view, we envision that the most likely action of the H3K9ac-specific scFv-58F intrabody is to competitively block the access of H3K9ac readers onto the K9 acetylated H3 protein (either chromatin-engaged or free). In this respect, the H3K9ac scFv-58F antibody domain works like a competing "*reader*" protein, a Chromatin Reader Antibody (CRA) devoid of the natural readers effector functions. The CRAs are a new class of molecules selective for specific chromatin marks that can be used as modular building blocks to build combinatorial marks readers (eg. AND gate, OR gate) or readers with novel effector functions. Besides, by occluding the acetylated K9 on H3 histone, the scFv-58F might prevent the release of acetyl groups from histones as acetate, thereby enabling subsequent acetylation. Several studies, in fact, have suggested the ability of acetate to influence histone acetylation in mammalian and yeast cells [239].
In recent years, reader inhibitors have been developed (e.g. the BET-bromodomain inhibitor JQ1 [27]). However, BET-bromodomains show a considerable target promiscuity [240]. Therefore, any BET-bromodomain inhibitor will intrinsically have pleiotropic effects and henceforth determine broad downstream effects on global gene expression. In any case, a comparison of the H3K9ac scFv-58F intrabody with BET-bromodomain inhibitors, in future work, will be undoubtedly informative. In the field of functional studies, in fact, BET inhibitors can only be used to address questions concerning broad effects of protein acetylation on transcription, rather than for the precise assessment of the role of a specific acetylation on a given protein.

2.3.3 Inhibition of K9-acetylated H3 histone by scFv-58F: effects on ribosome biogenesis

As for the specific PTM analyzed here, namely the acetylated lysine 9 of H3, the specificity of the H3K9ac-specific interfering intrabody over standard HATi drugs, allowed to demonstrate the transcriptional regulation of genes involved in ribosome biogenesis and function, as a likely downstream target of H3K9 acetylation in *Saccharomyces cerevisiae*. Specifically, the GO analysis has indicated a strong effect of scFv-58F in determining the downregulation of genes that encoded for proteins involved in processes such as rRNA/ncRNA processing and metabolism. This fundamental biological process is strictly linked to growth, since the availability of the protein construction machinery is indispensable in growing cells, as well as in proliferating cells.

It could be argued that the effect of the intrabody on ribosome biogenesis, could be a consequence of the specific PTM-selectivity in the context of a particular growth and metabolic landscape of the cell, in which an enrichment of the "intrabody target" (H3K9Ac) is probably

present on a defined subgroups of promoters such as those of ribosome-related genes. To test this hypothesis, it is possible to imagine an elegant experiment in which the expression of the scFv-58F would be induced in different growth phases of a cell, in order to assess if the effect on global gene expression changes with regard to the time interval in which the intrabody is expressed. Interestingly, links between H3K9 acetylation and the modulation of genes involved in ribosome function and biogenesis in yeast have been previously reported, on the basis of systematic mutagenesis of the acetylated lysines [229] or via ChIP and ChIP-seq analyses [230,231,239]. The latter studies established that induction of histone acetylation (and notably of H3K9 acetylation) occurs at several “growth-related genes”, upon stimulation of yeast cells into growth phase. The genome-wide locations of H3K9 acetylation, assessed by ChIP-seq analysis, showed that H3K9 acetylation was present almost exclusively at the promoter of “growth-related genes” (e.g., genes involved in ribosome function and biogenesis, translation and amino acid metabolism) to enable their transcription [230,239]. Accordingly, comparing the genes modulated by scFv-58F expression with the ones identified by H3K9ac ChiP-Seq after yeast growth induction [239], we found that 50 out of 89 genes specifically modulated by scFv-58F were previously identified as "growth-related" genes, characterized in previous work [239] by elevated H3K9ac ChIP-seq signal on their promoters. On the contrary, only 14 out of 151 (less than 10%) DEG genes specifically modulated by scFv-645 were genes characterized by elevated H3K9ac ChIP- seq signal on their promoters. Moreover, 49 out of the 50 "growth-related" genes modulated by scFv-58F were downregulated, reinforcing the hypothesis that the H3K9ac-binding intrabody acts by inhibiting transcription of a specific subset of genes in yeast (i.e., 84%, 49/58, genes downregulated by scFv-58F, identified as “growth-promoting”). These data from the literature provide an independent validation of the PTM-selective interference method and show that the superior precision of this method can generate new valuable data and hypotheses for future investigations.

2.4 The scFv-58F Project: Future perspectives

2.4.1 Further characterization of the specific silencing effect of the scFv-58F intrabody

The results obtained in this thesis project encourage us to move forward with the purpose to highlight the potentiality of the scFv-58F in the description of epigenetic processes but also to extend this approach to new biological applications. First, to integrate the work done in yeast cells and to more precisely compare different concentrations of the drugs with the intrabody a dose-response curve for the latter will be determined using an inducible promoter. Even if the scFv-58F expression is under the control of the ADH1 promoter, thus resulting in high protein level production, it could be argued that the smaller scale of the effect of the intrabody could be due to unknown intracellular events that may reduce the real activity of the intrabody (e.g.rapid degradation). Secondly, another elegant experiment to further characterize the scFv-58F would pass through the comparison with known bromodomain inhibitors, extending the demonstrated greater selectivity of the intrabody also to this different inhibition approach.
BET-bromodomains (BRD) show a considerable target promiscuity [240]. Therefore, any BET-bromodomain inhibitor will intrinsically have pleiotropic effects and henceforth determine broad downstream effects on global gene expression.
A comparison of the H3K9ac scFv-58F intrabody with BET-bromodomain inhibitors will be undoubtedly informative (specifically, we are planning to perform the comparison with one of the most used BET-bromodomain inhibitor named JQ1 [27]). Furthermore, pivotal information can be inferred throughout a competitive assay between scFv-58F and BRD containing proteins. We reasoned that the specific impact of the scFv-58F might be confirmed by evaluating the genomic expression of the yeast strain expressing the intrabody and overexpressing a BRD protein. In this scenario, we expected that the effect of the intrabody described in this work might be impaired or restrained due to the competition for the target, namely H3K9ac. Finally, to confirm that the effects of scFv-58F intrabody expression result from its target recognition, a key experiment would involve comparing the transcriptomic profile of yeast cells stably expressing the scFv-58F intrabody to that of another yeast line expressing a mutated form of the scFv-58F unable to bind the H3K9ac. The latter scFv-58F format could be obtained by introducing point mutations in the CD3 region.
Otherwise, in a more elegant manner, we could envisage to engineer a photoactivatable scFv-58F and compare the gene expression before and after photo-activation. Potentially, in fact,

photocaged amino acids can be introduced within the nanobody-binding interface, which, after photo-activation, show instantaneous binding of target proteins[241].

2.4.2 Functional characterization of the specific silencing effect of the scFv-58F intrabody in mammalian system

Having investigated the transcriptional consequences of H3K9ac-specific interference in yeast cells, we plan to explore if this PTM-specific intrabody approach is also effective in mammalian cells. We have already confirmed the physical association between scFv-58F and acetylated H3 histone in stably-transfected HeLa cell lines expressing via immunoprecipitation (IP) assay [32]. Thus, the next step will be to investigate the functional relevance of this interaction, by exploring the gene expression of candidate genes in stably-transfected HeLa cells expressing the scFv-58F in the nucleus (or the control intrabody scFv-645). We foresee that upon expression in the nucleus the intrabody scFv-58F induces specific transcriptional regulation of target genes in mammalian systems. Of course, toxicity assays should be performed to exclude any unwanted effect in the mammalian system.

As a first approach, we plan to measure the expression of transcripts which are human homologues of those most significantly regulated in yeast cells expressing scFv-58F by qPCR. Also, other genes whose expression had been shown to be regulated by HATi treatment, or by knock-down of the GCN5/PCAF complex[10], in HeLa cells, in previous reports will be investigated [7,229,242] will be tested.
These experiments will be particularly important to define if also in human cells the intrabody scFv-58F allows a more specific transcriptional modulation than the HAT inhibitors or than interfering with GCN5/PCAF through precisely targeting H3K9ac.
Finally, a genome-wide analysis in mammalian cells will definitely be the object of further investigation: the global transcriptome of Hela cells treated with scFv-58F will be profiled to hopefully open novel avenues for future studies in mammalian systems in vivo.

2.4.3 Application for anti-PTM-Histone Intrabodies: IntraChip and proteomics.

From a methodological point of view, one further application of anti-Histone antibodies is Chromatin Immunoprecipitation (ChIP). However, this technique generally foresees the immunoprecipitation of chromatin with the chosen probe only after several experimental steps

[10] Both GCN5 and PCAF are lysine acetyltransferases (KATs) incorporate in the SAGA complex (from Spt-Ada-Gcn5-acetyltransferase) that preferentially modifies histone H3 on Lys9 (H3K9) and therefore involved in gene activation. Both the HAT domain and the bromodomains of GCN5 and PCAF bridge interactions with both histones and transcription factors.

(i.e.crosslinking with paraformaldehyde, cell lysis, DNA sonication) that could lead to loss, or alteration, of information [243]. For instance, formaldehyde used before antibody probes could alter Histone Lysines. By using scFv-58F and and other anti histone PTM-selective intrabodies, it will be possible to investigate if a consistent improvement in quality of ChIP can be achieved by previously expressing the chromatin-precipitating antibody inside the cell. The idea is that of expressing the intrabody in cells and then proceed with Ip protocol. We called this application IntraChip **[Figure 2.12].** Replication of a ChiP study performed using intracellular expressed scFv-58F will help to compare gene reads between classic Ip and the proposed intraChiP. In this way the anti-AcH3K9 could be seen as a novel tool to study and directly manipulate the real epigenetic state of a cell. Furthermore, this opens the possibility to directly express an intra-probe only inside a desired cell type via a viral vector (using a lineage-specific promoter), for example allowing a cell-type specific ChIP from a heterogeneous tissue, such as the brain.

Also, the PTM-specific proteomic studies will greatly benefit from the ability to target intracellularly and in vivo the PTM pool of a protein. The intrabody could be fused to biotin ligase to identify the protein interactome at a specific chromatin mark on the basis of proximity biotinylation or to TAP tags for Affinity purification Mass Spectrometry [244].

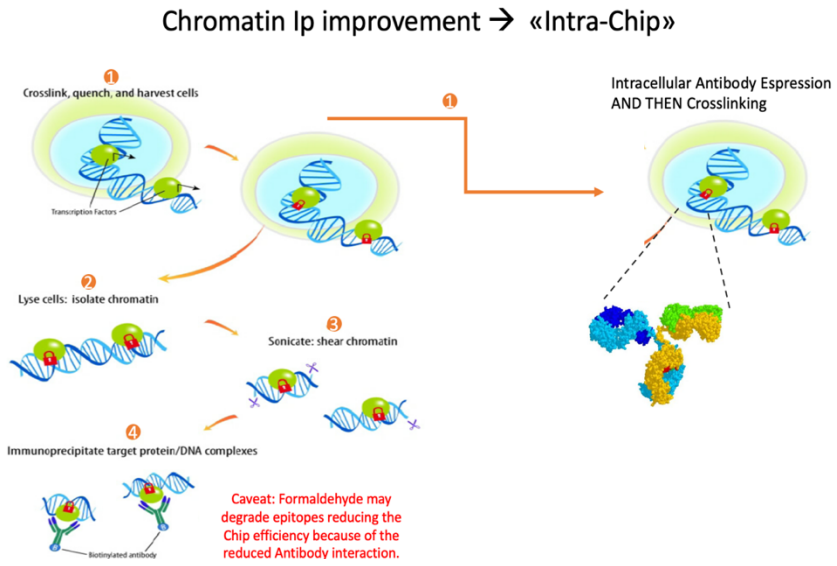


Figure 2.10. Intra-ChIP scheme: a possible application and proof of concept for an anti-Ac-H3 intrabody.

2.4.4 Application of anti-PTM-Histone intrabodies: sensors and tools to edit the epigenome

The potency and versatility of the anti-PTM approach, that relies on the universal, diverse and modular nature of antibody domains can be in fact further implemented and tailored by adding suitable effector or sensor functions to the H3K9ac binding moiety [29]. One could envisage to engineer the PTM-specific intrabody for live imaging of the PTM pool of the protein (used as sensors for a PTM-bursting event or for tracking acetylated chromatin *in vivo*) [245] or to recruit actuators for chromatin retargeting *in living cells*, similarly to an engineered Chromatin Reader (eCR) [246]. In this respect the H3K9ac scFv-58F antibody domains would work like Chromatin Reader Antibody modules that could be used as modular building blocks to build combinatorial marks readers or readers with novel effector functions allowing precision epigenome editing. On one hand, the interbody moiety working as a competitor reader would allow to directly assess the functional role of a deposited chromatin mark per se **[Figure 2.13a]**. On the other hand, the moiety appended to a catalytic domain (CD) of a DNA- or histone-modifying enzyme or a degron domain (allowing a PTM-specific target degradation) **[Figure 2.13a-b].** In recent years, proteomics-based studies have greatly enhanced current knowledge about interactions between proteins and chromatin marks [247]. However, the available methods rely on artificial chromatin, methods that require access to the underlying DNA, leading to chromatin disruption [248,249]. Conversely, the novel Antibody Chromatin Reader approach will enable detection of dynamic interactions between proteins and physiological chromatin in living cells.

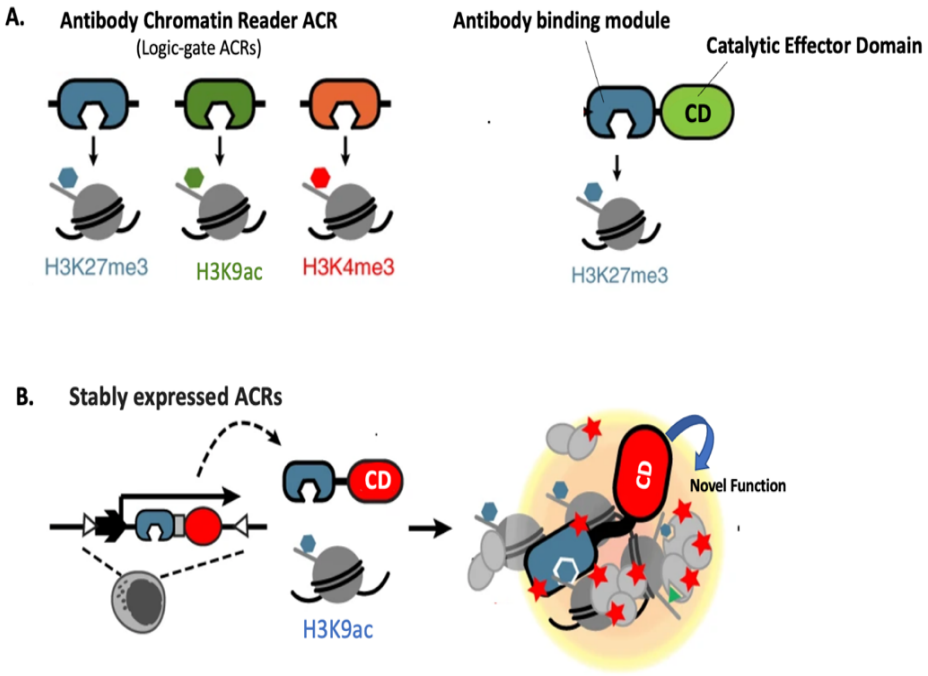


Figure 2.11.Antibody Chromatine Reader: a possible application of anti PTM-Histone-intrabodies. A. On the left, Chromatin Reader Antibodies (ACRs) and their specificities toward chromatin marks. The ACRs could be fused together to generate LOGIC-gate ACRs. On the right, schematics of an ACR (e.g. scFv-58F) fused to a novel effector function (ie. A Catalytic Domain (CD)). **B.** On the left, stably expressed ACRs. On the right schematic describing a possible use of ACRs fused to a CD in living cells.

3. The PISA 2.0 Project

3 The PISA 2.0 Project

As anticipated, recent results from my lab provided a solid and robust scientific and technological basis to isolate intracellular working binders against Post Translationally Modified proteins: the PISA (Post-translational Intracellular Silencing Antibodies) technologies. The platform represents a significant breakthrough in that it solves the gap between the huge potential provided by the PTM proteome of cells in physiology (and in pathology) and the possibility to validate this class of targets, implementing a PTM-selective intracellular interference (as well-exemplified by the work done with the scFv-58F (cfr. The scFv-58F Project).

In the present embodiment, PISA technology relies on tethered catalysis[105]: the yeast-two-hybrid (Y2H) host cells express a bait genetically fused with the catalytic domain of a modifying-enzymes to be constitutively PT-modified *in cis*. This approach turned out rather general and flexible, however, tethered catalysis PISA screening has some limitations providing only a partial solution to the possibility of *dissecting the contribution of each individual PTMs*. Possible limits of the methodology are:

i) large bait dimensions → degradation and difficult migration into the nucleus;
ii) the impossibility of *a priori* designing sites of single PTMs: chimeric baits can be modified at all the possible sites;
iii) many modifying enzymes are not known

A natural extension of the technology would be to perform PISA selections in 2H platform cells where the Genetic Code Expansion [113] permits the *site-specific genetically encoding of PT-modified amino acids* (e.g. Acetyl-Lysine, Phospho-serine) in baits for intrabody selections. In this thesis chapter, all the required steps and optimizations to adapt the Expanded Genetic Code technology to the PISA selection system will be detailed **[Figure 3.1].** This PISA 2.0 version development will be carried out both using the Yeast-Two-Hybrid platform (being the original host of the PISA selection) **[The PISA 2.0 Project in yeast]** and in *E.coli* exploiting the Bacterial-Two-Hybrid methodology **[The PISA 2.0 project in *E.coli*].**

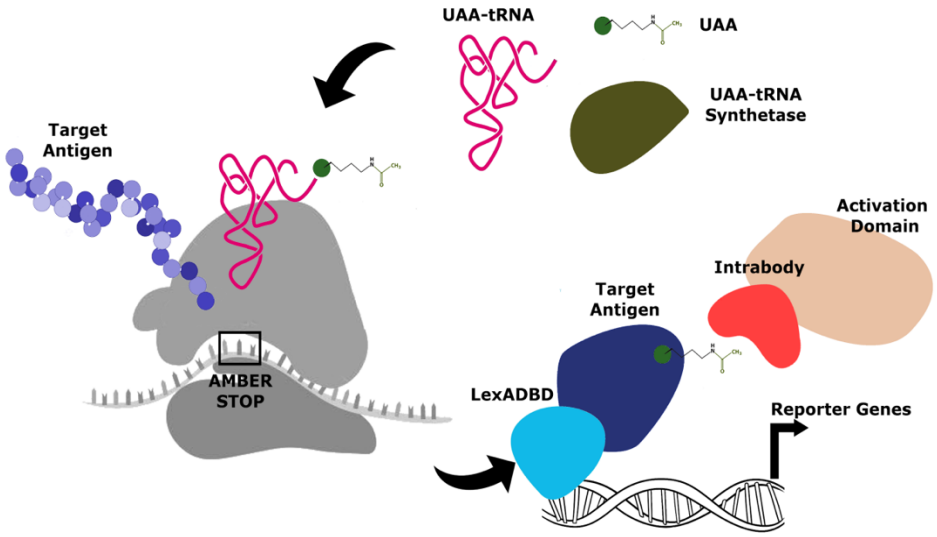


Figure 3.1. Schematic of PISA 2.0 in Yeast. The upgrade of the PISA technology foresees the substitution of tethered catalysis with expanded genetic code technologies to encode the PTM genetically in the target protein. Provided a tRNA synthetase matching set able to work in yeast, the pair will be integrated in both the S. cerevisiae L40 screening strain as an amber codon reallocation and in a Bacterial- Two-Hybrid strain. The new strains will then enable to perform PISA screening to select intrabodies against PTMs genetically encoded in the target antigen through the use of intragenic amber stops. UAA = unnatural amino acid

3.1 The PISA 2.0 Project in yeast: Material and Methods

In this chapter the materials and methodology employed to adapt the GCE to the PISA yeast strain will be described. In particular, all the required steps for the realization of yeast stable line, bait vectors, and the protocol to encode the PTM-AAs in yeast will be detailed.

3.1.1 PCR Protocols

Q5 polymerase (New England BioLabs (NEB), M0491). 1 ng of template plasmid DNA has been used for the initial PCR reaction with an appropriate couple of Forw + Rv primers. Q5 DNA polymerase ensures high-fidelity copies up to 5kb of amplicon length, with a speed of 1 kilobase every 30-60s. PCR using Hifi Q5 has the following protocol: 90°C×30s+(98°C×10s+T_{ann} X°C×30s+72°C×T_{el})×n°cycles+72°C×5min+4°C×∞[11]

Go Taq G2 polymerase (Promega, M7841). 1 ng of template plasmid DNA (or a bacterial colony) has been used for the initial PCR reaction with an appropriate couple of Fw + Rev primers. This polymerase has a speed of 1 kilobase per minute. In this work, Taq was used only for colony PCR testing. PCR using Taq has the following protocol: 95°C×3min+(95°C×30s+T_{ann} °C×30s+72°C×T_{el})×n°cycles+72°C×5min+4°C×∞

Reverse Transcription PCR protocol for transfer RNAs.

This method was used to retro-transcribe the orthogonal tRNACUAPyl and the Sc-tRNAArg. 1µg of total RNA was retro-transcribed into cDNA using the miScript II RT Kit (Qiagen). Briefly, The miScript Reverse Transcriptase Mix was inactivated for 5 min at 95 °C. 10 ng of cDNA has been used for PCR amplifications using specific primers.Script Reverse Transcriptase Mix is an optimized blend of poly(A) polymerase, reverse transcriptase, dNTPs, rATP, oligo-dT primers. tRNAs are polyadenylated by poly(A) polymerase and reverse transcribed into cDNA using oligo-dT primers. Polyadenylation and reverse transcription are performed in parallel in the same tube. The reaction was prepared according to the manufacturing conditions and incubated for 60 min at 37 °C.

Site-directed mutagenesis PCR protocol:

The parental plasmid containing cDNA insert is amplified using Q5 protocol with the mutagenic forward or reverse primer. The primers were designed complementary to each other with a length ranging from 10–20 nt of unmodified sequence on both sides of the mutation (The mutated bases were usually in the center of both primers). The PCR reaction was incubated

[11] Legenda: T_{ann} stands for "annealing temperature", T_{el} stands for time of elongation, and n° cycles is the number of cycles.

with DpnI (NEB) for 4h at 37°C in order to digest methylated non-mutated DNA (i.e. the parental plasmid). The reannealed plasmid was transformed into competent *E. coli* cells.

3.1.2 Cloning methods

Direct ligation of PCR products. All the constructs generated during this thesis project were cloned using either classical ligation methods or via Gibson assembly.
Backbone and HiFi-PCR fragments digestion. The digestion recipe for 1 µg of DNA was the following:

- DNA
- 10x NEB SmartCut® buffer
- 1 enzymatic unit (EU)
- Water (total reaction volume = 50 µl)

Small-scale plasmid ligation. These ligations were done by combining 50 ng of plasmid backbone with a 3:1 molar concentration of the fragment to be inserted, unless otherwise specified. The reaction was carried out with 1ul of T4 DNA Ligase (NEB, M0202) (along with its buffer) in a total volume of 20 µl. After incubating for 4 hours at RT or overnight at 16°C, the ligation products were transformed into DH5α (Invitrogen, EC0112) CaCl2 competent cells.
Gibson assembly. This technology is a robust exonuclease-based method for the easy assembly of multiple linear DNA fragments [250]. Regardless of fragment length or end compatibility, multiple overlapping DNA fragments can be joined in a single isothermal reaction. With the activities of three different enzymes, the product of a Gibson Assembly is a fully ligated double-stranded DNA molecule. This method has been widely adopted and is a major workhorse of synthetic biology projects worldwide. Following the protocol, the required overlapping fragments were created via PCR with primers that contain a 5′ end that is identical to an adjacent segment and a 3′ end that anneals to the target sequence. Adjacent segments should have identical sequences on the ends (sequences A and B in the figures). These identical sequences can be created via PCR with primers that contain a 5′ end that is identical to an adjacent segment and a 3′ end that anneals to the target sequence. The fragments were mixed in appropriate amounts (depending on their lengths in base pairs) and added to the provided 2x NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621) in a final volume of 10 uL. The reaction is incubated at 50°C for 15′ up to 1 hour. The assembly reaction products were directly transformed in bacteria by electroporation or in $CaCl_2$-transformation. Gibson assembly was exploited to clone the target cDNAs in pMICBD1 bait vector and tRNA/synthetase pair genes in either the pRSII402 or prSII422 plasmids (cfr. 3.1.3).

3.1.3 Plasmids and Constructs

Here follows a list of the most important vectors and constructs used in this study: i) the Y2H bait and prey vectors and ii) the orthogonal translation system specific vectors.

pMICBD1 Y2H bait plasmid [95]uses the strong alcohol dehydrogenase promoter (ADH1) to express bait proteins as fusions to the DNA-binding protein LexA **[Figure 3.2 A].**

pL220 Y2H prey plasmid was used for expression of recombinant antibody domains (or antibody domain library) as fusion to the VP16 activation domain **[Figure 3.2 B].**

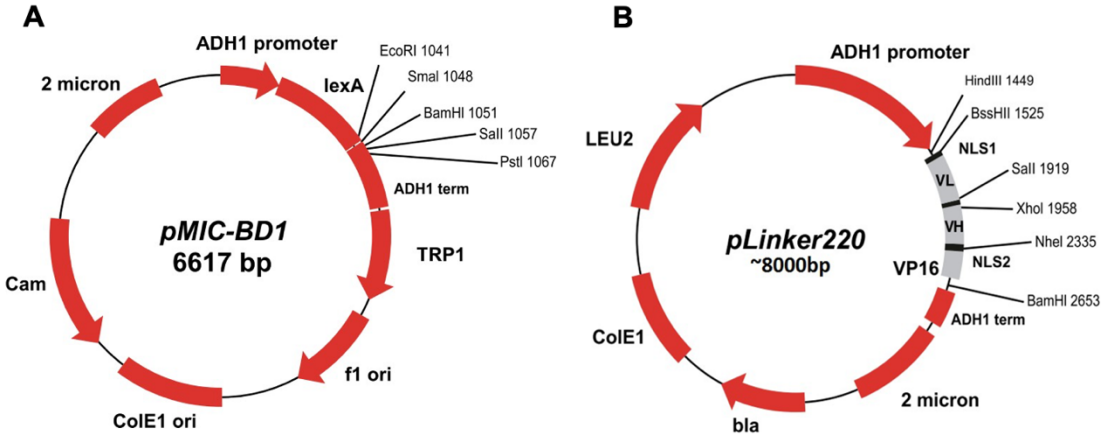


Figure 3.2 (A) pMICBD1 bait plasmid map and polylinker. ADH1 promoter= Alcohol Dehydrogenase promoter (yeast); lexA= LexA DBD; ADH1 term= alcohol dehydrogenase terminator; TRP1= TRP1 functional gene, restore Trp prototrophy; f1 ori= phagic replication origin; ColE1 ori= bacterial replication origin; Cam=gene for chloramphenicol resistance; 2 micron= yeast replication origin. Restriction sites are indicated in the polylinker. BTM116 FOR/back= sequencing primers. The plasmid contains a NLS within LexA. **(B) pLinker220 plasmid map and expression antibody cassette.** ADH1 promoter= Alcohol Dehydrogenase promoter (yeast); ADH1 term= alcohol dehydrogenase terminator; LEU2= LEU2 functional gene, restore Leu prototrophy; bla= beta lactamase, Amp resistance; ColE1 ori= bacterial replication origin; 2 micron= yeast replication origin; NLS= Nuclear Localisation Signal. Polylinker: Yellow= BssHII restriction site (RS); green= NheI restriction site. Between highlighted RS is shown a sample intrabody sequence in the scFv format ("DIQ…VSS"; linker between VL and VH is underlined). After NheI is shown VP16 sequence (in red). Initial translated M is in bold.

Plasmid for Functional Analysis pFA6a-KANMX6 (Addgene #39296). The plasmid for functional analysis (pFA) was used for both UPF1 and SER2 genes deletion. It contains the dominant marker *kan* gene from the *Escherichia coli* transposon Tn903 (*KANMX*; confers resistance to G418)[251] flanked by the *Ashbya gossypii TEF* promoter and terminator [252].This allows for expression of the marker in *S. cerevisiae* and virtually eliminates the possibility for

integration at random sites in the yeast genome since there are no homologous sequences. Sites in the plasmid backbone of pFA allow for annealing of forward (Forw) and reverse (Rv) primers for PCR amplification of the module **[Figure 3.3].** In this way, a common set of primers can be used to delete a gene with different markers. The site of integration is dependent on the gene-specific sequence in the 40-60 bp primer tails (shown in black) since there is no homology between the module and the yeast genome. Longer region of homology ~200 bp were used in this work to significantly enhance the targeting efficiency.

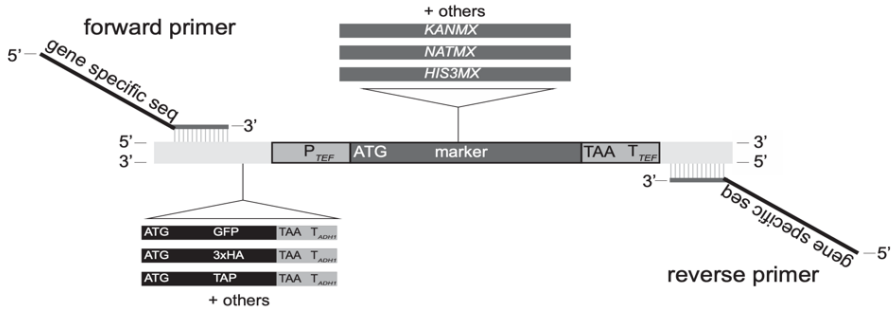


Figure 3.3. Schematic showing the modular design of plasmids for functional analysis (pFA) that was used for gene deletion.

pRSII402 vector (Addgene #35434)**.** The pRSII402 vector is an integration ADE2-carrying plasmid, lacking a yeast origin of replication [253]; It was employed to integrate the *Mb*-AcKRS3/tRNA$_{CUA}^{Pyl}$ pair in the ADE2 locus of the L40 yeast strain. Common features include the Ampicillin resistance and ADE2 gene, which permit L40 strain to acquire prototrophy for Adenine and to be selected in adenine-lacking media **[Figure 3.4 A].**

pRSII422 vector (Addgene #35462). The episomal version of the same vector (pRSII422**)** contains a 2μ circle replication origin and the cis-acting STB (stability) locus [254] **[Figure 3.4 B].** In the pRSII422 plasmid were subcloned all the synthetases/tRNAs pairs tested in this study **[see Table 3.2]**

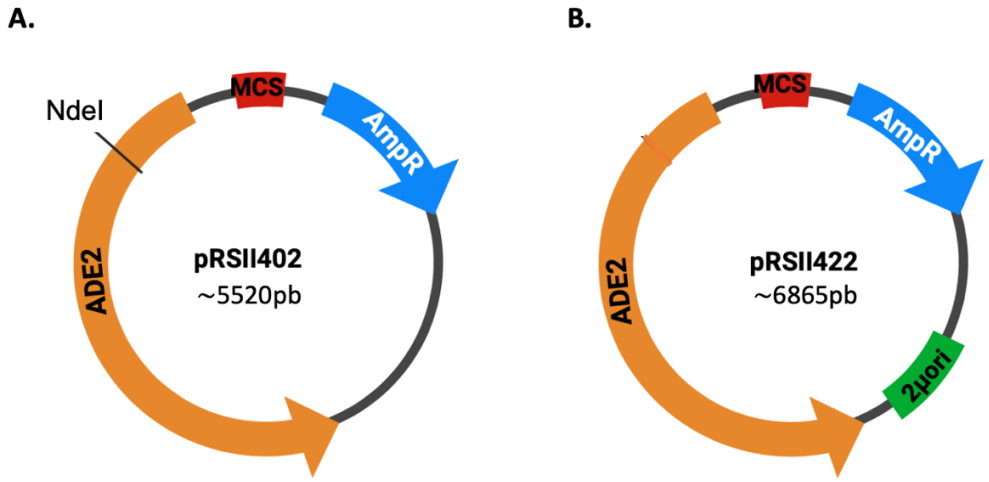


Figure 3.4. **(A) pRSII402 vector.** The integration plasmid pRSII402 was purchase from Addgene (Plasmid #35434); it is an integration ADE2-carrying plasmid, lacking a yeast origin of replication. Common features include the Ampicilline resistance and ADE2 gene, which permit L40 strain to acquire prototrophy for Adenine and to be selected in adenine-lacking media. **(B) pRSII422 vector**. pRSII422 was purchase from Addgene (Plasmid ##35462). It is an episomal ADE2-carrying plasmid, bearing 2μ yeast origin of replication. Common features include the Ampicilline resistance and ADE2 gene, which permit L40 strain to acquire prototrophy for Adenine and to be selected in adenine-lacking media.

3.1.4 Yeast strains

Yeast L40 strain.

Every experiment employed L40 strain of S.cerevisiae: mat-a hisΔ200 trp1-901 leu2-3, 112 ADE2, LYS2::(lexAop)4-HIS3 URA3::(lexAop)8-lacZ Gal4 [224].
This yeast strain is auxotroph for Tryptophan, Leucine and Adenine and will not grow in absence of these compounds. LYS2 and URA3 genes are disrupted by the LexA operator, but they are still functional (capital letters in the genotype). In PISA selection, for instance, bait plasmids contain the TRP1 gene and prey plasmids carry the LEU2 gene: interaction detection between prey and bait occurs on histidine-deprived media because the HIS3 gene transcription depends on LexA promoter activation. The ADE2 locus carries on a A112T transversion mutation in the 5′ of the gene which makes the protein inactive.

3.1.5 Generation of Yeast tSL40 strain: L40 stably expressing *Mb*AcKRS3/tRNA$_{CUA}^{Pyl}$ amber suppression pair

In this chapter all the required steps and constructs for the realization of yeast stable line will be detailed.
In order to create stable lines, L40 strain was transformed with the integration vector pRSII402 carrying *Methanosarcina barkeri* (*Mb*)-AcKRS3/*Mb*tRNA$^{Pyl}_{CUA}$ gene cassette. The final integration vector pRSII402- *Mb*AcKRS3/*Mb*tRNA$^{Pyl}_{CUA}$ was created with the insertion of the tRNA and synthetase cassette in the multi-cloning site via restriction enzyme (RE) ligation method. In addition, another couple of RE sites was designed in order to insert synthetases suitable for other UAAs.
The *Saccharomyces Cerevisiae (Sc)*-tDNA$_{UCU}^{Arg}$–*Mb*tDNA$_{CUA}^{Pyl}$ cassette containing the natural 5′-, 3′-, and 10-base pair linker sequence was amplified with High Fidelity PCR from pRS426 vector (provided by Dr. Chin Lab MRC-LMB) with appropriate couple of primers carrying protruding ends (Forw 5′-CGA GCT CAC CGG TAT CGA TGT GTG TTA T-3′ and Rv 5′-GGA CTA GTG CTA GCG TTA ACA AAG CAT TAA TG-3′) that have the SacI /SpeI sites to obtain a directional subcloning of the fragment into the SacI/SpeI sites of pRSII402, upstream from the synthetase **[Figure 3.5].**
Customized gene synthesis of the *Mb*AcKRS3 synthetase variant was chosen. A yeast codon-optimazed construct encoding for N-terminus Myc tagged *Mb*AcKRS3 synthetase variant was designed and cloned it into the SpeI/HindIII restriction sites present in the MCS of pRSII402 vector, downstream from tRNAs. ADH1 minimal promoter followed by yeast-codon-optimized Kozak sequence, and its terminator sequences were added to the AcKRS3 gene **[Figure 3.5].**

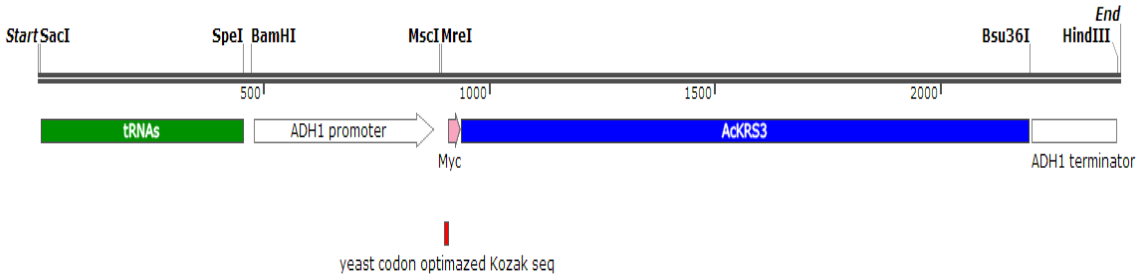


Figure 3.5. tRNAs/synthetase cassette: tRNAs are cloned in the multi-cloning site of PRSII402 via ligation between the restriction enzyme SacI and SpeI and the synthesized DNA cassette, encoding the synthetase AcKRS3, between BamHI and HindIII.

The only way pRSII402-*Mb*AcKRS3/*Mb*tRNA$_{CUA}^{Pyl}$ vector can be propagated in yeast is by integration into the host genome at *ADE2* gene locus via homologous recombination, introducing in the process both the required protein to revert auxotrophy of L40 and the *Mb*AcKRS3/*Mb*tRNA$_{CUA}^{Pyl}$ gene cassette. The ADE2 locus of L40 carries on A12T transversion mutation in the 5′ of the gene which makes the protein inactive in L40 strain. The recombination process will produce two copies of the gene, recombining the 5′-3′ portions of the genes with each other; this will lead to one functional copy of the ADE2 gene on the yeast genome and a non-functional one (still carrying the mutations present in the ADE2). To provide a guarantee for this type of 5′-3′ recombination, the vector was digested with NdeI at the 3′ terminus of ADE2 marker bearing on it [255] **[Figure 3.6 A].**

1 µg of pRSII402 vector digested with 1 enzyme unit (U) of NdeI (NEB) at 37°C for 1 h in the appropriate enzyme buffer [Figure]. DNA is then purified with KIT QUIAGEN MINelute. 200ng of digested pRSII402-*Mb*AcKRS3/*Mb*tRNA$_{CUA}^{Pyl}$ were transformed in L40 using the Lithium Acetate method. Selection on transformed cells was performed with adenine deprivation. The resultant strain able to genetically encoding N-acetyl-Lysine was designated tSL40 or “the acetylator”.

To verify that the recombination has taken place the integrating construction was sequenced to check the duplication of *ADE2* gene in yeast genome, called to simplify ADE2#1 and ADE2#2. ADE2#1 consists of 5′ region of genomic locus and 3′ of vector ADE2 marker, carrying on A-T transversion mutation; ADE2#2 vice versa consists of the 5' vector WT ADE2 sequence, so it is functional. Genomic DNA template from one surviving clone has been used for the initial PCR reaction (see Material and Methods) with appropriate couple of primers. Forw ADE2#1 the forward primer complementary to the region upstream genomic ADE2 locus 5′-

AGCAGATTGTACTGAGAGTGCACCATAGATCTGAATTAA-3′ and the reverse to vector sequence upstream ADE2 marker 5′-CTATGTATGAAGTCCACATTTGATGTAATCATAACAAAGCC-3′; for ADE2#2 Forw matches to the vector 5′-CAATCAAGAAAAACAAGAAAATCGGACAAAACAATCAAGT-3′ and Rv to the genomic region downstream ADE2 locus 5′-TGCGGTATTTCACACCGCATAGATCTTATGTATGAAATT -3′. After purification, PCR fragments were sequenced using Sanger method ("genome walking" approach) (cfr. BLAST Alignment Appendix II) **[Figure 3.6].**

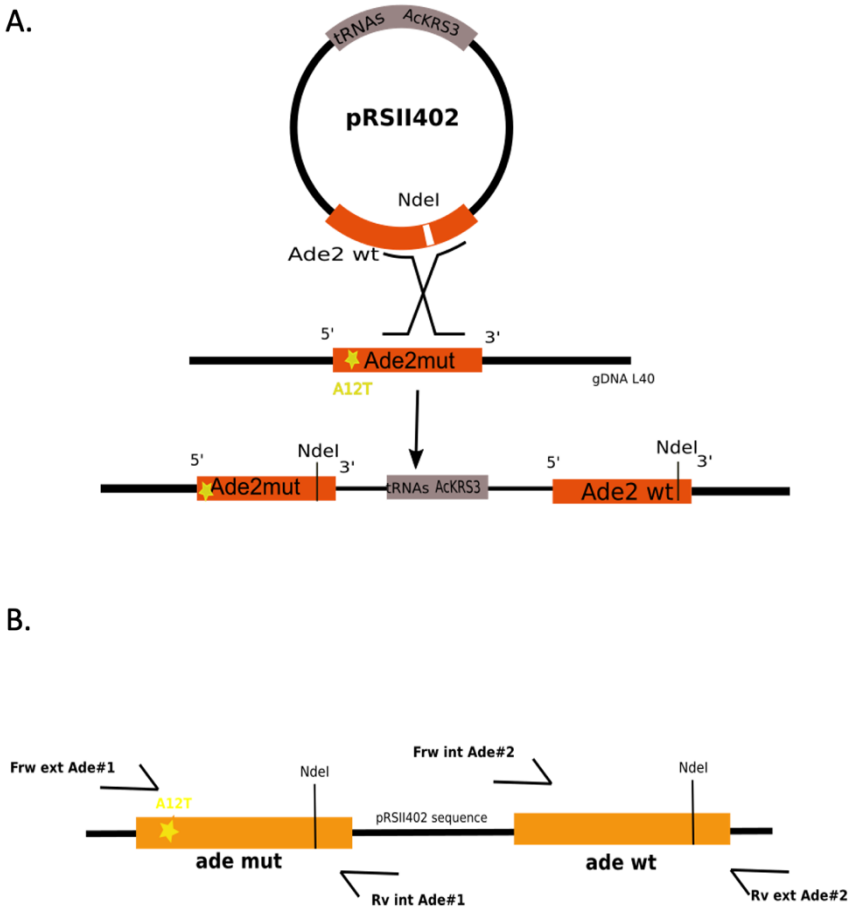


Figure 3.6. (A) Recombination of an integrating plasmid with the yeast genome. The recombination event is directed by cleaving the transforming DNA, leading to a direct duplication in the chromosome. ***(B)* Sequencing scheme of pRSII402 vector integration.** For ade#1 amplification, the forward primer is complementary to the region upstream genomic ADE2 locus and the reverse to vector sequence upstream ADE2 marker; for ade#2 Forw matches to the vector and Rv to the genomic region downstream ADE2 locus.

3.1.6 Generation of tSL40ΔUPF1 Yeast strain

The Nonsense-Mediated Decay (NMD) pathway deficient strain was made by one-step UPF1 gene replacement using a dsDNA fragment that contains ends of the target gene (homology regions required to target the integration), flanking a selectable marker. The KANMX marker module lacking homology with any region of the *S. cerevisiae* genome was chosen: it contains the *kan* gene from the *Escherichia coli* transposon Tn903 (*KANMX*; confers resistance to G418) flanked by the *Ashbya gossypii TEF* promoter and terminator [256].
The genomic regions containing about 200 bp upstream of UPF1 and about200 bp downstream of UPF1 were amplified from tSL40 genomic DNA preparation using the following primers (Forw UP SalI, 5'- gcaggTCGACAATGAAAAGCTTACCAGAAACTTA-3' / Rv UP BamHI, 5'-CATCAATCATTGTCATTATCAAggatccccgggt-3' and Forw DW SacI, 5'-taaacgagctCTTCGGTGAACCCTGTTAAAATAAAATGTTAAAC-3'/Rv DW SpeI, 5'-GGGCATGGACTTGATATCCTAGCCActagtggccta-3'). Resulting PCR products named 200pbUP and 200pbDW were digested with the proper REs and cloned via ligation method in the pF6a vector at the 5' and 3' of the KANMX cassette respectively.
The final cassette containing 200 bp upstream of UPF1, the KANMX, and 200 bp downstream of UPF1 was amplified using primer Forw 5'-AAT GAA AAG CTT ACC AGA AAC TTA CG-3' and primer Rv 5'-GGC TAG GAT ATC AAG TCC ATG CC-3'. 50 µl of PCR product was transformed into tSL40 yeast strain using the lithium acetate method. Transformed cells were plated on YPAD plates at 30°C for 6h and then replica plated on G418 YPAD plates for selection. Transformants were visible in 3-4 days at 30°C. 5-50 transformants are expected.
After transformation and selection, integration of the cassette and concurrent deletion of the chromosomal target sequence was first analyzed by performing colony PCR to amplify the locus of the integration using with primers 300 bp away from the UPF1 gene (Frw 300up 5'-GAT TTG GGA GGG ACA CCT TTA TAC GC-3', Rv 300 Dw 5'-TTC ATT AGA AGT ACA ATG GTA GCC C-3').
Finally, the PCR products were sequenced to confirm that the UPF1 gene was replaced with the KANMX6 through homologous recombination. The resultant UPF1Δ strain was designated as tSL40ΔUPF1 strain.

3.1.7 Generation of L40ΔSER2 Yeast strain

Phosphoserine Phosphatase-deficient strain (L40ΔSER2) was created by knocking out the *SER2* gene from the L40 yeast genome using a DNA fragment that contains ends of the target gene (homology regions required to target the integration), flanking the KANMX selectable marker.

This strain was employed to express the phospho-serinyl-RS/tRNA$_{CUA}^{Sep}$ (SepRS/tRNA$_{CUA}^{Sep}$) pair.
The genomic regions containing about 200 bp upstream of SER2 and about 200 bp downstream of SER2 were amplified from L40 genomic DNA preparation using the following primers (Forw UP, 5′-cccggccagcTGACAACGTTTCTCGAATCAGG-3′ / Rv UP, 5′-gcctccatgtcTTTTTACAGCTGTCGTGCTTAATG-3′ and Forw DW, 5′-tggtcgctatactgAGAAAATAATGATAGATAGATGTAATAG-3′/ Rv DW, 5′-tatcgaatcgacagCATGCCTACCCTAATCTTCTAC-3′). Resulting PCR products named 200 pb UP and 200 pb DW were subcloned using Hi-Fi assembly kit from NEB in the pF6a vector at the 5′ and 3′ of the KANMX cassette respectively.
The final cassette containing 200 bp upstream of *SER2*, the KANMX, and 200 bp downstream of *SER2* was amplified using primer Forw UP / Rv DW. 50 µl of PCR product was transformed into L40 yeast strain using the Lithium Acetate method. Transformed cells were plated on YPAD plates at 30°C for 6h and then replica-plated on G418 YPAD plates for selection. Transformants were visible in 3-4 days at 30°C. 5-50 transformants are expected. To confirm that the SER2 gene was replaced with the KanMX6 through homologous recombination I opted to directly measure the increase in the intracellular level of phospo-serine in one of the survived cloned compared to the WT L40 **[see Table 3.4]**

3.1.8 Yeast Genomic DNA preparation

The following protocol (Auble Lab) is a reliable and low-cost genomic DNA extraction protocol from yeast cells for the variety of PCR-based applications described in this thesis (including colony-PCR, RT-PCR onto tDNAs, and DNA sequencing of ADE2) and for amplification of DNA fragments at least up to 3500 base pairs. This method is especially powerful for simultaneous analysis of a large number of samples. DNA can be efficiently extracted from different yeast species. The protocol involves:

Required reagents:
Sorbitol Buffer: 1 M sorbitol, 0.1 M sodium citrate (pH 5.8), 0.01 M EDTA (pH 8.0)
Saline-EDTA Buffer: 0.15 M NaCl, 0.1 M EDTA (pH 8.0), 0.1% SDS
Lyticase (or Zimolase)
25 ml of yeast cells were grown in YPD to 5x10^7 cells/ml. Cells were centribuged in a 50-ml conical tube at 1,000 x g for 5 min and resuspended in 10 ml of sterile distilled (SD) water, then spinned again at 1,000 x g for 5 min. After resuspension in 5 ml of SD water, 100 µl of β-mercaptoethanol was. Samples were incubated at room temperature for 15 min then centrifuged at 1,000 x g for 5 min and resuspended in 5 ml of sorbitol buffer. ~1.0 mg of Lyticase

(or 0.25 mg Zymolase) were added and incubated at 37°C for 1 hour (Check efficiency of spheroplasting after 30 min and if less than 80% incubate another 30 min). The spheroplasts were spinned at 1,000 x g for 3 min and resuspended in 1.2 ml Saline-EDTA Buffer (divide sample into two tubes). Phenol:Chloroform extraction was repeated 3 times (make sure to avoid interface). Immediately after, Ethanol precipitation (adding 1/10 the volume of 3 M sodium acetate and 3X the volume of ethanol) was performed. The sample was placed at -20°C for 60 min. The sample was then centrifuged for 30 min at maximum speed at 4°C. Final Wash with 70% Ethanol was performed and the dried pellet was resuspended in 100 μl of SD water.

3.1.9 Yeast RNA extraction

Acid guanidinium thiocyanate-phenol chloroform extraction method (TRIzol RNA Isolation Reagents, Thermo Fisher Scientific) was used to recover total RNA from yeast samples. TRIzol was also used for isolating small RNAs, such as tRNAs.
25 ml of yeast cells were grown in YPD to $5x10^7$ cells/ml. Cells were centrifuged in a 50-ml conical tube at 1,000 x g for 5 min and resuspended in 10 ml of sterile distilled (SD) water, then spin down again at 1,000 x g for 5 min. Cell lysis was performed with 1 ml TRIzol Reagent by repetitive pipetting. Samples were subsequently incubated for 5 min at RT to permit the complete dissociation of nucleoprotein complexes. In the "phase separation" step, 0.2 ml of chloroform per 1 ml of TRIzol Reagent was added. Samples were mixed vigorously for 15 seconds and incubated RT for 2 to 3 min. The samples were then centrifuged at no more than 12,000 x g for 15 min. The upper aqueous phase was transferred carefully without disturbing the interphase into a fresh tube. Then the RNA precipitation phase follows. 0.5 ml of isopropyl alcohol per 1 ml of TRIzol Reagent was used for the initial homogenization. Samples were incubated at 30°C for 10 min and centrifuged at not more than 12,000 x g for 10 min at 4°C. The RNA pellet was washed once with 75% ethanol. The samples were mixed by vortexing and centrifuged at no more than 7,500 x g for 5 min at 4°C. The above washing procedure was repeated once more. The air-dried RNA pellet was dissolved in DEPC-treated water.

The RNA quality was first assessed on DEPC/MOPS/formamide 1% agarose gel where it is possible to distinguish two main bands representing the 28S and 18S ribosomal RNAs, but also low molecular weight RNAs (for example tRNAs and 5S rRNA) could be detected. RNA was finally quantified using Nanodrop spectrophotometer.

3.1.10 Yeast Culture Reagents and Media

Wild type L40 yeast strain and L40ΔSER2 are grown at 30 °C in YPD medium, whereas the stable yeast lines tSL40 and tSL40ΔUPF1 (with integrated plasmid carrying ADE2 marker) are grown in SD medium ad hoc, without adenine in the amino acid mix.
Preservation of yeast strain is ensured by glycerol stocks (GS) (33% (v/v) glycerol and medium and stored at -80°C. Yeast is streaked onto fresh plates from glycerol stock 3 day before usage. L40 colonies should appear pinkish because of the accumulation of ade2 red precursor, while integrated L40 ones should be white and wet and grow to >2 mm in diameter. It is recommended to prepare fresh plates for experiments and store them at 4°C no longer than 2 weeks. Use yeast recovered from GS. All material codes can be found in Visintin et al 2004 [95].

YPD medium

10 g yeast extract (BD # 211931)
20 g bacto peptone (BD # 211840)
20 g bacto-agar (BD # 214010)
2% glucose (SIGMA # G-7021)
Add H2O to 950 ml. Adjust pH to 5.8 then adjust to 1 liter. Autoclave 121°C for 15 min.

YC medium

1. YNB w/o aa & (NH4)2SO4:

1.2 g yeast nitrogen base, w/o amino acids and ammonium sulfate (SIGMA # Y-1251)
20 g bacto-agar (BD # 214010) for YC plate only (also referred to as SD plate)
Add H2O to 800 ml. Autoclave 121°C for 15 min.

2. SALTS:

5.4g NaOH (Riedel-de Haen #30620)
10g succinic acid (SIGMA # S-7501)
5g ammonium sulfate (Fluka # 09980)
22g D-glucose (SIGMA # G-7021)
Add H2O to 100 ml and dissolve all components one by one and H2O to obtain a final volume of 150 ml.

3. amino acids MIX

5.8 g NaOH (Riedel-de Haen #30620)
1 g Adenine hemisulfate salts (SIGMA#A-9126)
1 g L-Arginine HCl (SIGMA # A-5131)
1 g L-Cysteine (SIGMA # C-6852)
1 g L-Threonine (SIGMA # T-8625)
0.5 g L-Aspartic acid (SIGMA # A-4534)

0.5 g L-Isoleucine (SIGMA # I-2752)
0.5 g L-Methionine (SIGMA # M-9625)
0.5 g L-Phenylalanine (SIGMA # P-2126)
0.5 g L-Proline (SIGMA # P-0380)
0.5 g L-Tyrosine (Sigma # T-3754)
0.5 g L-Serine (SIGMA # S-4500)
Dissolve in 80 ml H2O

Yeast Transformation reagents

50% PEG

250 g Polyethylene glycol, avg. mol. wt.=3350 (Fluka # 95904)
Add H20 up to 500 ml. Filter sterilize. Solution must be kept in a tightly sealed glass bottle to avoid evaporation.

1M LiOAc

25.5 g Lithium acetate dehydrate (SIGMA # L-6883)
adjust to pH 7.5 with Glacial acetic acid, volume to 250 ml, autoclaved.

10X TE

1.21 Tris
0.37 g EDTA
adjust to pH 7.5 with HCl, volume to 100ml, autoclaved.

LiOAc-TE-PEG solution (prepare freshly for immediate use)

10 ml 1M LiOAc
10ml 10XTE
80 ml 50%PEG

3.1.11 Yeast LiOAc transformation

Here is described an efficient protocol for chemical transformation that is able to generate more than 10^6 transformants. It can be used, thus, to create stable yeast lines and for the initial transformation of the bait plasmid. Following protocol is a modification of Visintin et al 2004[95].
Required Materials
10X LiOAc buffer: 1M LiOAc, pH 7.5 adjusted with diluted glacial acetic acid
autoclave-sterilized
50% (w/v) autoclave-sterilized PEG 4000:
10X TE buffer: 100mM Tris, 10mM EDTA, pH7.5, autoclave-sterilized

100% DMSO: dimethyl sulfoxide (Fluka #41639)
10 mg/ml denatured carrier DNA from fish sperm (Roche # 1467140)

1. Inoculate 1-3 single colonies in 50 ml of medium with a sterile loop. Grow the culture overnight at 30°C, shaking 220 r.p.m to place the culture at mid log phase the next day (O.D600 > 1.5).
2. Measure OD600 from 10-fold YC dilution. Inoculate the necessary volume of O/N culture to obtain 300 ml of OD600=0.3 culture. Grow until the culture reaches OD600=0.6. Important: absorbance should never exceed 0.7, otherwise efficiency will dramatically drop down. Transform between 0.5-0.6.

3. Pellet down the culture at 1,000 x g for 5 min

4. Remove supernatant (also by pouring) and resuspend pellets in 10 ml of sterile ddH20.

5. Centrifuge again at 1,000 x g for 5 min

6. Remove the supernatant and resuspend the pellet in 1ml or less (if you cultured less than 300ml calculate you need 10ml culture/100ul) of TE/ LiOAc solution (fresh prepared, on the same day). These are competent cells, keep them on ice. They are stable for at least 1h. Use freshly made transformation solutions because this will strongly affect your transformation efficiency.

7. Put 100 µl of culture into an eppendorf tube containing (each tube=one transformation):

   a. 100 µg of denatured salmon sperm DNA (100°C for 10min)
   b. 200 ng of bait plasmid

8. Vortex for 30sec
9. Add 600 µl of freshly made TE-PEG-LiOAc solution to each tube
10. Incubate at 30°C for 30min with shaking 220 r.p.m.
11. Add 70 µl of DMSO. DO NOT VORTEX, but mix gently by inversion.
12. Incubate for 15 min in a water bath at 42°C
13. Immediately put tubes on ice for 2 min
14. Fill the microtube with sterile ddH20 and centrifuge down at 3,000 x g for 5 min, RT.
15. Remove supernatant and resuspend in 1ml of sterile TE 1x.
16. Plate out 100-200 µl on YC selective plates and incubate 3d at 30°C.

These plates are stable for a week if stored at 4°C and sealed with parafilm.

Creation of master plates: from there, pick single clones (testing more than one is needed, because different clones have different expression levels and bait activity behaviors) and restreak on new YC plates (I suggest by resuspending in TE 1x and plating out again).

Creation of glycerol stock: master plates can be stored and used for 2 weeks at 4°C for every experiment; however, it is strictly necessary to create a GS from the master plate as soon as possible, that will be our starting point in future.

3.1.12 PISA 2.0 baits production and characterization

pMICBD1-MCS-LexA: N-terminal LexA fusion creation. A 1076 bp HindIII– HindIII fragment of plasmid pMICBD1 **[Figure 3.2],** contains a truncated ADH promoter, the LexA coding region, a polylinker and an ADH terminator. The *Hin*dIII–*Pst*I fragment containing the LexA ORF and the multicloning site was replaced by a fragment generated by PCR and encoding a multicloning site followed by the LexA ORF and a stop codon. The multicloning site allows the generation of fusions at the N-terminus of LexA according to the sequence: GAA TTC GGA TCC GTC GAC GGT GGT ATG, where ATG is the original LexA initiation codon. Notice that (i) *Eco*RI, *Bam*HI and *Sal*I are available as cloning sites and (ii) two glycines are used as spacers between the fused proteins. The resulting plasmid was named pMICBD1-MCS-LexA.

PISA 2.0 bait construction

H3 Histone-LexA bait was created by amplifying H3 from pDG plasmids described in Guo et al [105] and cloning it at the C-terminus of LexA protein in pMICBD1 plasmid using EcoRI-BamHI restriction sites or Gibson assembly. HA tag has been introduced successively in the NheI site by ligating in vitro annealed oligonucleotides (from Integrated DNA Technologies). The H3 Histone baits bearing the TAG stop codon either at position 9 or 27 were generated using site- directed mutagenesis primers (with the desired mutation). Syn-LexA bait was created by amplifying the a-synuclein coding sequence from pMICBD1-synuclein and subcloning into the SalI restriction sites of pMICBD1-MCS-LexA. The yeast monomeric GFP ymUKG1 coding sequence was synthesized (from Integrated DNA Technologies) and cloned in frame with LexA in pMICBD1. The LexA-ymUGK1150* bearing the TAG at position 150 was obtained using site-directed mutagenesis primers. All constructs were verified by colonyPCR and sequencing. Complete list of bait sequences can be found in **Table 3.1**.

Table 3.1. Sequence of PISA 2.0 bait used in this study.

>LexA-H3-HA MKALTARQQEVFDLIRDHISQTGMPPTRAEIAQRLGFRSPNAAEEHLKALARKGVIEIVSGASRGIRLLQEEE EGLPLVGRVAAGEPLLAQQHIEGHYQVDPSLFKPNADFLLRVSGMSMKDIGIMDGDLLAVHKTQDVRNG QVVVARIDDEVTVKRLKKQGNKVELLPENSEFKPIVVDLRQQSFTIEGLAVGVIRNGDWLEFRTKQTARKST GGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTEGSYPYDVPDYA*
>LexA-H3K9TAG-HA MKALTARQQEVFDLIRDHISQTGMPPTRAEIAQRLGFRSPNAAEEHLKALARKGVIEIVSGASRGIRLLQEEE EGLPLVGRVAAGEPLLAQQHIEGHYQVDPSLFKPNADFLLRVSGMSMKDIGIMDGDLLAVHKTQDVRNG QVVVARIDDEVTVKRLKKQGNKVELLPENSEFKPIVVDLRQQSFTIEGLAVGVIRNGDWLEFRTKQTAR*ST GGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTEGSYPYDVPDYA*
>LexA-H3K27TAG-HA MKALTARQQEVFDLIRDHISQTGMPPTRAEIAQRLGFRSPNAAEEHLKALARKGVIEIVSGASRGIRLLQEEE EGLPLVGRVAAGEPLLAQQHIEGHYQVDPSLFKPNADFLLRVSGMSMKDIGIMDGDLLAVHKTQDVRNG

QVVVARIDDEVTVKRLKKQGNKVELLPENSEFKPIVVDLRQQSFTIEGLAVGVIRNGDWLEFRTKQTARKST
GGKAPRKQLATKAAR*SAPATGGVKKPHRYRPGTVALREIRRYQKSTEGSYPYDVPDYA*

>HA-H3K9TAG-LexA

MYPYDVPDYAGSRTKQTAR*STGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTER
ITVVMKALTARQQEVFDLIRDHISQTGMPPTRAEIAQRLGFRSPNAAEEHLKALARKGVIEIVSGASRGIRLL
QEEEEGLPLVGRVAAGEPLLAQQHIEGHYQVDPSLFKPNADFLLRVSGMSMKDIGIMDGDLLAVHKTQDV
RNGQVVVARIDDEVTVKRLKKQGNKVELLPENSEFKPIVVDLRQQSFTIEGLAVGVIRNGDWL*

>α-Synuclein-LexA

MDVFMKGLSKAKEGVVAAAEKTKQGVAEAAGKTKEGVLYVGSKTKEGVVHGVATVAEKTKEQVTNVGG
AVVTGVTAVAQKTVEGAGSIAAATGFVKKDQLGKNEEGAPQEGILEDMPVDPDNEAYEMPSEEGYQDYEP
EAVDGGMKALTARQQEVFDLIRDHISQTGMPPTRAEIAQRLGFRSPNAAEEHLKALARKGVIEIVSGASRGI
RLLQEEEEGLPLVGRVAAGEPLLAQQHIEGHYQVDPSLFKPNADFLLRVSGMSMKDIGIMDGDLLAVHKT
QDVRNGQVVVARIDDEVTVKRLKKQGNKVELLPENSEFKPIVVDLRQQSFTIEGLAVGVIRNGDWL*

>LexA-ymUKG1

MKALTARQQEVFDLIRDHISQTGMPPTRAEIAQRLGFRSPNAAEEHLKALARKGVIEIVSGASRGIRLLQEEE
EGLPLVGRVAAGEPLLAQQHIEGHYQVDPSLFKPNADFLLRVSGMSMKDIGIMDGDLLAVHKTQDVRNG
QVVVARIDDEVTVKRLKKQGNKVELLPENSEFKPIVVDLRQQSFTIEGLAVGVIRNGDWLEFMVSVIKEEMK
IKLHMEGNVNGHAFVIEGDGKGKPYDGTQTLNLTVKEGAPLPFSYDILTNAFQYGNRAFTKYPADIPDYFK
QTFPEGYSWERTMSYEDNAICNVRSEISMEGDCFIYKIRFDGKNFPPNGPVMQKKTLKWEPSTEMMYVRDG
FLMGDVNMALLLEGGGHHRCDFKTSYKAKKVVQLPDAHKIDHRIEILSHDRDYSKVKLYENAVARNSLLP
SQASK

>LexA-ymUKG1-150TAG

MKALTARQQEVFDLIRDHISQTGMPPTRAEIAQRLGFRSPNAAEEHLKALARKGVIEIVSGASRGIRLLQEEE
EGLPLVGRVAAGEPLLAQQHIEGHYQVDPSLFKPNADFLLRVSGMSMKDIGIMDGDLLAVHKTQDVRNG
QVVVARIDDEVTVKRLKKQGNKVELLPENSEFKPIVVDLRQQSFTIEGLAVGVIRNGDWLEFMVSVIKEEMK
IKLHMEGNVNGHAFVIEGDGKGKPY*GTQTLNLTVKEGAPLPFSYDILTNAFQYGNRAFTKYPADIPDYFK
QTFPEGYSWERTMSYEDNAICNVRSEISMEGDCFIYKIRFDGKNFPPNGPVMQKKTLKWEPSTEMMYVRDG
FLMGDVNMALLLEGGGHHRCDFKTSYKAKKVVQLPDAHKIDHRIEILSHDRDYSKVKLYENAVARNSLLP
SQASK

3.1.13 Orthogonal-aaRS/tRNA pair constructs:

pRSII422-*Mb*AcKRS3/*t*RNA$_{CUA}^{Pyl}$.The *Mb*AcKRS3 gene block (containing the ADH1 regulatory element) was PCR-amplified and cloned into the pRSII422 vector using Gibson Assembly. The resulting plasmid was then double restriction enzyme digested with SacI and SpeI that flank a stuffer region. The *Sc*tRNA$_{UCU}^{Arg}$-tRNA$_{CUA}^{Pyl}$ region of plasmid pRS426 (gifted by the Chin Lab at the UK Medical Research Council Laboratory of Molecular Biology) was inserted in the digested pRSII422. The plasmid containing the correctly assembled insert pair was named pRSII422-*Mb*AcKRS3/tRNA$_{CUA}^{Pyl}$.
All the other pRSII422 series of plasmid were constructed starting from the pRSII422-*Mb*AcKRS3/tRNA$_{CUA}^{Pyl}$ using Gibson assembly as follows

pRSII422-*Mm*AcKRS/tRNA$_{CUA}^{Pyl}$. The gene for *M. mazei* synthetase was amplified from pRST-*Mm*PylRS/tRNA$_{CUA}^{Pyl}$ plasmid (gifted by Chin Lab) and cloned to replace *the M.barkeri* synthetase, yielding plasmid pRSII422-*Mm*PylRS/tRNA$_{CUA}^{Pyl}$. The *Mm*PylRS that aminoacylate *Mm*tDNA$_{CUA}^{Pyl}$ with *N*-acetyl-L-lysine (AcKRS) was created by transferring mutations (L266M/L270I/Y271F/L274A/C313F) identified in *M.barkeri* into the *Mm*PylRS gene using site-directed mutagenesis primers.

pRSII422-MeHisRS/tRNA$_{CUA}^{Pyl}$pairs. The *Methanomassiliicoccus luminyensis*, *Methanomassiliicoccaceae archaeon* RumEn M1, *Methanomassiliicoccaceae archaeon* PtaU1.Bin030 PylRS synthetases bearing (L121M, L125I, Y126F, M129A, V168F) mutations were amplified from pRST plasmids (gifted by Chin Lab) and subcloned in pRSII422 to replace the *M.barkeri* synthetase. Cloning of their cognate tRNA proceeded similarly: each one of the tRNA sequences was embedded into the amplifying primers and cloned to replace *Mb*tDNA$_{CUA}^{Pyl}$ via Gibson Assembly. The final vectors were called *Ml*MeHisRS/tRNA$_{CUA}^{Pyl130C}$, *Ma*RumMeHisRS/tRNA$_{CUA}^{Pyl6B03}$, *Ma*Bin03MeHisRS/tRNA$_{CUA}^{Pyl6C}$.

pRSII422-*Mmp*SepRS/tRNA$_{CUA}^{Sep}$. The gene for *Mmp*SepRS synthetase was amplified from pRST- *Mmp*SepRS/tRNA$_{CUA}^{Sep}$ plasmid (gifted by Chin Lab) and cloned into pRSII422 to replace *the M.barkeri* synthetase together with its cognate tRNA in one-step Gibson reaction (as previously described).

pRSII422-*Ec*TyrRS/*Ec*tRNA$_{CUA}^{Tyr}$ described in Hancock et al 2010 was kindly provided by Dr. Susan Hancock.

Complete List of orthogonal aminoacyl-tRNA synthetases and tRNAs used in this study (in yeast) can be found in **Table 3.2.**

Table 3.2. Sequences of aminoacyl -RNA syntheses and tRNAs used in this study

>*Mb*AcKRS3 MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKTCKR CRVSGEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVSAKASTNTSRSVPSPAKSTPNS SVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDREDYLGKLERDITK FFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMMAPTIFNYARKLDRILPGPIKIFEVGP CYRKESDGKEHLEEFTMVNFFQMGSGCTRENLEALIKEFLDYLEIDFEIVGDSCMVYGDTLDIMHGDLELSS AVVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRSESYYNGISTNL*
>MmPylRS MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRHHKYRKTCKR CRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVST QESVSVPASVSTSISSISTGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFR ELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDK NFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLESIITDFL NHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDFKNIK RAARSESYYNGISTNL*
>MmpSepRS MFKREEIIEMANKDFEKAWIETKDLIKAKKINESYPRIKPVFGKTHPVNDTIENLRQAYLRMGFEEYINPVIV DERDIYKQFGPEAMAVLDRCFYLAGLPRPDVGLSDEKISQIEKLGIKVSEHKESLQKILHGYKKGTLDGDDL VLEISNALEISSEMGLKILEDVFPEFKDLTAVSSKLTLRSHMTSGWFLTVSDLMNKKPLPFKLFSIDRCFRREQK EDKSHLMTYHSASCAIAGEGVDINDGKAIAEGLLSQFGFTNFKFIPDEKKSKYYTPETQTEVYAYHPKLKEW LEVATFGVYSPVALSKYGIDVPVMNLGLGVERLAMISGNFADVREMVYPQFYEHKLNDRNVASMVKLDKV PVMDEIYDLTKELIESCVKNKDLKSPCELAIEKTFSFGKTKKNVKINIFPKFEGKNLLGPSILNEIYVYDGNVIG IPESFDGVKEEFKDFLEKGKSEGVATGIRYIDALCFKITSKLEEAFVSNTTEFKVKVMWVRSLSDINLKIDDIAL KQIMSKNKVIDVRGPVSLNVEVKIE*
>*Ml*MeHisRS MDTRLTPAQAQRIREMGGTVDPSLAFSSEAERESAFQRISADLQGANLAKIRRCAEAPERHPIGSLENTLAC ALAAKGFIEVKTPMMIPADGLVKMGIDESHPLWNQVFWVGPKKALRPMMAPNIFFLARHLRRSVPAPLLL FEIGPCFRKESRGSNHLEEFTMLNLFELAPQADATERLKEHIATVMNAVGLPYELVVEGSEVYGTTIDVEVD GVELASGAVGPLPMDKPHGITEPWAGVGFGLERIALMRTKEQNIKKVGRSLVYVNGARIDI
>*Ma*RumMeHisRS MTIEWTPSQKQRLKELGIDSDQDYTINNIQEREEVFSRLVTRRQSEGRRAIRSMMEHPVRHKLAQLEQDLAQ ALVDDGFLEFRTPTIITRSALEKMGIGREHPLHEQVFWLDEKRCLRPMMAPNIFYVARHLKRNAKGPVKLFE IGTCYRKESHGSNHLEEFTMLNLFELDPAGDAREQLRKHISTIMNTIGLDYELVSCSSDVYVETTDVEVNGVE VASGAIGPHKLDPAHGIKAPWAGVGFGLERLLMLKHGEDNVKKVGRSLIYLQGVRLDI
>*Ma*Bin03MeHisRS MVIEWSPSQKQRLRELGRADEGGMEFETVVERDEAFTKEVAYYQSINRKEIRNIQERRERHLLAKVEENIAE ALIADGFLEVRTPTIISGNALVKMGIDHNHPLREQVFWLDGSRCLRPMMAPNIFFLARHLKRNVRMPLQMF EIGTCYRKESHGSNHLEEFTMLNLFEMASMDDPAVRLRHHIQTVMGAIGLEYELSECESDVYGRTIDVEVN GVEVASAALGPHKLDPAHGITDAWSGVGFGLERLLMVKNAENNIKKVGRSLIYLGGARLDI
>*Ec*TyrRS MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALYCGFDPTADSLHLGHLVPLLCLKRFQQAGHKPVAL VGGATGLIGDPSFKAAERKLNTEETVQEWVDKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRD IGKHFSVNQMINKEAVKQRLNREDQGISFTEFSYNLLQGYDFACLNKQYGVVLQIGGSDQWGNITSGIDLT

RRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTSPYKFYQFWINTADADVYRFLKFFTFMSIEEIN ALEEEDKNSGKAPRAQYVLAEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEMEK GADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLFGRFTLLRRGKKNYCLICWK*
>*Mm*tDNA$_{CUA}^{Pyl}$ gggaacgtgatcatgtagatcgaatggactctaaatccgttcagtggggttagattccccacgtttccg
>*Ma*tDNA$_{CUA}^{Pyl}$ ggaaacgtgatcatgtagatcgaatggactctaaatccgttcagtggggttagattccccacgtttccg
>*Sc*tDNA$_{UCU}^{Arg}$-*Mb*tDNA$_{CUA}^{Pyl}$ cassette accggtatcgatgtgtgttatatgtacctctgctttgcagtataagaaatttacatttatttctgactaataacaccttggtgccccaacggtaaacaacttgtatcagtt ctcataagtgcggccattttatgcaatacaggctgcattatttcaccagccgtgaaaatccgaaaattgtagtaattgaaagcgtaattaggttttactataataaagt agtaaaaccttcaacaaatagtagctcgcgtggcgtaatggcaacgcgtctgacttctaatcagaagattatgggttcgacccccatcgtgagtgctttgtttctgg aaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgatttttggctactcctgtagttattcttcattaatgctttgtta acgctagc
>tDNA$_{CUA}^{Sep}$ gtccgcgtagtctaggggttaggcagcagtctctaaaattgccttacgtgggttcaaatcccaccgcggact
>tDNA$_{CUA}^{Pyl\ 13C10}$ ggtgttctggtccgggaccaccgggcctctaaagccacggttagccgggttcaactcccgggaacatcg
>tDNA$_{CUA}^{Pyl\ 6C10}$ ggtgaactggtccgggaccaccaggcctctaaagccacggttagccgggttcaactcccgggttcatcg
>tDNA$_{CUA}^{Pyl6B03}$ ggtgaactggtccgggaccaccaggcctctaaagccacggctagccgggttcaactcccgggttcatcg
>tDNA$_{CUA}^{Tyr}$ ggtggggttcccgagcggccaaagggagcagactctaaatctgccgtcatcgacttcgaaggttcgaatccttcccccacca

3.1.14 Protocols for Non-canonical-AAs incorporation in *S.cerevisiae*

ncAA acid liquid stocks. All ncAA stocks were prepared at a final concentration of 1 M L-isomer. Deionized water was added to the solid ncAA to approximately 90% of the final volume needed to make the stock, and 6.0 N NaOH was used as needed to fully dissolve the ncAA powder in the water. Water was added to the final volume and the solution was sterile filtered through a 0.2 micron filter. No pH adjustment of additional ncAAs was performed unless otherwise noted. Filtered solutions were stored at -20 °C or at 4 °C for up to four weeks for less labile ncAAs.

Growth culture protocols for ncAAs incorporation. Accordingly with standard protocol [149] the appropriate selective medium supplemented with 10 mM of ncAA was inoculated with a stationary phase culture of the required yeast line to yield an OD600 ≈ 0.2 in a 250 ml flask. Parallel cultures with no ncAA were also prepared as negative controls. Cultures were grown at 30°C in an orbital shaker for 24-48 h.
As for agar plates, 10mM of the ncAA stock solution was added to the proper agar medium (cooled at 55°C) before pouring the plates.

Intracellular ncAA uptake measurment: method for yeast. Following Chin et al.[214], one colony of L40 yeast was grown in both presence and absence of 10mM of ncAAs overnight. Then, the metabolite extraction protocol was performed and the samples were analyzed using liquid chromatography–mass spectrometry (LC–MS).
1. Inoculate 1 colony of L40 wt cell (or L40ΔSER2) in 5 ml YC medium supplemented with 10 mM ncAA in a falcon tube. Agitate cells at 30°C overnight.
2. Measure the OD600 from 10-fold YC dilution. Document OD600 and calculate the total cell count using OD600=1=1,5X10^7 cells
3. Centrifuge the cultures at 3,500 x g for 10 min at 4°C to harvest the cells
4. Resuspend the cells in 1 ml of ice-cold YC media by gently pipetting the cells. Transfer the cell culture to a 2 ml round-bottom centrifuge tube (Note that the round bottom tube is easier for cell lysis in Step 5 compared to the V-shape 1.7 ml tube). Centrifuge the cells at 5,000 x g for 5 min at 4 °C. Repeat this step 2 more times (for 3 washes total).
5. Suspend the final cell pellet in 400 µl of water/methanol (40:60). Add 0.1 cm glass beads to the tube (approximately 300 mg per OD600= 2). Vortex the tubes for 15 min on the capped vortexer to fully lyse the cells.
6. Centrifuge the vortexed suspension at 21,000 x g for 30 min at 4 °C. Pipette the supernatant into a fresh 1.5 ml conical tube.
7. Pipette the supernatant (200 - 300 µl, make sure no glass beads are taken) into a 3K MWCO amicon centrifugal tube (1.5 ml) (The centrifugal tube should be pre-hydrate by centrifuging water). Centrifuge the centrifugal filter tube for 15 min at 14,000 g at 4 °C.
This step is very important. If it is not done, the LC-MS can be easily clogged.
8. Transfer the flow through (NOT the protein, high MW layer) into an LC-MS insert.

9.Analyze the samples by LC-MS in the SIM mode for your UAA. In order to make sure you measure the correct peak, UAA standard in water (100 μM), UAA standard spiked in cell lysate (100 μM in cell lysate obtained from no UAA control) should be run before your samples as positive controls.
10. Analyze five standards to make a calibration curve[12].

ncAA Intracellular Concentration analysis: LC-MS method (SIM). An Agilent 1260 Infinity equipped with an Agilent 6130 Quadrupole LC-MS unit was used for analysis of all samples. A HILIC-Z column (4.6 × 150 mm) equipped with a guard column (Agilent) with a flow rate of 0.5 ml min–1 was used to elute amino acids. Buffer A (10 mM ammonium formate in water) and buffer B (acetonitrile:water (9:1 (vol/vol)) solution with 10 mM ammonium formate) were used for reverse-phase HPLC. Five microliters of amino acid-containing solution were injected and eluted using a gradient of 100% buffer B to 70% buffer B in buffer A over 10 min at 30 °C. The mass spectrometer was set to selected-ion monitoring (SIM) mode. Pure amino acids, purchased from Sigma-Aldrich, were run as standards for comparison. Then, A linear fit of the standard samples was used to determine the concentration of ncAAs in the lysate.

[12] The peak area of the Positive or Negative SIM mode should be used to calculate the concentration based on the calibration curve. The value calculated based on the standard curve is not the real intracellular concentration. You have to recalculate the intracellular concentration based on the volume of the cell pellet, the OD600. The estimated numbers routines used are as follows: OD600 of 1= 1,5 x 10^{7} cells, 1 yeast cells= 42 x $10^{(-15)}$ L

3.1.15 Randomized mutagenesis Library of AcKRS3 synthetase: construction and characterization

Error-prone PCR protocol. Error prone PCR (epPCR) was performed following the error prone PCR method (Wilson & Keefe, 2001) [257].

*Mb*AcKRS3 was used as a template for epPCR reaction. The forward and reverse primers (Forw BamHI AcKRS3 mut, 5'-tacttaGGATCCAAAAAAATGGATAAAAAACCGCTGGATG-3' and Rv PstI AcKRS3 mut, 5'-gcatgcCTGCAGTTACAGGTTCGTGCTAATGCCG-3')for the mutagenic PCR contained the restriction sites to allow digestion of the PCR product for cloning, and, respectively, the start and stop codons of the synthetase to guarantee the correct translation of the protein. By keeping the start and stop codons in the primers, these codons are protected from the mutagenesis, whilst all the nucleotides in between the primers will be subjected to amplification with reduced fidelity.

The reaction mix contained Tris pH 8.3 10 mM, KCl 50 mM, $MgCl_2$ 7 mM, dCTP 1 mM, dTTP 1 mM, dATP 0.2 mM, dGTP 0.2 mM, Forw primer 2 μM, Rv primer 2 μM, template DNA 20 pg/μl, $MnCl_2$ 0.5 mM (added just before reaction starts), Taq G2 DNA polymerase (Promega M784A) 0.05 U/μl (added just before reaction starts).

The epPCR reaction was carried out in serial reactions of 4 cycles in 100 μL in the recommended supplier reaction conditions and with an annealing temperature of 56 °C. This was sufficient to maintain the amplicon level constant throughout the reaction, as it can be seen in a staining of polyacrylamide gel electrophoresis of various stages of the amplification process. In the first reaction tube, the DNA template was a gel purified digested AcKRS3 fragment 20 pg/μl, while subsequent reactions were fed with 10 μL of the previous PCR product. So, reactions were run on the thermal cycler at 95 °C for 30 s followed by 30 cycles of 95 °C for 30 s, 56°C for 30 s, 72 °C for 180 s. Once cycles were complete, samples underwent a 10 min 72 °C final extension and were held at 4 °C until they were removed from the thermal cycler.

Following PCR with mutagenic dNTPs, the library derived from the reaction stopped after the 16th cycle was amplified again via PCR at a higher volume to prepare enough DNA for transformation into yeast. PCR was performed by combining 10X Q5 Pol Buffer, 0,5 mM dNTP, Q5 polymerase (0,05 U/μL), 10 μL epPCR-mutated DNA template, forward and reverse primer at 2 μM and sterile water to bring the total volume to 50 μL. Reactions were run on the thermal cycler at 95 °C for 30 s followed by 30 cycles of 95 °C for 45 s, 56 °C for 30 s, 72 °C for 60 s. Once cycles were completed, samples underwent a 10 min 72 °C final extension and hold at 4 °C until they were removed from the thermal cycler. After this second PCR the band of the correct molecular weight was gel purified.

Library construction: Ligation strategy and optimization.

The purified AcKRS3 library PCR product (100ul) was digested with *BamHI* and *PstI* restriction enzymes for 3h in the CutSmart buffer (NEB 3.1). One hour before the end of the reaction, an appropriate amount of calf intestinal phosphatase (CIP) (NEB M0290S) was added following the supplier's instruction. Adding CIP during insert digestion strongly reduced the formation of insert concatemers, guaranteeing a single synthetase mutant per plasmid. After gel purification to remove the CIP, the ligation between the fragment and the *BamHI*/*PstI* digested backbone of pRSII422 was performed in a 1:1 insert:vector ratio. Formation of backbone concatemers was expected and unavoidable, but did not hinder the selection efficiency. The ligated library was purified and then transformed in yeast (using large scale transformation protocol).

3.1.16 In vitro Assay

Yeast cell Lysis and Protein. After bait transformation or vector integration proteins were extracted with the following protocol, using Sigma glass beads:

1. Culture a single colony from the master plate in 10ml of liquid YC medium O/N.
2. The next day dilute the culture with YPD to OD600=0.3 and grow until 0.6. You will need at least 10-15 ml of final culture per extract.
3. Spin down the cells at 1,000 x g for 5 min
4. Remove the supernatant and resuspend in 10ml of sterile distilled water
5. Spin down again 1000 x g for 5 min
6. Remove the supernatant and resuspend in 200-300 µl T.E. 1X with protease inhibitors. Keep cells on ice from now on.
7. Add 1g of Sigma Glass Beads.
8. Vortex 1 min, put samples on ice. Repeat it four times.
9. Transfer all the supernatant you can take (usually are 200-300 µl) to a 1.5 ml Eppendorf tube and add 25 µl of Triton-X-100 20%
10. Vortex 30s
11. Spin down for 15 min (14,000 x g) at 4°C
12. Transfer the supernatant to a new eppendorf tube
13. Quantify with Bradford method
14. Boil 5 min and load the samples on an SDS-PAGE.

Western blot assay. Protein extract concentration was measured through the Bradford assay (Bradford, 1976) preparing a new standard curve with known concentrations of Bovine Serum Albumin (BSA). The desired amount of protein (50-100 µg) was then diluted in Laemmli buffer 4X and denatured at 95°C for 10 min. The samples were then rapidly centrifuged and loaded onto polyacrylamide gels for SDS-PAGE. Resolved proteins were transferred onto a nitrocellulose membrane, blocked with 5% milk for at least 1 hour at R.T. Then the membranes were incubated ON at 4°C or for 4 hours at R.T. with a primary antibody (the working concentration of every antibody used is reported above), followed by a further incubation of 1 hour with a secondary antibody HRP-conjugated. After each antibody incubation the membranes were washed with TBS-T, three times x 6 min. Protein detection was performed using a chemiluminescent substrate (Clarity or SuperSignal). When necessary, membranes were stripped with Re-Blot Plus Mild Solution 1X and probed with housekeeping specific antibodies (anti-PGK and anti-mouse) to monitor equal loading among samples. PrecisionPlus Protein Dual color was used as protein standard marker in the western blots. Chemidoc XRS was used for the chemiluminescent signal acquisition.

Antibody probes used in this thesis project were all diluted in 2,5% milk and are reported below:

-anti-HA Rat monoclonal High affinity (1:1000) – Roche, clone 3F10

- anti-H3Ac Pan Rabbit polyclonal (1:10000) – EMD Millipore 06-599

-anti-Histone H3 Rabbit (1:1000) Abcame 1791

-anti-LexA Rabbit polyclonal (1:1000) –EMD Millipore

-anti-Myc tag Mouse monoclonal (1:1000) –Qiagen

-anti-VP16 Rabbit polyclonal (1:1000)-Sigma V4388

-anti-PGK Mouse monoclonal (1:10000) – Abcam 22C5D8

-anti-Rat-HRP Goat (1:2000) – Santa Cruz Biotechnologies

- anti-Rabbit-HRP (1:2000) – Santa Cruz Biotechnologies sc2004

-anti-Mouse-HRP (1:5000) – Santa Cruz Biotechnologies sc2005

Immunoprecipitation (anti-HA)

Immunoprecipitation was conducted with 0.8 μg of an anti-HA monoclonal antibody (clone 3F10, Roche) in 200 μg of total yeast protein. The volume was adjusted to 200 μl with 50 mM Tris-HCl, pH 7.4, 250 mM NaCl,50 mM NaF, 5 mM EDTA, 0.1% NP-40 (vol/vol), 1 mM PMSF and complete protease inhibitors. Reactions were rotated at 4 °C overnight, followed by the addition of 30 μl protein G Sepharose 4 Fastflow 1:1 slurry (Sigma) and rotated at 4 °C for 2 h. The beads were pelleted by pulse-spin , washed six times with the binding buffer TBS +tween 0.05%The final immune complex was boiled in SDS-PAGE loading dye and resolved and blotted for western blot analysis.

Sample preparation for in-gel digestion Mass Spectrometry analysis. Mass Spectrometry (MS) analysis was used to confirm the incorporation of N-Acetyl-Lysine at the genetically encoded site in the protein bait of interest. The *In-gel digestion* method followed by MALDI-MS was chosen. tSL40 yeast expressing the LexA-H3K27*-HA bait with the TAG codon in correspondence with Lysine 27 was grown in both presence and absence of N-Acetyl-Lysine. After 24 h the protein extraction and immunoprecipitation of the bait was performed.

1. After immunoprecipitation, the pull-down fraction was loaded in a 15% SDS page with proper control.
2. After electrophoresis, protein bands are visualized using Coomassie.
3. Protein band corresponding to the expected MW of the full length H3 is then excised from the gel and de-stained.
4. Gel band samples were sent frozen on dry ice to the MS service (ToscanaLifesceince).
5. The in gel-digestion protocol with trypsin was performed by the MS service.

MALDI-TOF experiment. 2uL of sample solution (protein digests) was mixed with 2uL saturated solution of HCCA inTA30 solvent ((30:70[v/v] acetonitrile:0.1%TFAinwater)) and then deposited onto a stainless-steel plate and the solvents were evaporated (HCCA Dried Droplet, Ground Steel Targets). MALDI-TOF mass spectra were recorded using the MALDI TOF/TOF UltraflexIII (Bruker Daltonics, GmBH Germany) equipped with a nitrogen UV laser in linear positive-ion Mode (acquisition range 500-3500 m/z).

3.1.17 Functional Assay

Retrotrascription and Real Time PCR. Oligonucleotides for real-time PCR (RT-qPCR) **[Table 3.3]** were designed using BLAST Primer software, which included a BLAST analysis against *S. cerevisiae* Genome sequence for specificity confidence, and analysis using the server to avoid positioning on risky secondary structures. One microgram of total RNA was reverse-transcribed into cDNA in a 20 µL reaction mixture using Reverse Transcriptase Core Kit 300 Reverse Transcriptase Core Kit 300 (Eurogentec RT-RTCK-03). The cDNA levels were then analyzed using Rotar-gene Q (Quiagen). Each sample was tested in triplicate in a 96-well plate. A blank (No Template Control) was also incorporated in each assay. The thermocycling program consisted of 2 min at 50°C, 10 min at 95°C (Taq-pol activation) followed by 40 cycles with 15 sec at 95°C (denaturation) and 1 min at 60°C (annealing and extension). Quantitative values for cDNA amplification were calculated from the threshold cycle number (Ct) obtained during the exponential growth of the PCR products. Threshold was set automatically by the built-in Rotor-Gene Q relative quantitation analysis function. Relative expression was calculated with the 2ΔΔCt methods normalizing to the housekeeping TAF 10 gene for yeast RNA samples. Specifically, relative quantification of target transcripts has been calculated as follow: 2-[ΔCt(target)- ΔCt(control)] = 2-ΔΔCt, where ΔCt is the difference Ct(target)-Ct (TAF10), and Ct is the threshold cell cycle in the PCR.

Table 3.3. List of Primers used for RT-qPCR

TAF10	Forw,ATATTCCAGGATCAGGTCTTCCGTAGC; Rev,GTAGTCTTCTCATTCTGTTGATGTTGTTGTTG
H3K9*	Forw, AGATTACGCTGGGTCCCGTA; Rv, TGAGGTTTCTTGACTCCGCC
Pre-CYH2	Forw, GAGGTCACGTCTCAGGTATGT; Rv, CGTCTCAACAGTGAGATGGT
CYH2	Forw, TGTCCAAGTTCAAGACTGGCT; Rv, GTCACGTCTCAGCCGGTAAA

Mav203 phenotyping yeast assay.

The *S.cerevisiae* MaV203 strain is MATa; leu2-3,112; trp1-901; his3D200; ade2-101; cyh2R; can1R; gal4D; gal80D; GAL1::lacZ; HIS3UASGAL1::HIS3LYS2; SPAL10::URA3.

MaV203 contains deletions in the endogenous *GAL4* and *GAL80* genes for use with most GAL4-based two-hybrid systems. This strain has been constructed with three GAL4-inducible reporter genes for identification of interacting fusion proteins: *GAL1::lacZ*, *HIS3UASGAL1*::*HIS3* and the counter-selectable *SPAL10::URA3*. MaV203 is also auxotrophic for leucine (*leu2*) and tryptophan (*trp1-901*)- MaV203 also contains the recessive resistance alleles *cyh2*R and *can*R which are useful for plasmid shuffling.

In particular, the MaV203:pGADGAL4(2TAG) cell from Hancock et al. was employed [149]. This yeast strain expresses a full length GAL4 transcriptional activator gene bearing 2 amber codons in correspondence with T44 and R110, besides the HIS3 gene on GAL4-activated promoters. When a functional amber suppression system is transformed into this strain, full length GAL4 is produced, leading to activation of *HIS3* genes. Transcription of this gene allows cells to grow in the absence of histidine. To test the phenotypes conferred by various plasmid combination on Mav203::pGADGAL4(2TAG) yeast colonies from YC plates of each transformation were resuspended in 15 µl of sterile TE 1X and streaked or serial diluted (over 5 order of magnitude) on the selective media of interest. Each phenotype was confirmed with at least 3 independent colonies.

Super folded GFP read-through assay. The pMICBD1-ymGFP-39* can be used as the reporter plasmid to quickly verify the incorporation of unnatural amino acids by the orthogonal tRNA/synthetases on agar plates, and to quantify the incorporation efficiency of the ncAAs into GFP by measuring the fluorescence intensity.GFP expression for Fluorescence quantification.

1. Yeast cells containing the gene encoding yeast monomeric GFP with an in-frame amber stop codon at position 39 were transformed with plasmid pRSII422 bearing one of the aaRS/tRNA pair by LiOAc method. The transformation was performed onto selective plates YC/−Ade/−Trp.

2. After 3/4 days Pick 3 single colonies (3 biological replicates) from the YC/−Ade/−Trp agar plate, suspend in 1 ml YC/−Ade/−Trp medium, and vortex to disperse.

3. Add 500 µl of the volume to a different well of a 24-well deep-well plate (Riplate SW 24, PP, 10 ml) containing 4.5 ml of appropriate YC medium for selection, with or without 5 mM-10mM ncAA, as necessary.

Perform steps 1–3 for both yeast cells transformed with pMICBD1-ymGFP39* only in YC/−Trp medium as negative control and yeast cells transformed with pMICBD1-ymGFP full length only in YC/−Trp medium as positive control.

4. Cells were incubated at 30 °C at 250 r.p.m. for 22 h, and 500 µl of medium was transferred to clear-bottom 24-well plates. The GFP fluorescence in this standard volume was measured with the PHERAstar FS (BMG Labtech) plate reader using the optic module with an excitation wavelength of 485 nm and an emission wavelength of 520 nm. The gain was set to 8 (arbitrary units). Measurements were performed on each sample grown in individual wells.

* Use the negative and positive control samples to adjust the detection sensitivity of the flow cytometer so that the fluorescence intensity readings of all samples are in a good dynamic range.

5. Measure the mean fluorescence intensity of cells transformed with pRSS422-aaRS/tRNA and pMICBD1-ymGFP39* grown in presence of ncAA (Int1), and those grown in the absence (Int2). The net incorporation efficiency is defined by (Int_1−Int_2). Note that this GFP reporter can be generally used to evaluate the incorporation of many unnatural amino acids, as the 39 TAG site is permissive for GFP fluorescence. Also note that the incorporation efficiency of an unnatural amino acid can be protein-dependent and site-dependent

3.1.18 tRNA Extension experiment

RNA extraction for tRNA Extension analysis

To perform the tRNA extension (tREX) analysis, total tRNAs extraction described was adapted from Cervettini et. al. 2020 [214]. A preculture of the desired strain was incubated overnight at 30 °C with shaking at 220 r.p.m., and was diluted into 50 ml of prewarmed proper medium and cells were incubated at 30 °C with shaking at 220 r.p.m. for 3-4 h. At OD600 = 0.6, cells were pelleted at 14,000 x g for 5 min.

The cells pellet was resuspended in 800 µl of buffer D (50 mM sodium acetate pH 5, 150 mM NaCl, 10 mM MgCl2 and 0.1 mM EDTA) and transferred to a 2-ml tube. Cells were pelleted at 5,000 x g for 2 min and the pellet was resuspended in 450 µl of buffer D. Fifty microliters of liquefied phenol (90% phenol in water; Sigma-Aldrich, cat. no. P9346) was added to the cells, and lysis was performed by head-over-tail rotation (15 r.p.m., 15 min, room temperature). The lysed cells were pelleted for 25 min at 20,000 x g at 4 °C. Approximately 500 µl of supernatant containing the RNAs was recovered and transferred to a clean tube, and 500 µl of chloroform was added to the solution. The tube was thoroughly vortexed for 1 min until a cloudy emulsion

was formed. The emulsion was separated by centrifugation for 1 min at 20,000 x g. The top layer containing the tRNAs was recovered (~480 µl). This solution was either frozen at −20 °C or immediately processed with periodate as follows.

tREX probe design. The tREX probes were DNA probes composed of two sections. The 3′ section of the probes was the reverse complement of the sequence of the tRNA of interest between canonical position 45 (the second nucleotide of the variable loop) and canonical position 76 (the 3′-CCA-5′ end of the tRNA). The 5′ section of the probe was a poly(A) sequence whose length scaled with the length of the 3′ section as follows:

Length of the 3′ section	30 to 31	32 to 33	34 to 36	37 to 38	39 to 41	42 to 43	44 to 46	47 to 48	49 to 51
Number of As	12	13	14	15	16	17	18	19	20

Probes were purchased from Sigma-Aldrich as PAGE-purified oligonucleotides and were labeled at the 3′-OH with Cy5 dye.

tREX protocol. The tRNA extract was aliquoted into three 136 µl samples (A, B and C). Sample A (positive control for full extension) was brought to 160 µl with buffer D, and tRNAs were precipitated with 375 µl of absolute ethanol. After incubation for 1 h at 10 °C, the sample was centrifuged for 25 min at 16,100 x g, the supernatant was removed and the pellet was dried. The pellet was then resuspended in buffer D to a final concentration of 1 µg µl–1, as measured by NanoDrop.

Sample B (negative control for no extension) was deacylated by adding NaOH (8 µl of 300 mM NaOH, 42 °C, 1 h) and then neutralized by addition of sodium acetate (8 µl of 3 M sodium acetate, pH 5) and oxidized with NaIO4 (8 µl of 100 mM NaIO4, 1 h, 10 °C). Next, 375 µl of absolute ethanol was added and tRNAs were precipitated and resuspended as for sample A. Sample C was brought to 152 µl with buffer D and then oxidized with NaIO4 (8 µl of 100 mM NaIO4, 1 h, 10 °C). Next, the tRNAs were ethanol precipitated with 375 µl of absolute ethanol and resuspended as for samples A and B. The control for probe specificity was tRNA extract from untransformed L40 yeast treated as for sample C.

The enzymatic reactions were assembled as follows:

2 µl of tRNA (1 µg µl–1; from sample A, B or C, plus a control for probe specificity),
1 µl of dNTPs (10 mM each)
5 µl of NEBuffer 2.1,
1 µl of Cy5-labeled DNA probe (1 µM) and 40.5 µl of ddH2O.

The reactions were annealed in a thermocycler using the following settings: 95 °C for 1 min, 70 °C for 2 min and 50 °C for 2 min followed by 4 °C. After annealing, 0.5 µl of Klenow fragment (3′→5′ exo–; NEB, M0212S) was added to each reaction. The samples were incubated at 37 °C

for 20 min. Next, each reaction was mixed with 50 μl of loading dye (8 M urea and 0.04% Orange G), and 10 μl was run on an 8% acrylamide (19:1) gel (1× TBE, 200 V, approximately 45 min). Gels were imaged on an RGB Typhoon Imager (635 nm).

3.2 The PISA 2.0 Project in yeast: Results

In this Chapter, all the collected results obtained in order to adapt the Genetic Code Expansion method to the Y2H platform will be described **[Figure 3.7].**

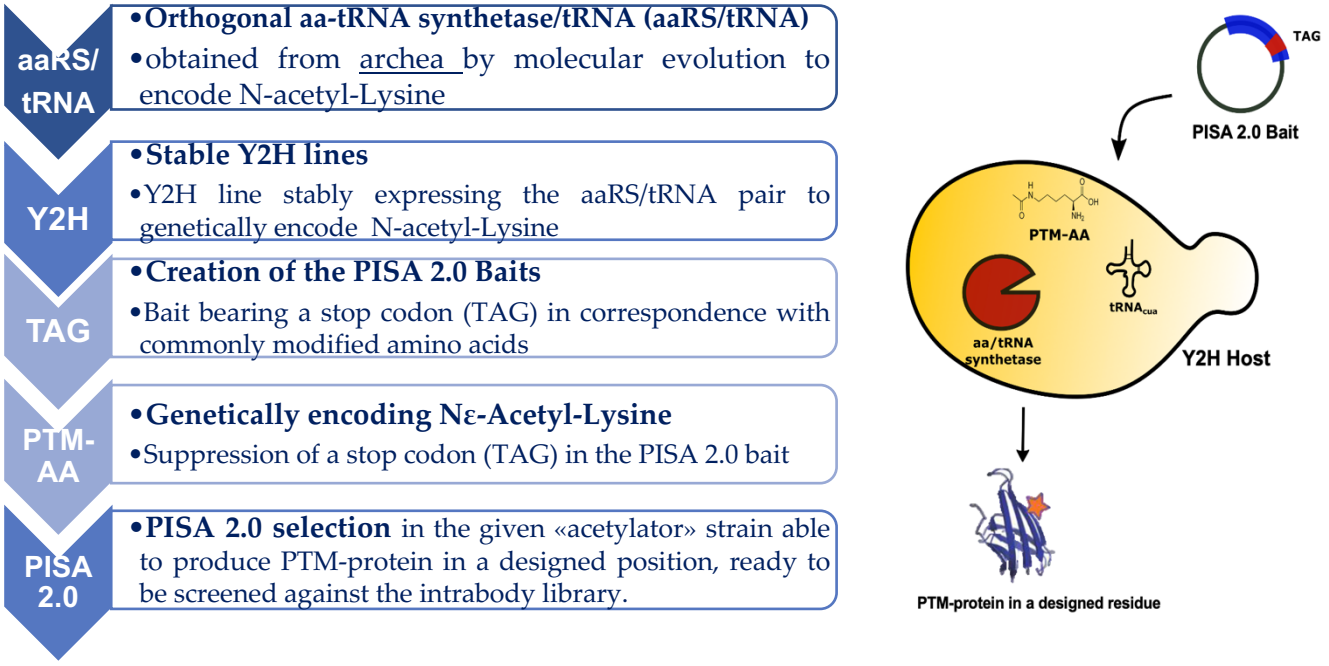


Figure 3.6. Schematic pipeline of the PISA 2.0 Project in yeast.

3.2.1 Generation of "acetylator" yeast stable lines

Efficient incorporation of ncAAs is limited to transient expression settings that lead to very heterogeneous expression [170]. To address the limitations of transient transfection approaches, I decided to stably integrate the amber suppressor pairs of our interest in the yeast L40 strain: i) the *M.barkeri* pair to encode N-acetyl-Lysine (*Mb*AcKRS3/tRNA$_{CUA}^{Pyl}$)and the *Methanococcus maripaludis* phosphoseryl-tRNA synthetase (*Mmp*SepRS/*Mm*tRNA$_{CUA}^{Sep}$) used to incorporate O-Phospho-L-Serine.

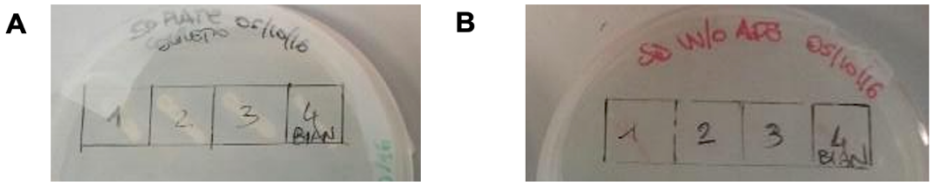
In the following sections all the required steps for the realization of the yeast stable line stably expressing the M*b*AcKRS3/tRNA$_{CUA}^{Pyl}$ pair will be described. This pair has already been

demonstrated to be functional in yeast [149], therefore I decided to set up the main PISA 2.0 selection first in the given "acetylator" yeast. The same steps will be followed for the *Mmp*SepRS/tRNA$_{CUA}^{Sep}$ pair, which will be first tested episomally. To the best of my knowledge, in fact, this pair has not been proven to be functional in yeast so far.

Test of *S.cerevisiae* L40 strain

Every experiment employed the L40 strain of *S.cerevisae*. This yeast strain is auxotroph for Tryptophan, Leucine, Histidine and Adenine and will not grow in absence of these compounds. In the PISA selection bait plasmids contain the *TRP1* gene and prey plasmids carry the *LEU2* gene. The *HIS* gene instead is the selection marker used during the screening (the *HIS3* transcription depends on LexA promoter activation). By process of elimination, Adenine auxotrophy was exploited as a selection marker to create the stable yeast cell lines with the integrated amber suppressor pairs.

Before starting, the L40 auxotrophy for Adenine was confirmed by observing/ not observing growth on selective minimal media plates YC w/o Adenine (Ade). As protocols, few single L40 colonies were restreacked onto YC w/o Ade and as control onto complete YC medium at 30 °C for four days **[Figure 3.8].**



	YC plate	YC-Ade plate
S.cerevisiae L40	Growth	No Growth

Figure 3.7. Test of L40 Auxotrophy. Four L40 clones restreacked onto YC w/o Ade and as control onto complete YC medium. On the left growth onto YC complete medium. On the right absence of growth onto YC w/o Adenine.

As expected, only a few senescent or dying pinkish colonies grew onto selective plate w/o Ade, mainly thanks to the starvation metabolic pathways. To confirm this hypothesis, cells picked from these colonies were in part restreacked again on the selective medium plate without Ade. Plates were monitored for up to 10 days and no survival clones could be detected.

The acetylator yeast stable line tSL40

In order to create the acetylator yeast line, the strain L40 was trasformed with the integration ADE2-carrying plasmid (pRSII402) bearing the *Methanosarcina barkeri (Mb)* AcKRS3/tRNA$_{CUA}^{Pyl}$ gene cassette. The only way pRSII402-*Mb*AcKRS3/tRNA$_{CUA}^{Pyl}$ can be propagated in yeast is by integration into the host genome at *ADE2* gene locus via homologous recombination, introducing in the process both the required protein to revert auxotrophy of L40 and the *Mb*AcKRS3/tRNA$_{CUA}^{Pyl}$ gene cassette. The *ADE2* locus of L40 carries on A12T transversion mutation in the 5′ of the gene which makes the protein inactive in L40 strain. As shown in **Figure 3.8 A**, after yeast transformation, the recombination process will produce two copies of the gene, recombining the 5′-3′ portions of the genes with each other; this will lead to one functional copy of the *ADE2* gene on the yeast genome (besides of the not-functional one) so that selection on transformed cells could be performed with adenine deprivation.

The Lithium Acetate method was used for yeast cells transformation. The protocol, specialized media and reagent are reported in the chapter "Material and Methods". After 3-4 days single colonies were visible and the transformation efficiency test[13] was performed. A good transformation efficiency (10^2/ug) was obtained, just one order of magnitude lower than what is usually obtained with episomal vectors **[Figure 3.8 B].** Thanks to the advances in yeast transformation protocols and the fact that the recombination in yeast occurs extremely efficiently, I was able to successfully generate the acetylator yeast line (also referred as tSL40). To verify that the recombination had taken place the integrating construction was sequenced using the "genome walking" approach to check the duplication of *ADE2* gene in the L40 yeast genome, called to simplify ADE#1 and ADE#2. BLAST alignment showed a copy of ADE2 gene still carrying A12T transversion and a functional wild type one (cfr. BLAST Alignment in Supplementary Material). Furtherly, the tSL40 growth was monitored over time: the strain was viable and we did no observe any evident growth phenotype alteration **[Figure 3.8 C]** compared to WT. Colonies appeared to be normal in color and size.

[13] Transformation efficiency is assessed by counting the number of cells grown onto YC w/o Ade plates, taking into account the appropriate dilution factor and quantity of plasmid DNA used.

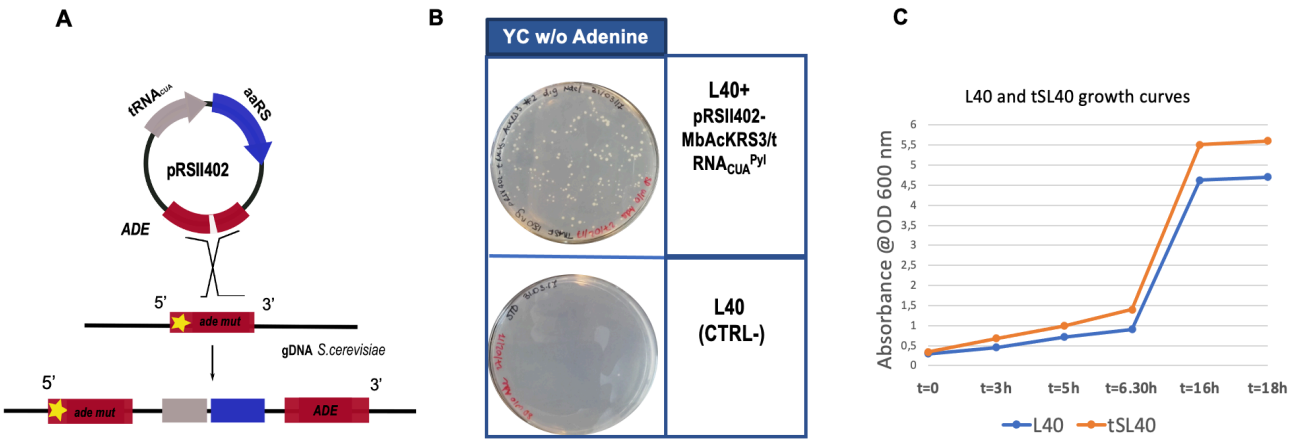


Figure 3.8. A) Schematic of the integration of a functional copy of the ADE2 gene in the yeast L40 genome. The recombination event is directed by cleaving the transforming DNA in between the Adenine marker, leading to a direct duplication in the chromosome in which only one of the two copies of the gene is active. **B) tSL40 yeast growth on selective medium w/o Adenine.** On top, L40 cells were able to grow on selective mediums without Adenine indicating the acquisition of the pRSII402-*Mb*AcKRS3/tRNA$_{CUA}^{Pyl}$ plasmid. Below, Absence of revertants was also assessed: no growth was detectable when the cells were transformed with the carrier DNA only. C) L40 WT strain and tSL40 strain growth curve comparison in YPD medium at 30°C. tSL40 strain followed the typical growth kinetic with no significant difference with respect to the WT strain. At 30℃, L40 wild-type yeast strains have a doubling time of ~180 minutes in YPD.

At this point I had to establish whether the amber suppression aaRS/tRNA pair was fully functional for the future PISA 2.0 selection. Accordingly, the expression of each of the components was verified together with the correct N-Acetyl-Lysine incorporation.

o-Prokarytotic tRNA cassette expression

General methods to express bacterial tRNAs in yeast involve placing an external Pol III promoter containing the consensus eukaryotic A- and B-box sequences upstream of the target bacterial tRNA gene (the CCA trinucleotide excluded) followed by a 3′-flanking sequence of an endogenous yeast tRNA [199,258]. Specifically, two yeast Pol III promoters, the RPR1 promoter and the SNR52 promoter, have been shown to efficiently drive the expression of *Escherichia coli* tRNAs in yeast. Alternative method using the yeast tRNA^{Arg} fused upstream of the target tRNA [259] has also been developed. Yeast in fact possess an unusual dicistronic tDNA$_{UCU}^{Arg}$ - tDNA$_{GUC}^{Asp}$ gene in which the two mature tRNAs are generated from a single precursor RNA. The tDNA$_{UCU}^{Arg}$ in fact provides the A and B-box sequences required to transcribe tDNAs inserted in place of tDNA$_{GUC}^{Asp}$.

Here I chose to take advantage of the latter method, as it works with different *E. coli* tRNAs and has been reproducibly used in different laboratories [124,260].

Specifically, I employed a *Saccharomyces cerevisiae Sc*tDNA$_{UCU}^{Arg}$–*Mb*tDNA$_{CUA}^{Pyl}$ cassette containing the natural 5′-, 3′-, and 10-base pair linker sequences. The cassette was subcloned in the MCS of the pRSII402 (see Material and Methods) **[Figure 3.9 A].** After integration in the genome, the detection of tRNAs was performed to establish whether they were transcribed in L40 strain or they needed to be modified. Although Northern blot analysis is the one of the most common methods used to verify tRNAs transcription, due to technical issues I chose to perform a reverse transcription-PCR analysis, that indeed turned out to be sufficient to detect solely the presence of tRNAs **[Figure 3.9 B].** A couple of primers for *Methanosarcina barkeri* tDNA^{Pyl} detection and one for *S. cerevisiae* tDNA^{Arg} (as positive control that tRNAs extraction has been successful) were designed. An additional Rv primer was, also, placed on the genomic upstream region from tDNA^{Pyl}, so that the same PCR protocol can be used on genomic DNA as negative control. Since tRNAs molecules fold in a secondary structure which is stable even above the annealing temperatures, primer pairs were designed so that they were complementary to an arm and half a loop of the tRNAs, as it shown in the scheme, avoiding intra-intermolecular interactions. Furthermore, to improve specificity, the 3' end of each primer corresponds to tRNA anticodon. The primer specificity was first tested in silico against the appropriate database http://gtrnadb.ucsc.edu. **Figure 3.9** shows the major result achieved via PCR: the specific detection of the *M.barkeri* tRNA^{Pyl} was obtained from cDNA of tSL40 and not from the WT L40 strain .

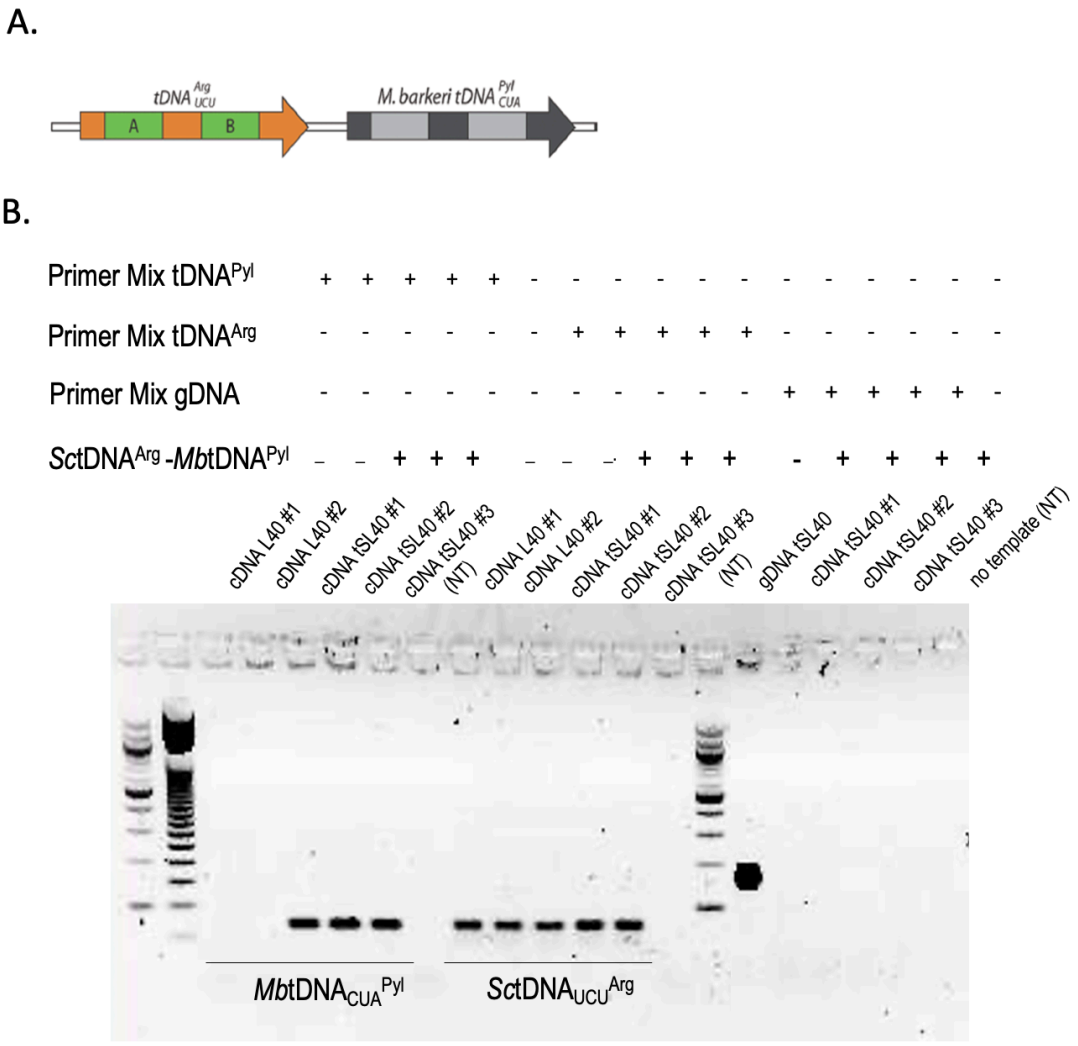


Figure 3.9. o-Prokarytotic tRNA cassette expression in yeast (A) Schematic of the MbtDNA$_{CUA}^{Pyl}$ expression constructs created and examined in this work. **(B)** Retro Transcription-PCR analysis was chosen to detect the tDNA^{Pyl} in cDNA samples from two tSL40 colonies expressing the SctDNA^{Arg}-MbtDNA^{Pyl} cassette. L40 cDNA was used as control. A couple of primer (Primer Mix tDNA^{Pyl}) for tDNA^{Pyl} detection and one (Primer Mix tDNA^{Arg}) for S. cerevisiae tDNA^{Arg} as positive control were designed. An additional Rv primer was also placed on the genomic upstream region from tDNA^{Pyl}, so that the same PCR protocol can be used on genomic DNA from the tSL40 colonies as a control. Gel electrophoresis result from PCR with the primer sets: Primer Mix tDNA^{Pyl}, Primer Mix tDNA^{Arg}, Primer Mix gDNA from either cDNA of L40 strain or cDNA from L40 stably expressing SctDNA^{Arg}-MbtDNA^{Pyl}. The location and label of each predicted PCR products are indicated. Legend NT= no template

o-Synthetase expression

The yeast codon optimized Myc-tagged AcKRS3 expression was then assessed. SDS-page and immunoblot analysis on a stably integrated tSL40 clone's lysate using the Myc-tag as detector was performed. **Figure 3.10** shows Western Blot analysis performed with anti-Myc antibody after glass-beads-protein extraction from tSL40 yeast (see "Material and Methods).

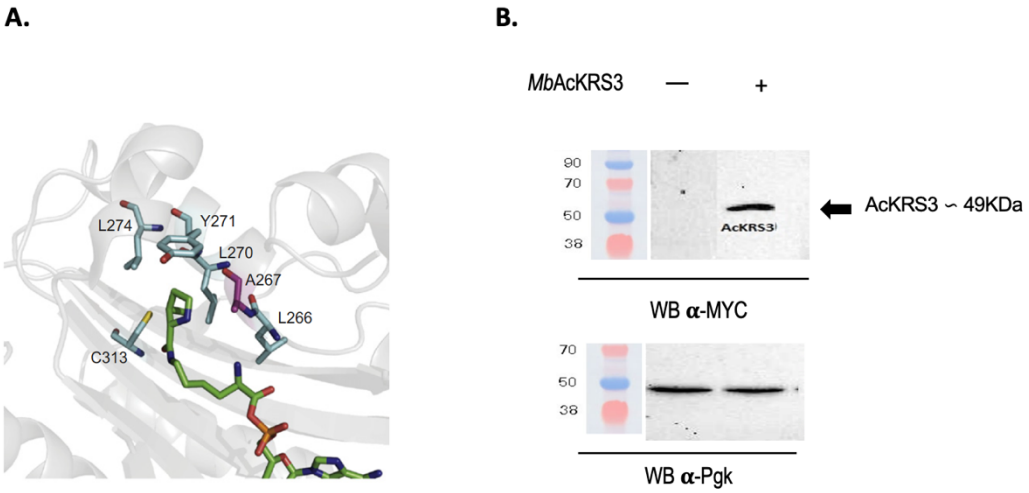


Figure 3.10. Orthogonal procaryotic synthetase expression in yeast.A) The active site of PylRS bound to pyrrolysine (figure created using PyMOL [http://www.pymol.org] and PDB file 2Q7H). The residues mutated in Neumann et al. 2009 relative to the wild-type sequence are shown as sticks. **B)** WB analysis using anti-Myc antibody to detect the *Mb*AcKRS3 on protein extracts of tSL40 strain (lane 2). L40 WT strain was used as negative control (lane 1). The same extracts were reprobed with anti-PGK antibody to confirm equal loading of each lane (panel below).

N–ε-Acetyl-Lysine Intracellular concentration measurement

Another prerequisite for the ncAAs incorporation methodology is that the N-Acetyl-Lysine (AcK) would be efficiently transported into the cytoplasm when added to the growth medium. Accordingly, the intracellular concentration $[c]_{int}$ of the AcK in tSL40 strain was assessed by Mass Spectrometry (MS) (The protocol, specialized media and reagent are reported in the chapter "Material and Methods"). The $[AcK]_{int}$ measured in yeast samples grown in minimal media supplemented by 10mM AcK was approximately 1mM, the expected working range concentration [261] [cfr.Table 3.4]

3.2.2 Encoding site-specific bait protein acetylation

PISA 2.0 baits construction: Protein target bearing the TAG codon to encode PTM-AAs

After the generation of the tSL40 line, I had to design PISA 2.0 baits carrying the amber stop codon in correspondence with those amino acids commonly modified by modifying enzymes. As first proof of concept, the Histone H3 recombinant proteins bearing the TAG stop codon in correspondence with either Lysine 9 or Lysine 27 were chosen as PISA 2.0 baits. The H3 protein target emerges more than the others because of the importance of histone acetylation in fields of translational interest, such as cancer epigenetic. Lysine 9 and 27 are among the most biologically relevant sites. Furthermore, the H3K9ac turned out to be an excellent testbed to address with PISA 2.0 considering the previous work done by the lab [32]. Since the PTM-AA incorporation efficiency is strictly dependent on the sequence context of the amber codon and its position in a gene, the H3K27ac was also tested.

The H3 bait constructs were designed using site directed mutagenesis as described in Material and Methods and subcloned in pMICD1 bait plasmid, so that the final H3 bait were fusion proteins of LexA DBD (necessary for PISA 2.0 screening) at the N-term and a HA tag at the very C-term that serves for Immunoprecipitations (Ip) and Western Blot detections. This conformation was chosen to test the functionality of the orthogonal translation system in vitro: the detection of the read-through protein via Ip assay is strictly dependent on the incorporation of the AcK in response to the amber stop codon (The installing N-Acetyl-Lysine experiment). Otherwise, cells translate the bait truncated form, which could not be immunoprecipitated **[Figure 3.11].**

Here follows the PISA 2.0 bait protein sequence and description **[Figure 3.13 A-C]:**

1)LexA-H3-HA: full-length HA tagged H3 was used as positive control in all the experiments.
LEFPGIRRPAAKRTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALR EIRRYQKSTE YPYDVPDYA

2)LexA-H3K9*-HA: H3 bearing the TAG codon in correspondence with Lysine 9.
LEFPGIRRPAAKRTKQTAR*STGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALRE IRRYQKSTE YPYDVPDYA

3)LexA-H3K27*-HA: H3 bearing the TAG codon in correspondence with Lysine 27.
LEFPGIRRPAAKRTKQTARKSTGGKAPRKQLATKAAR*SAPATGGVKKPHRYRPGTVALRE IRRYQKSTE YPYDVPDYA

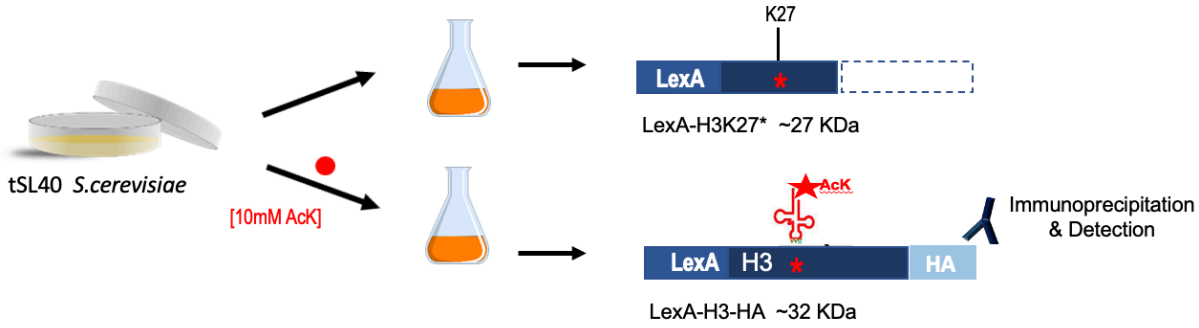


Figure 3.11. Schematic of the Read-Through bait protein Immunoprecipitation Assay. One colony of yeast tSL40 stably expressing the AcKRS/tRNA$_{CUA}^{Pyl}$ pair and bearing the bait pMICBD1-H3K27*-HA was grown in liquid medium supplemented with or without 10mM of Ac-Lys. Then, protein extraction and Immunoprecipitation using anti-HA agarose beads were performed to assess the constitutive acetylation. The detection of the read-through H3 protein via Ip assay is strictly dependent on the incorporation of the AcK in response to the amber stop codon. Otherwise, cells translate the bait truncated form, which could not be immunoprecipitated. AcK= N-Acetyl-Lysine

Each bait (controls included) was transformed in tSL40 with LiOAc method and a series of control experiments were performed. First the transformation efficiency and toxicity were measured: after 3-4 days single colonies were visible onto selective media and a good transformation efficiency for LexA-H3-HA, LexA-H3K27*-HA, LexA-H3K9*-HA was obtained. The growth rate, number, and morphology of colonies of newly established lines were comparable with those of wild type L40 yeast strains, so results show that no particular toxicity can be observed for the tSL40 line.

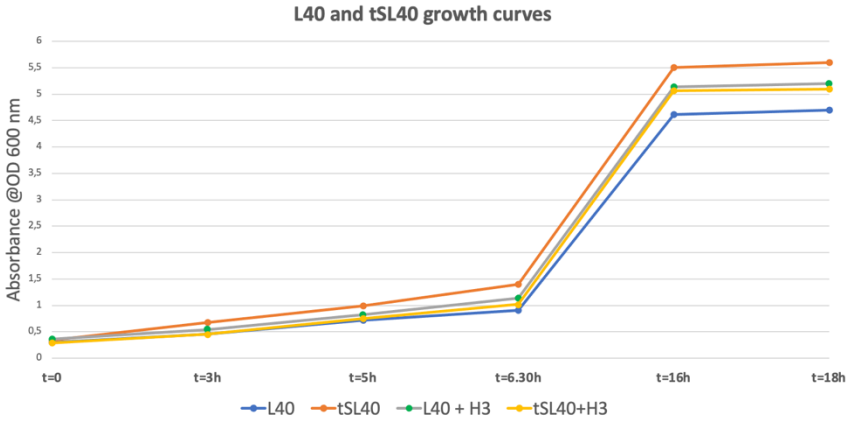


Figure 3.12 L40 WT strain and tSL40 strain growth curve comparison. Both tSL40 strain and tSL40 transformed with the bait (H3) followed the typical growth kinetic with no significant difference respect to the WT strain.

The expression level was also evaluated: baits must show good expression level to perform a screening. Three clones were analyzed for each construct by Western Blot analysis using anti-LexA antibody **[Figure 3.13 D].**

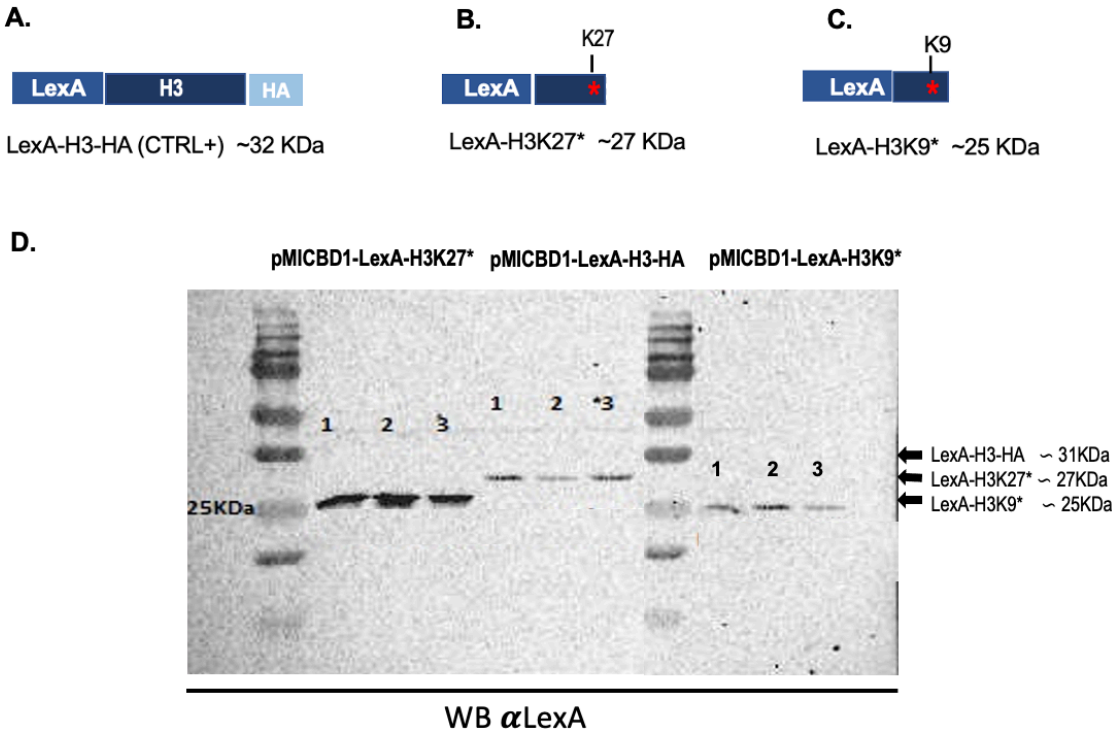


Figure 3.13. PISA 2.0 baits bearing the TAG codon. A-C) Schematic representation of PISA 2.0 baits. In this example (B-C), the target antigen Histone H3 bearing the TAG stop codon-in correspondence with those amino acids commonly modified by modifying enzymes- is fused to C-term of LexA DBD (necessary for PISA 2.0 screening) and at the N-term HA tag that serves for Immunoprecipitations (Ip) and Western Blot detections. The full-length HA tagged H3 was used as positive control (A). **D)** WB analysis using anti-LexA antibody on protein extracts of tSL40 cells expressing in. order pMICBD1-LexA-H3k27*-HA, pMICBD1- LexA-H3-HA and pMICBD1-H3K9*-HA baits (3 clones were tested).

Installing N-Acetyl-Lysine

Having established that the components required for genetically encoding N-Acetyl-Lysine were stably transfected and produced in tSL40, before moving on to the main selection, the functionality and the orthogonality of the *Mb*AcKRS3/tRNA$_{CUA}^{Pyl}$ pair needed to be validated. To determine the incorporation of Ac-Lysine in response to the TAG codon in the protein of interest the Immunoprecipitation assay and Western Blot analysis were used. As previously described, the expression and immunoblot detection of both the read-through H3K27*-HA and H3K9*-HA baits via the C-terminus tag were strictly dependent on the incorporation of N-

acetyl-lysine in response to the amber codon **[Figure 3.11 and 3.13 A-C].** To begin with, three colonies (three biological replicates) of tSL40 bearing pMICBD1-H3K27*-HA were grown in liquid medium supplemented with 10mM of Ac-Lys. As negative control, parallel growths without AcK were also performed. Then, both yeast extracts for direct immunoblot detection and Immunoprecipitated proteins were employed to assess the constitutive acetylation **[Figure 3.14].**

The immunoprecipitation (Ip) assay was conducted using anti-HA agarose beads and WB was performed with anti-LexA antibody. The Ip results-left panel **[Figure 3.14 D]** - clearly showed the read through H3 expression (from pMIC-H3K27*-HA in tSL40) only in the culture supplemented by Ac-Lys, at the expected molecular weight like unmodified H3-HA (used as positive control). I was not able to observe the read-through in the input direct immunoblot (right panel) since the detection of full-length acetylated H3 proteins occurred only if they were previously immunoprecipitated. Additional WB analysis using an anti-H3K9Ac antibody will be addressed in future work. MALDI-Mass Spectrometric analysis of the Histone H3 (with the unnatural amino acid incorporated) isolated by SDS-PAGE, confirmed the presence of Ac-Lysine at the genetically encoded site[14] **[Figure 3.15].** Importantly here, *the N-Acetyl-Lysine dependent amber suppression* of one of the PISA 2.0 baits was fully demonstrated in tSL40 yeast ("the acetylator" yeast).

[14] Specifically, Peptide mass fingerprinting (PMF) by MALDI-TOF and subsequent sequence database searching was used to reveal the presence of Ac-Lysine at Lysine 27. In this approach, peptides are generated by digesting H3K27ac readthrough protein isolated by SDS-PAGE with the sequence-specific enzyme trypsin. And then H3K27ac peptides are analyzed by MALDI-TOF mass spectrometry to get the peptide masses. The experimental masses are compared against a database containing theoretical peptide masses with the same sequence-specific protease. After that, the presence of PTMs can be detected by peptide analysis either as a mass increment or mass deficit (Δm) of the modified peptide relative to the unmodified species. Of course, the PMF analysis should be followed by amino acid sequence by tandem MS (MS/MS) in order to reveal the presence of PTMs at individual amino acid resolution. Unfortunately, for histone H3 proteins, which have many basic amino acids, trypsin generates too many short fragments that cannot be perfectly analyzed by tandem MS.

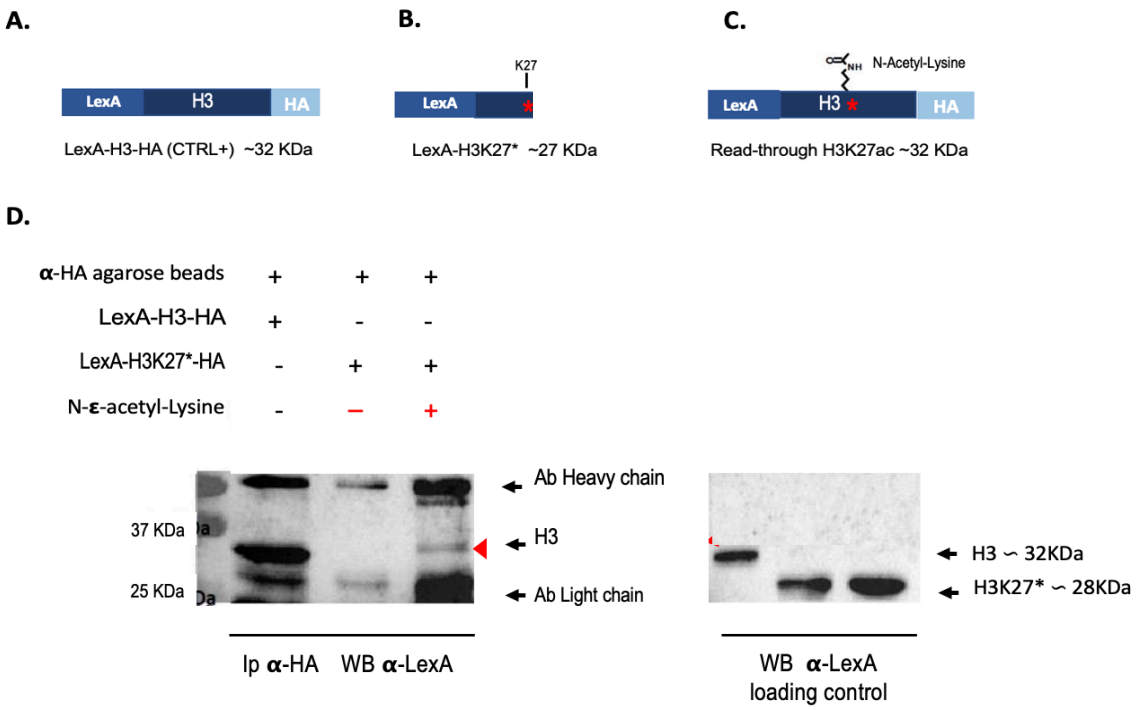
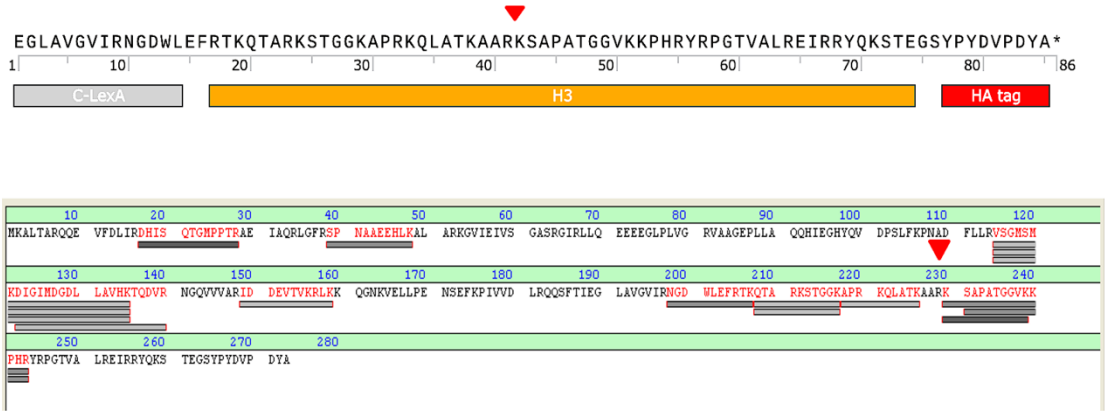


Figure 3.14. N-Acetyl-Lysine dependent amber suppression of Histone H3 with a TAG codon in correspondence with Lysine 27. A-C) Schematic representation of PISA 2.0 baits. The full-length HA tagged H3 was used as positive control (A). Histone H3 bait bearing the TAG stop codon-in correspondence with Lysine 27 (B) Acetyl Lysine dependent amber suppression of the Histone H3 bait bearing the TAG stop codon-in correspondence with Lysine 27 (C). **D)** Immunoprecipitation assay. Left Panel Immunoprecipitation with anti-HA agarose of soluble protein extracts from tSL40 yeasts intracellularly expressing either LexA-H3K27*-HA (∽ 27KDa) grown in both presence and absence of 10mM N-Acetyl-Lysine or LexA-H3-HA (∽32KDa positive control). Membrane blotted for LexA. H3 read through is pulled-down only in culture supplemented with N-Acetyl-Lysine (red arrow head). Right panel: The last three lanes show a fraction of the input used.

A.



B.

Peptide Sequence	Identified Acetylation	Range	Meas. m/z	Calc. MH+	Intensity	Dev. (Da)
VSGMSMKDIGIMDGDLLAVHK	2: Acetylation 4: 0xidation 7: acetylation	115-135	2349.130	2349.113	1099.030	0.016
KSAPATGGVKKPHR	1: Acetylation	230-243	1475.809	1475.844	13140.768	-0.0036
KSAPATGGVK	1: Acetylation 2: Acetylation 6: Acetylation 10: Acetylation	230-239	1083.612	1083.568	5038.531	0.044

Figure 3.15. MALDI-Mass Spectrometry: H3K27ac Peptide Mass Fingerprinting (PMS) and PTM mapping. A. Top panel. Partial amino acid sequence of Histone H3K27ac expressed in tSL40 strain in presence of acetyl Lysine. A portion of LexA sequence is also indicated at the N-terminus of the H3 sequence**. Bottom panel. Peptide map of H3 generated by tryptic digestion.** The sequence of H3 was analyzed by the bottom-up approach. Sequence coverage was shown. Gray bars indicate the tryptic peptides obtained for one of the H3K27ac samples. Red arrows indicate K27. (Three biological replicates. The complete array of data is reported in Supplementary Section). **B. Peptide and PTMs mapping.** Table shows the post-translationally modified peptides obtained from H3K27ac PMF. The Peptide sequence of interest containing the K27 is highlighted in bold, and it resulted to be acetylated at position K1 correspondent to K27 of H3, given a list of putative modifications and their Δm values and the list of predicted (column Meas m/z) and observed peptide masses (column Calc MH+). Legend. Meas.= measured m/z.Calc= calculated MH+

The same procedure was performed from tSL40 expressing pMICBD1-H3K9*-HA, nonetheless I was not able to recapitulate the previous results. No read-through protein was observed neither via Ip nor via direct immunoblotting. This data might not be of concern since the PTM-AA incorporation efficiency is reported to be dependent on the sequence context of the amber codon and its position in a gene [184].

What was obtained here was a promising result, but good detection of read through proteins occurred only if they were previously immunoprecipitated; therefore, I was not sure whether the expression level of constitutive acetylated protein was suitable for PISA 2.0 screening or

not. A robust expression of the bait was paramount for the selection. Therefore, I decided to perform an *in vivo* binding test with a known read-through H3 bait interactor to assess if the acetylated protein yield was sufficient to allow the yeast growth.

PISA 2.0 bait vector: reversed LexA polarity pMICBD1 vector

Regarding the PISA 2.0 selection, I opted for modifying the bait plasmid pMICBD1 for a technical reason. Specifically, I constructed a new pMICBD1-MCS-LexA bait plasmid, in which the LexA DNA Binding Domain (DBD) polarity (i.e., N-bait-LexA-C) was reversed compared to the canonical PISA pMICBD1 bait polarity (N-LexA-bait-C) **[Figure 3.16].** Most Y2H bait plasmid used N-terminally fused LexA DBD [262], on the contrary because here LexA is fused to the target C-terminus, the identification of antigen-binders would occur only if the PTM-AA is incorporated, greatly accelerating the PISA 2.0 selection. The suitability of pMICBD1-MCS-LexA vector for Y2H interaction was tested performing the IACT screening with a couple of bait-prey interactor previously tested in the lab (exploiting the canonical pMICD1) **[Cfr. Appendix I]**. Accordingly, the H3 baits were subcloned in the new PISA 2.0 vector and transformed via LiOAc methods in the acetylator strain to set the PISA 2.0 in vivo binding test.

Figure 3.16. Optimization of the PISA 2.0 Bait Vector. A novel pMICBD1-MCS-LexA bait plasmid, in which the LexA DNA Binding Domain (DBD) polarity (i.e., N-bait-LexA-C) (B) was reversed compared to the canonical PISA pMICBD1 bait polarity (N-LexA-bait-C) (A) was constructed for this thesis**.** In the PISA 2.0 configuration, LexA is fused to the target C-terminus so that the identification of antigen-binders would occur only if the PTM-AA is incorporated, greatly accelerating the PISA 2.0 selection.

3.2.3 The PISA 2.0 in vivo binding test: Testing the interaction between the H3K9* bait and the anti-H3K9ac scFv-58F in the tSL40 acetylator stain

Although the in vitro Ip assay showed a low acetylated-protein yield, I reasoned that it could have been sufficient to allow the in vivo selection and that the selective pressure might also have boosted the system. Specifically, to assess whether the expression levels of the read-through acetylated H3 proteins were suitable for the PISA 2.0 selection, I performed a Y2H interaction test using the PISA 2.0 bait H3K9*-LexA and the anti-H3K9ac scFv-58F (previously identified in [[32]) in tSL40 yeast 2 hybrid platform.

This screening test works by co-transforming the tSL40 (L40 stably expressing the pair for encoding AcK) with the LexA fusion protein H3K9* and the anti-H3K9ac scFv-58F. The positive bait-prey interaction activates the transcription of the HIS3, allowing survival of the yeast clone on selective media (supplemented with AcK for this experiment). Because here LexA was fused to the C- terminus of H3K9*, the production of the full-length acetylated protein and so the interaction with the scFv-58F would occur only if the Ac-Lys was incorporated. tSL40 expressing the H3K9* protein target was transformed using the LiOAc method with the anti-H3K9ac scFv-58F or control scFvs (the unrelated scFv-88 and the a-LexA scFv-15). Cells were plated onto minimal medium plates (-WL) and selective plates (-WHL +/- 10mM AcK).

Picture shows that tSL40 H3K9* was not able to grow onto selected media (supplemented with AcK) when co-transformed with the scFv-58F. A few tiny colonies were observed in presence of the a-LexA scFv-15: this is probably due to a little translation of LexA transcripts starting from one of its numerous ATG codons at the 5′ end (this result was corroborated by WB analysis **[Figure 3.17 A].** Colonies from -WL plates were therefore taken into account for further confirmations: WB analysis was performed to detect and verify the correct scFv-58F expression levels **[Figure 3.17 B].** Together, since the scFv-58F was a specific interactor of the H3K9ac protein [32,112], it can be stated that the expression level of the PISA 2.0 bait H3 bearing the TAG at position 9 seemed to be not suitable to perform an in vivo interaction assay. However, this result was not conclusive. I surmised, in fact, that this bait construct presented the nonsense codon too close to the 5′end: this would markedly reduce its mRNA level by the Nonsense Mediated RNA decay pathway in yeast impairing our test screening. As described in Chapter 1.8.2 differently to higher eukaryotes , NMD in yeast shows a polar effect of nonsense codon positions [206,207]. The steady-state mRNA level is reduced by NMD more significantly when the nonsense codon is closer to the 5′ end than to the 3′ end of an mRNA. Consistently, the unnatural amino acid incorporation efficiency in the yeast correlates with the position of the UAG codon[142,183].

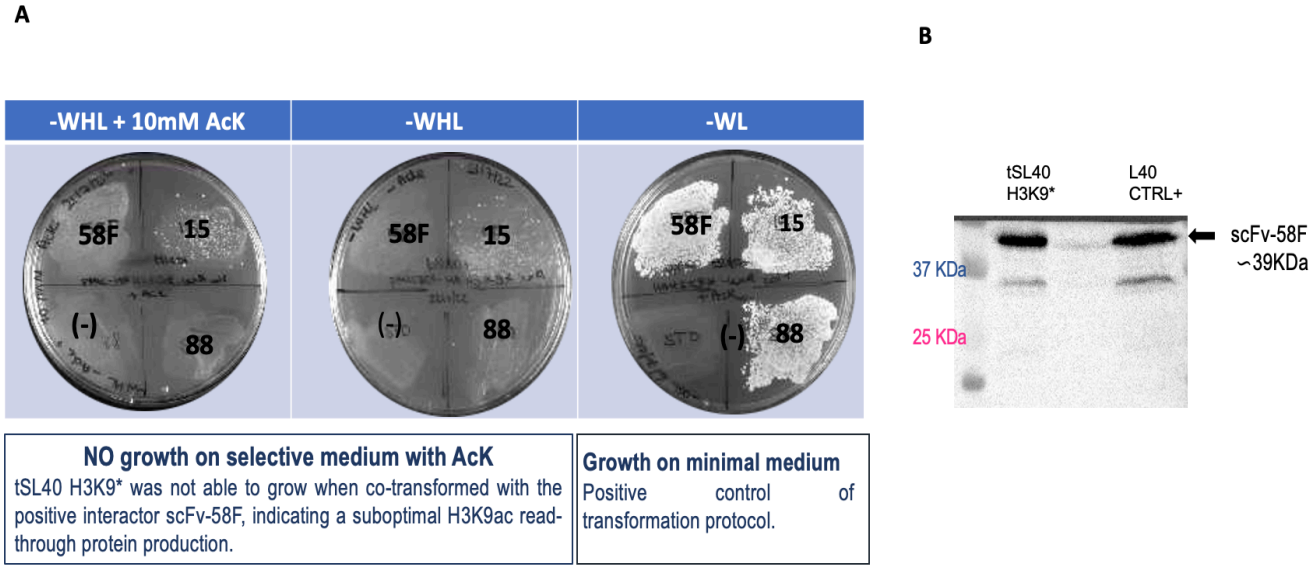


Figure 3.17. PISA 2.0 in vivo interaction test: the expression level of the readthrough H3K9* PISA 2.0 bait is not suitable to perform an *in vivo* interaction assay. (A) Yeast growth on histidine-lacking plates (-WHL) and control plate (-WL). In particular, it is shown how the tSL40 expressing H3K9* was not able to grow onto selected media (supplemented with AcK) when co-transformed with the scFv-58F indicating a suboptimal read-through protein yield. Legend. 58F = scFv-58F, 15= scFv-15 anti-LexA used as positive control, 88= scFv-88 unrelated scFv used as negative control, (-)= carrier ssDNA only. **(B).** WB analysis using anti-VP16 antibody on protein extract of one tSL40 colony expressing pMICBD1-HA-H3K9*-LexA and pL220-scFv-58F-VP16 picked from –WL screening plate. A protein extract from L40 expressing pL220-scFv-58F-VP16 was used as positive control.

3.2.4 Methods to improve ncAAs incorporation efficiency

Despite the robust preliminary results on the generation of the yeast strains carrying the *Mb*AcKRS/tRNA pair, data also showed very low yield of protein bait readthrough, indicating a suboptimal translational activity of the amber system. So, before moving on to the main selection, I reasoned on how to increase the acetylated-protein yields as strong expression level of protein targets was paramount for the PISA 2.0 screening. All the state-of-art methods to improve the amber suppression system were systematically followed as described in the following chapters (i. increasing the o-tRNA copy numbers; ii. Disabling the NMD pathway; iii. changing the synthetase variant) [Figure 3.18].

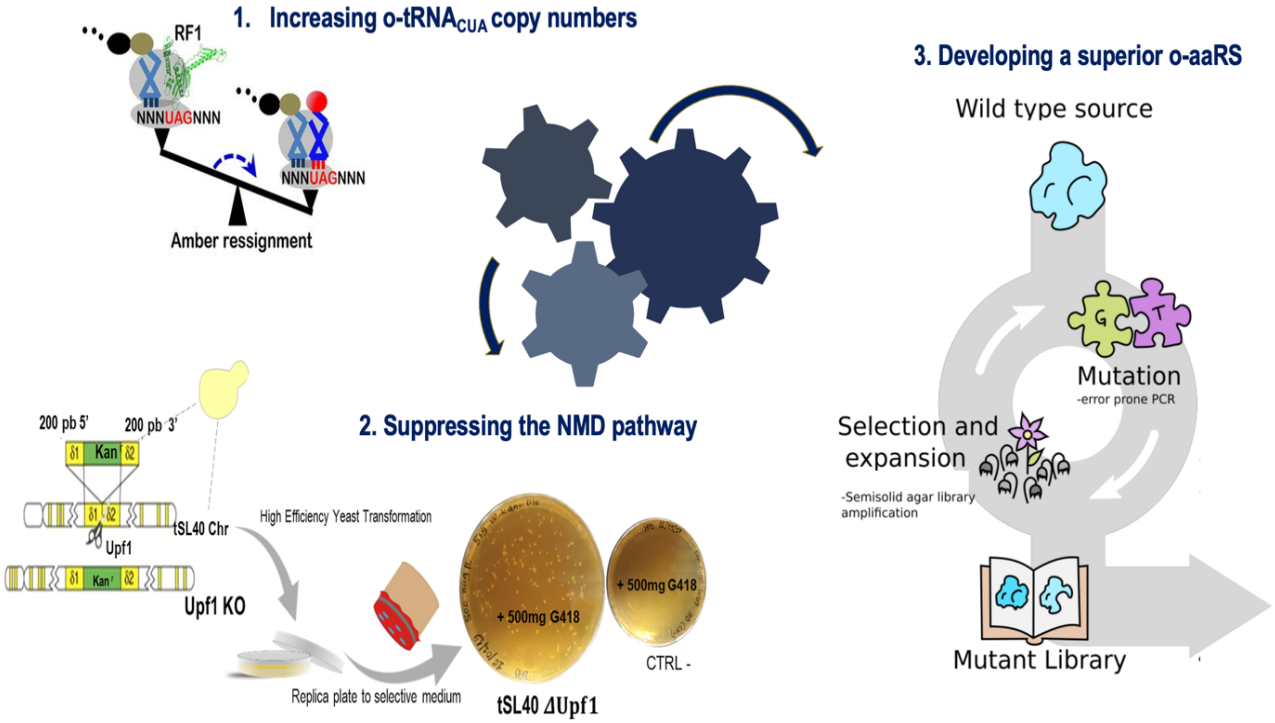


Figure 3.18. Schematic pipeline of the up-to date methodologies applied to the "acetylator strain" in order to increase the read-through protein yield. i) increase the tRNA$_{CUA}^{Pyl}$ gene copy number to minimize competition with the eRFs. ii) Impairing the nonsense-mediated mRNA decay (NMD) by Knocking Out an essential gene for the pathway. iii) Evolving a more active AcKRS synthetase using directed evolution technique.

Increasing tRNA copy numbers

One reason amber suppression is less efficient is that the amber-suppressor tRNA competes with release factors (RFs) in recognizing the TAG codon. Accordingly, I decided to increase the $tRNA_{CUA}^{Pyl}$ gene copy number to minimize competition with the eRFs. To begin with, I subcloned another copy of the tRNA cassette in the bait plasmid to test this strategy episomally. This vector, called pMICBD1/$tRNA_{CUA}^{Pyl}$ was used from now on.

Impairing the nonsense-mediated mRNA decay (NMD) Pathway in tSL40

Another important factor for unnatural amino acid incorporation in eukaryotes is the stability of the target mRNA. Eukaryotic cells have an mRNA surveillance mechanism, the nonsense-mediated mRNA decay (NMD), to identify mRNAs containing premature stop codons and target the mRNA for rapid degradation. Use of the amber stop codon for encoding PTM-UAA may subject the target protein mRNA to the NMD, thus decreasing the protein yield. In particular, the NMD pathway in yeast shows a polar effect of nonsense codon positions [263]. The steady-state mRNA level is reduced by NMD more significantly when the nonsense codon is closer to the 5′ end than to the 3′ end of an mRNA, similarly to our case.

Hence, I generated an NMD-deficient tSL40 yeast strain (tSL40ΔUPF1) by knocking out the *UPF1* gene, an essential component for NMD, from the yeast genome [198,199]. The NMD pathway deficient strain were made by one-step UPF1 gene replacement using a dsDNA fragment that contains ends of the target gene (homology regions required to target the integration), flanking a selectable marker. A gene cassette containing about 200 bp upstream of *UPF1* (PCR extracted from the genome of L40), the G418 resistance, and about200 bp downstream (PCR extracted from the genome of L40) was transformed into tSL40 yeast strain using the Lithium Acetate method. Transformed cells were plated on YPD medium for 6h and then replica plated on G418 YPD plates for selection. After transformation and selection, integration of the cassette and concurrent deletion of the chromosomal target sequence was analyzed by whole colony PCR first. Using genomic primers 300 bp away from the UPF1 gene together with specific primers either for KanR or UPF1 gene, integration of the deletion module (as the presence of the wild-type gene) was detected based on the size of the PCR products **[Figure 3.19].** Finally, the genomic DNA of one positive clone (colony 1) was amplified and sequenced with Sanger methods to confirm the UPF1 gene replacement. The resultant *UPF1* deficient strain was designated as tSL40ΔUPF1.

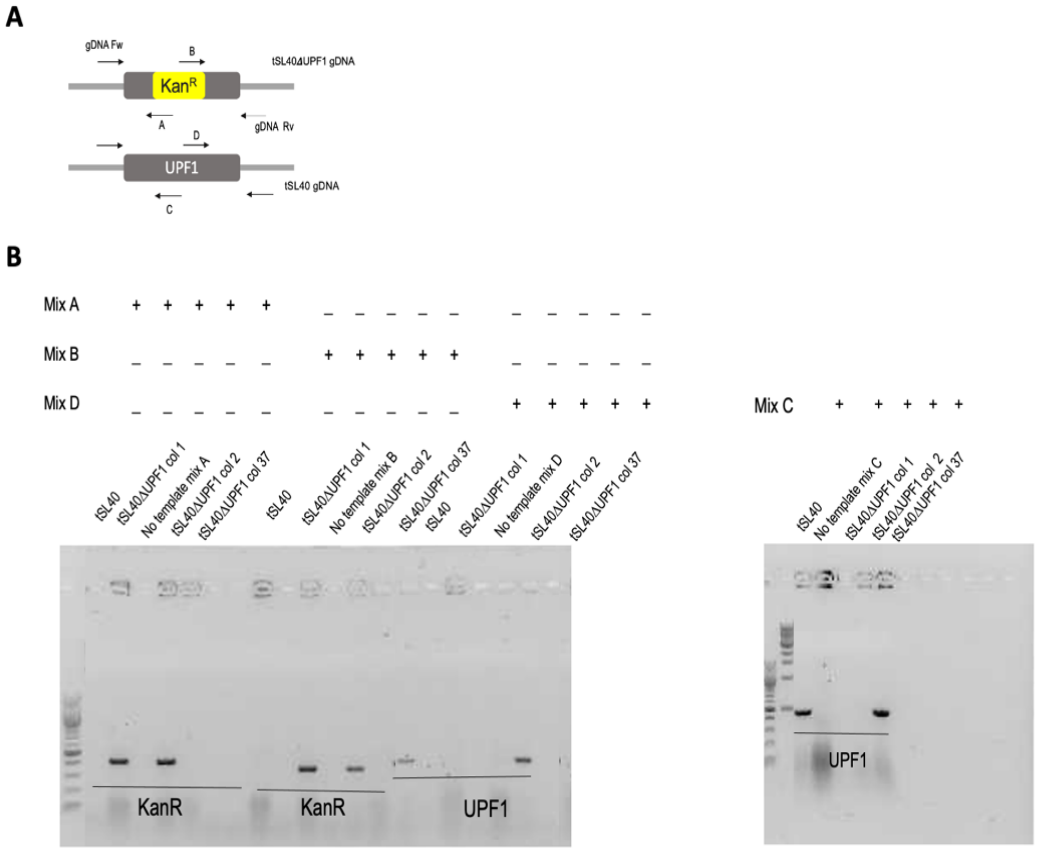


Figure 3.19.The tSL40ΔUPF1 strain: an NMD-deficient strain. (A) The design of primers for gene deletion verification is depicted. Short primers designed ~200 bp 5′ to the start codon on the coding strand (gDNA Forw) and ~200 bp 3′ to the stop codon on the reverse strand (gDNA Rv). Additional primers ~100 bp downstream of the start codon on the reverse complement of both KanR cassette (int A) and UPF1 gene (intC); ~ 100pb primer upstream to the stop codon of KanR cassette (int B) UPF1 gene (intD). **(B) Integration of the cassette and concurrent deletion of the chromosomal target sequence is analyzed by colony PCR using the gDNA Forw coupled with both intA primer (mix A) and intC primer (mix C).** For the sake of completeness, also the gDNA Rv primer together with either intD primer (mixD) or intB primer (mixB) was used. Integration of the deletion module can be detected (as can the presence of the wild-type gene) based on the size of the PCR products. If the deletion was done the following PCR products will be present: a ~400 bp band using the primer mixA and a 300pb band using the primer mix B. One colony of tSL40 was used as control. tSL40**Δ**UPF1 colonies n°1 and 2 showed the expected band pattern, while the colony 37 turned out to be a false positive showing the same PCR products using primer mix D and C as the tSL40 control.

The novel tSL40ΔUPF1 strain is bound to be an essential resource in order to set up the PISA 2.0 Technology. By using the NMD-deficient tSL40, the incorporation of ncAA is expected to increase more than 2-fold (in a manner that correlates to the UAG position) compared to the WT strain[183,199].

Accordingly, I exploited the tSL40ΔUPF1 to test whether the stability of the mRNA containing the UAG was the factor that prevented the efficient amber suppression of the H3K9* bait. To test this hypothesis, a Real-Time qPCR (RT-qPCR) was performed in order to measure the

transcript abundance of the H3K9* target in the tSL40ΔUPF1 strain compared to the WT UPF1+. In the experiment a natural target of the NMD pathway in *S.cerevisiae*, namely CYH2 gene[15], was also included to independently validate the NMD pathway inactivation in the tSL40ΔUPF1 [211,264]. The *S. cerevisiae* CYH2 pre-mRNA (the NMD-dependent transcript) is normally inefficiently spliced, transported to the cytoplasm, and degraded by the NMD pathway for the reason that it contains an early in frame stop codon: in a strain lacking UPF1 function, the CYH2 pre-mRNA is normally stabilized 2- to 5-fold [211].
Accordingly, primers specific for both i) the CYH2 cDNA sequence containing the intron between exon 1 and exon 2 (referred to as pre-CYH2) and ii) the CY2H cDNA coding sequence (correctly spliced/NMD independent transcript) were designed and used as internal controls for the Real-Time qPCR experiment.
As shown in **Figure 3.20 B** the analysis revealed that the pre-CYH2 transcript was stabilized and increased by approximately 3-fold in the KO strain compared to the WT, whereas the CY2H mRNA level remained stable as expected. This experiment allowed me to confirm the lack of the NMD activity in the generated tSL40ΔUPF1 strain. However, no difference in the expression level of the H3K9* was detected comparing the WT to the tSL40ΔUPF1 strain.

For the sake of completeness, I also repeated the encoding N-Aceyl-Lysine incorporation experiment and WB analysis to verify the potential production of the H3K9ac protein. tSL40ΔUPF1 cells bearing pMICBD1-HA-H3K9*-LexA/2xtRNA$_{CUA}^{Pyl}$ were grown in liquid medium supplemented with 10mM of Ac-Lys, as protocol. Then, both yeast extract for direct immunoblot detection and Immunoprecipitated proteins were employed to assess the constitutive acetylation. No significant increase was detected while using the new strain. No detectable level of the read-through protein was found neither in the direct input nor in the immunoprecipitated extract **[Figure 3.20 A].** Taking everything into account, these results indicated that the stability of the target mRNA was not the factor that was hampering the amber suppression of the H3K9* bait, strongly suggesting that the knot in the inefficiency of the system might lie in the activity of the AcKRS synthetase variant itself.

[15] Natural mRNAs regulated by the NMD pathway have been identified in yeast. For instance, some intron-containing pre-mRNAs are NMD targets. The generation of a functional mRNA depends on precise processing of the pre-mRNA. Pre-mRNAs containing introns undergo splicing prior to export to the cytoplasm. In some cases, pre-mRNA processing events result in the retention of an inefficiently spliced intron. The majority of intron-containing pre-mRNAs that are exported to the cytoplasm are degraded by the NMD pathway in *S. cerevisiae*, because these pre-mRNAs are likely to have an in frame stop codon. For example, the S. cerevisiae CYH2 pre-mRNA is inefficiently spliced, transported to the cytoplasm, and degraded by NMD.

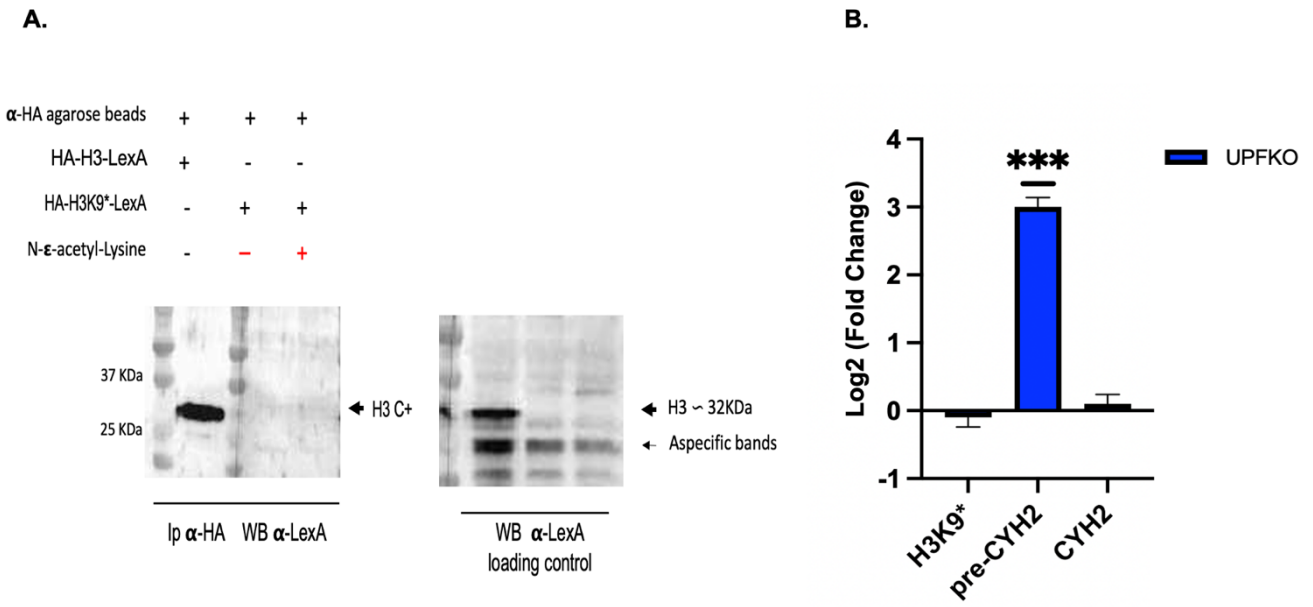


Figure 3.20. Validation of the NMD-deficient tSL40 yeast strain A. Immunoprecipitation with anti-HA agarose of soluble protein extracts from tsL40ΔUPF1 yeast intracellularly expressing either HA-H3K9*-LexA grown in both presence and absence of 10mM N-Acetyl-Lysine or HA-H3-LexA (~32KDa positive control). Membrane blotted for LexA. The last three lanes show a fraction of the input used. No detectable level of the H3 read-through protein was found neither in the direct input nor in the immunoprecipitated extract. **B. Validation through RT-qPCR of the UPF1KO strain and quantification of the H3K9* bait relative expression.** Log2FC +/- SD values of real-time PCR for the chosen targets (H3K9*, pre-CYH2, CYH2) in the tSL40ΔUPF1 strain. Values for each treatment are representative of n=3 biological replicates. For real-time PCR data: ΔCt = Ct GENE – Ct TAF10. log2(FC) = -ΔΔCt normalized on thousekeeping gene (TAF1O) and tSL40 WT control samples. SD = radq(varΔCt(KO))+(varΔCt(WT)). Student t-test (UPF1K0 vs WT, two tails) was performed on ΔCt values. Calculated P values were adjusted with the Benjamini-Hochberg procedure.

3.2.5 Screening of the AcKRS3 mutant Library obtained by Randomized Mutagenesis

Next, in the attempt to improve the translational activity of the MbAcKRS3/tRNA$^{Pyl}_{CUA}$, I aimed to evolve a more active synthetase using directed evolution. This technique, actually my personal favor, mimics natural evolution and involves two steps:
i) Diversification: the gene encoding the protein of interest is mutated at random to create a large (high- complexity) library of gene variants. ii) Selection: The library is tested for the presence of mutants (variants) that have the desired property using a selection that automatically eliminates all non-functional mutants.
As a mutational strategy, I decided to use error prone PCR (ePCR), a modification of standard PCR methods. The reaction mixture was modified to enhance the natural error rate of the polymerase that is amplified for a set number of cycles according to the required level of mutagenesis. The protocol of ePCR I decided to employ (Wilson & Keefe, 2001) took into serious consideration the issue of the saturation of the reaction and depletion of reagents and solved it in a very simple manner. A small aliquot of the reaction solution was taken after a set number of cycles and diluted in fresh reaction mixture to reduce the amplicon level and maintain the duplication process in a continuous exponential phase. In this embodiment, the average number of mutations per template could be controlled as a function of ePCR dublings performed. I was attempting to generate an enhanced catalytic activity, so I required multiple mutations per template. Thanks to pilot experiments [265], I decided to use the library derived from the reaction stopped after the 16th cycle, in which I expected more or less 15 aa mutations per ORF taking into account the expected number of mutations under similar conditions for a random mutation in a DNA stretch of similar size [257] **[Figure 3.21 A]**. Mutagenesis of the gene encoding for AcKRS3 was achieved in a mutation prone buffer with manganese ions, low magnesium, unbalanced dNTPs concentrations and a low fidelity DNA polymerase. Both low magnesium and the presence of manganese ions affect the efficiency of magnesium ions as cofactors of the polymerase by competition or by sheer low availability, while the unbalanced dNTP concentration favors mutations by scarcity of substrate and the deliberate usage of a low fidelity polymerase further increases the mutation rate.
Then I performed a large-scale ligation to clone the library in the pRSII422 vector (see Material and Methods). During the preliminary experiments, I observed a high degree of concatemerization of the reagents during ligation. These byproducts must be contained because, they would damage the transformation efficiency by reducing the quantity of the effective properly ligated product. Concatemers of the insert were particularly of concern because a plasmid with multiple copies of the insert would actually carry an antibiotic

resistance gene that allows a cell to survive. After screening for functional variants, it would be impossible to discriminate which of the copies of the synthetase gene present in these objects was the one that conferred the yeast the ability to survive under the selective pressure. So, to obtain the relative maximal yield of the desired ligation products, I designed an unusual protocol where the insert was subjected to the CIP treatment instead of the common practice of CIP-treating the plasmid backbone. In addition, the insert was employed in an equimolar ratio to the backbone vector. The result of this procedure constituted the pRSII422-AcKRS3 mutant library **[Figure 3.21 B].**

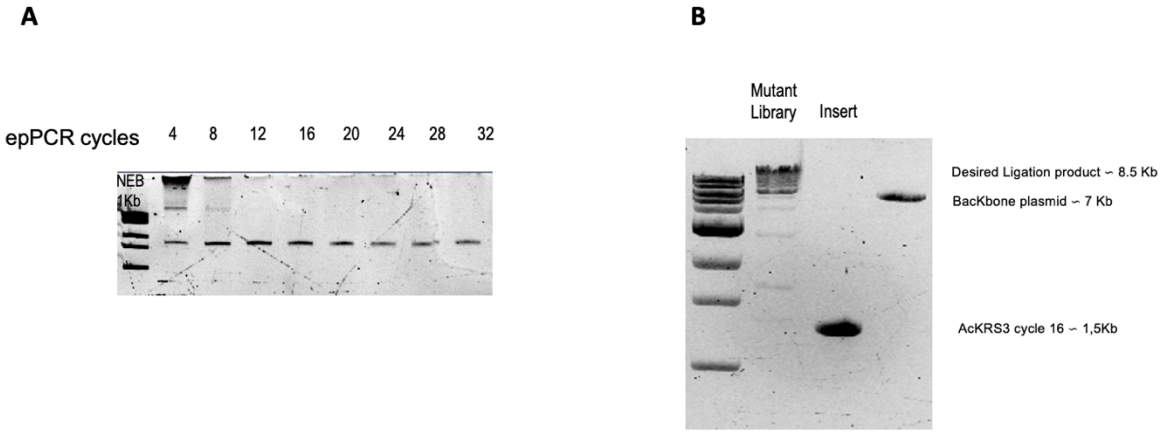


Figure 3.21. Creation of AcKRS mutant Library obtained by Randomized Mutagenesis. (A) Amplicons present in the error prone PCR reaction solution at different time points of the PCR amplification. Only bands at the expected molecular weight (1.5Kb) for the synthetase amplicon were present. **(B) epPCR Library after ligation.** Gel electrophoresis of the ligation products (Library) of mutant AcKRS obtained after 16th cycles of epPCR to the backbone pRSII42-tRNA^{Pyl} vector. The insert fragment used in the ligation was treated with Calf Intestinal Phosphatase (CIP) after the restriction enzyme digestion. Formation of backbone concatemers was expected and unavoidable, but did not hinder the selection efficiency.

As a selection method, I decided to use a life-death assay. Survival assays are the most straightforward type of assays, given that only the cells that carry a functional copy of the mutated gene are allowed to survive. I took advantage of the TRP essential gene. The TRP1 gene encodes phosphoribosylanthranilate isomerase, an enzyme that catalyzes the third step in tryptophan (W) biosynthesis. W was particularly attractive for two main reasons: i) L40 was auxotroph for W, ii) the low rate of reversion mutation occurring in this gene [266]. Accordingly, to set up a selection approach which would couple the survival of the cell with the suppression by the pair, I created a pMICBD1-W36* reporter plasmid in which a TAG codon replaced the Lys at position 36 of the Tryptophan (W) marker, essential for growth.

As a first attempt, the mutagenized library was directly co-transformed in yeast together with the reporter plasmid. Then I performed selection and replica plated on W deprived medium plate supplemented with AcK. A dozen of colonies appeared only after > 10 days. Among all sequenced clones, I found only revertants (false positive), probably due to the severe selection method. Sanger sequencing was used. Of course, this kind of phenotype-based selection are sensitive to many factors and further improvements are needed. In particular a deep sequencing of the library capable of covering the full sequences of genes encoding aaRSs and the development of a more efficient selection method will be the objects of future study. Nonetheless, all these findings suggested that the actual factor limiting the amber suppression efficiency may lie in the specific aminoacyl-tRNA synthetase variant employed.

3.2.6 Finding a superior o-aaRS/tRNA candidate: aaRS/tRNA pair comparison in *S.cerevisiae*

Since all the collected results in the tSL40 strongly indicated a suboptimal catalytic activity of the *Mb*AcKRS variant employed until now, I decided to address this issue by changing the synthetase/tRNA variant: other aaRS variants reported in literature [123,145,172] were analyzed in order to compare them under the same experimental conditions and choose the best working one in our strain. A brief description of these orthogonal translational systems (OTSs) follows. Then, after having checked the transcription and expression of the components and the bioavailability of their specific ncAAs, two assays were performed to measure their translational activity.

First, I changed the synthetase for encoding AcK, using the *M.mazei* AcKRS/$tRNA_{CUA}^{Pyl}$ to investigate if it would outperform the *M.barkeri* one (used until now). At the same time, I also tested the newly discovered single domain PylRS/$tRNA_{CUA}^{Pyl}$ pairs from *Methanomassiliicocales* that were shown to be very active and orthogonal in *Escherichia coli*[162,267]. The aminoacyl-tRNA synthetases of these pairs lack the N-terminal domain, so they are smaller in size and more soluble than the *Methanosarcina* PylRSs but share a homologous active site with them. In particular I employed the Methanomassiliicocales evolved variants (Me-HisRS-1/ $tRNA_{CUA}^{Pyl13C}$, Me-HisRS-2/ $tRNA_{CUA}^{Pyl6C10}$, Me-HisRS-3/ $tRNA_{CUA}^{Pyl6B03}$) kindly provided by Dr. Jason Chin (MRC-LMB, Cambridge UK), to verify if they indeed increased the ncAAs incorporation also in yeast. These synthetases were previously evolved to direct the incorporation of 3-*methyl*-L-*histidine* (Me-His) in mammalian cells (unpublished data). If they had been found to display a significant activity, I would have tried to evolve them to recognize

the ncAAs of our interest, using methods previously used [116,117,162]. Additionally, to further explore the potential of the genetic code expansion technology in our yeast strain, I seized the opportunity to analyze also the activity of the pair to encode Phospho-serine (*Mmp*SepRS/tRNA$_{CUA}^{Sep}$) [145]. As positive control the functional amber suppressor *Ec*TyrRS/tRNA$_{CUA}^{Tyr}$ pair was used [125].
From a methodological point of view, the pairs were expressed using the pRSII422 2μ vector in L40 (or L40ΔSER2 respectively). The aaRS expression was assessed by Western Blot with an anti-Myc tag mAb: all of the proteins were expressed at the expected molecular weight (data not shown). Northern Blot analysis confirmed the expression of the heterologous tRNAs. Then, before measuring the translational activity, the intracellular concentration of their corresponding ncAAs was assessed in the yeast cells.

Testing ncAAs intracellular concentration

Before proceeding with the testing of the new pairs, I thought that it would be important also to assess the bioavailability of the novel ncAAs to yeast cells. NcAAs structurally close to canonical amino acids (i.e.: acetyl-Lysine) may enter cells through endogenous amino acid transporters when added to the growth media, yet those deviating significantly or highly charged (e.g.: phosphorylated ncAAs) may not [142]. Thus, before testing the above mentioned new aaRS/tRNA pairs, the intracellular concentration $[c]_{int}$ of the PTM-AAs in our yeast strains was measured by Mass Spectrometry (MS), as previously done for the AcK solely [Chpt. 3.2.2].

In the following experiments (beside acetyl-Lysine) phospho-serine and its non-hydrolysable-analog 2-amino-4-phosphonobutyric acid were used. Methyl- Histidine (Me-His) was also employed as it is the substrate of the Methanomassiliicocales variants. Table 3.4 lists all the PTM-AAs tested and the $[ncAAs]_{int}$ measured in yeast samples grown in minimal media supplemented by those PTM-AAs. The intracellular concentration was in the working range (high micromolar range) for all the PTM-AAs tested in this study, except for phosphoserine (Sep). Phosphoserine is an intermediate in the biochemical pathway for the biosynthesis of Serine and is therefore rapidly depleted in vivo in wild-type L40 cells: phosphoserine is synthesized from 3-phosphohydroxypyruvate by phosphoserine aminotransferase (encoded by *SER3*), and converted to serine by phosphoserine phosphatase (encoded by *SER2*) [145,176]. Besides, its non-hydrolysable-analog 2-amino-4-phosphonobutyric acid poorly entered the L40 cell. For these reasons, to increase the endogenous level of Sep, a Phosphoserine Phosphatase-deficient strain L40ΔSER2 was generated by knocking-out the SER2 gene from the L40 genome following the same procedure as per the tSL40ΔUPF1. As predicted, the [pSer]int in L40ΔSER2 raised up to 3,7mM. The strain will be used as a platform to express the SepRS/tRNA$_{CUA}^{Sep}$ in all the experiments.

Table 3.4. List of the Non-Canonical Amino Acid intracellular concentration used in this thesis.

Yeast Strain	NcAAs	$[ncAA]_{ext}$	$[ncAA]_{int}$
S.cerevisiae L40	Acetyl-Lysine	10 mM	1.1 mM
S.cerevisiae L40	Phospho-Serine	N.A[#]	9 uM
S.c ΔSer2 L40	Phospho-Serine	N.A[#]	3.7 mM
S.cerevisiae L40	Metyl-Histidine	5mM[#]	4.4mM

Legend. [#]N.A.= Not Applicable

aaRS/tRNA pair comparative assays in *S.cerevisiae*

At this point, the new pairs were tested under the same experimental condition using two fast and well stated functional assays: i) a phenotyping assay that coupled the survival of the cells with the suppression by the pairs and ii) the highly sensitive read-through GFP assay.

Phenotyping assay with Mav203 cells

I decided to replicate a previously reported and well characterized yeast assay to investigate the amber suppressor activity of the orthogonal aaRS/tRNA of our interest [117,120,125]. The method is based on activation of GAL4-responsive HIS3, URA3, or lacZ reporter genes by suppression of amber codons in GAL4 transcriptional activator protein; it was initially established to isolate new aminoacyl-tRNA synthetases in *Saccharomyces cerevisiae* by Dr. Jason Chin and coworkers [120].

In this work, the *S.cerevisiae* selection strain MaV203:pGADGAL4(2TAG) cells was used; this strain contains a GAL4 transcriptional activator gene bearing two amber codons and a HIS3 GAL4-activated promoter, hence it is auxotrophic for histidine. When a functional amber suppressing pair, such as the *Ec*TyrRS/$tRNA_{CUA}^{Tyr}$ pair, is transformed into this strain, full

length GAL4 is produced, leading to activation of HIS3 genes. Transcription of this gene allows cells to grow in the absence of histidine **[Figure 3.22A].**

Phenotyping was performed as described in Chin et al. [120]. Briefly, *S. cerevisiae* MaV203:pGADGAL4(2TAG) was transformed by the Lithium Acetate Method with the pRSII422 bearing the aaRS/tRNA$_{CUA}$ pairs under investigations. Overnight cultures were serially diluted and replica plated onto selective media (YC- His/-Ade) in the presence or absence of 10 mM of ncAAs, as necessary **[Figure 3.22 B].** Mav containing pGADGAL4(2TAG) and *Ec*TyrRS/tRNA$_{CUA}^{Tyr}$ grew robustly at 10^5 cell dilution, as expected. The cells expressing the AcKRS variants were almost unable to grow in media lacking histidine supplemented with 10mM AcK, suggesting that these constructs are not fully functional. A better growth phenotype was observed for Me-HisRS pairs, since individual clones were present at lower dilutions. Nevertheless, this growth phenotype was still too poor and suboptimal compared to the *Ec*TyrRS/ tRNA$_{CUA}^{Tyr}$ mediated amber suppression **[Figure 3.22 B]**.

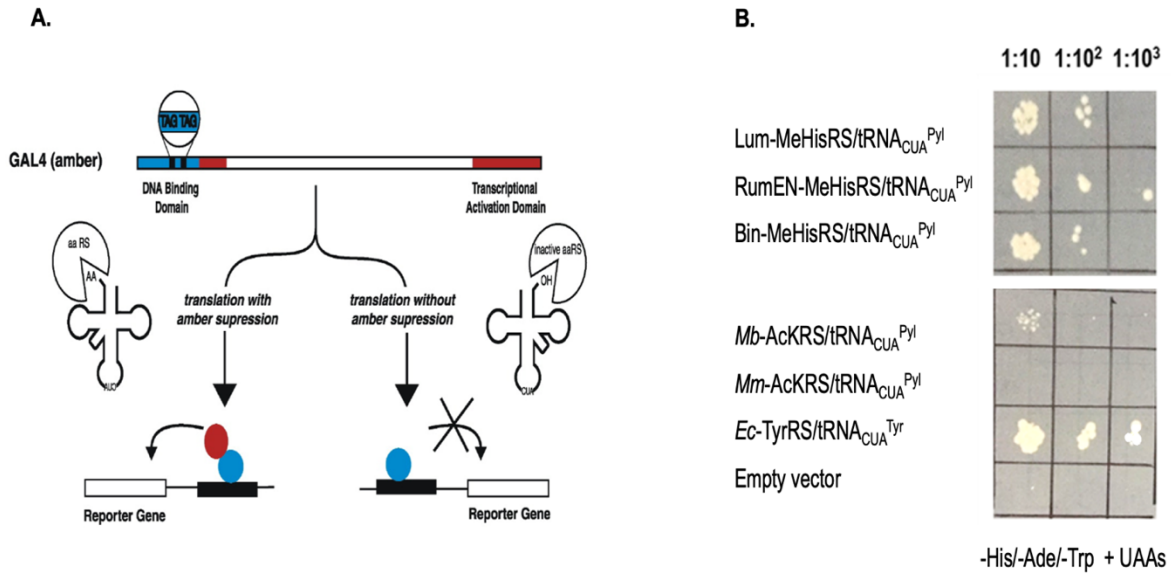


Figure 3.22. Phenotyping constructs for amber suppression in MaV203:pGADGAL4(2TAG) cells. The growth phenotype in presence of the OTSs under investigation was suboptimal compared to that obtained with the functional OTS *Ec*TyrRS/ tRNA$_{CUA}^{Tyr}$. A) Schematic of the GAL4 reporter. The reporter (encoded from a vector with Trp selection marker) is derived from full-length GAL4 (1–881) and contain amber codons at positions in the DNA binding domain. Activated transcription of His gene is driven by amber suppression of TAG codons in GAL4 by an active aminoacyl tRNA synthetase (aaRS)/tRNA pair. **B) Phenotyping experiment.** Serial dilutions of cells contained the appropriate aaRS/tRNA expression constructs were spotted onto selective plate, where ncAAs were used at 10 mM.

Yeast monomeric GFP read-through assay

Besides Western blot analysis, to quantify the ncAAs charging activity of the OTSs in yeast, the GFP read-through assay was chosen. This is a fluorescence highly sensitive test based on a GFP reporter with an in-frame TAG codon at position 39 (a site known to be permissive for diverse ncAAs). The PTM-AAs incorporation is verified with the generation of the read-through GFP fluorescence directly in intact cells: green fluorescence of GFP should be detected in cells only in presence of PTM-AAs. In the absence of the unnatural amino acid, no green fluorescence is detected **[Figure 3.23 B].** Surprisingly, the most used GFP (EGFP from *Aequorea* species) exhibited both extremely low expression and fluorescence levels in *S. cerevisiae*. So, instead of the commercially available EGFP, I made use of the codon-optimized monomeric green fluorescent protein mUkG1 from soft coral (referred for simplicity as ymGFP) [268], which showed a more intense fluorescence (and green color was observed visually under natural light conditions in cells expressing these proteins) **[Figure. 3.23 A].**

Each of the pRSII422 plasmid containing the orthogonal pairs and the vector expressing the ymGFP39* were co-transformed in the wt strain (or L40ΔSerB for testing the *Mmp*SepRS/tRNA$_{CUA}^{Sep}$) with LiOAc method. The transformation was performed onto selective plates. After 4 days, 3 single colonies (3 biological replicates) were grown in the appropriate YC liquid medium for selection, with or without 10mM ncAA, as necessary. L40 cells transformed with either the pMICBD1-ymGFP39* or the pMICBD1-ymGFP full length solely were used as negative and positive controls respectively to adjust the detection sensitivity of the spectrophotometer. After 24h, the fluorescence intensity was measured, and the data reported in the grouped bar plot **[Figure 3.21 C].** Compared with the WT control, translation of UAG codon by *Ec*TyRS was 65%. None of the other pairs produced a detectable level of GFP when used to suppress the TAG codon in the GFP gene. Also, this comparative study indicated that the aaRS/tRNA pairs tested showed a suboptimal translational activity in the L40 yeast.

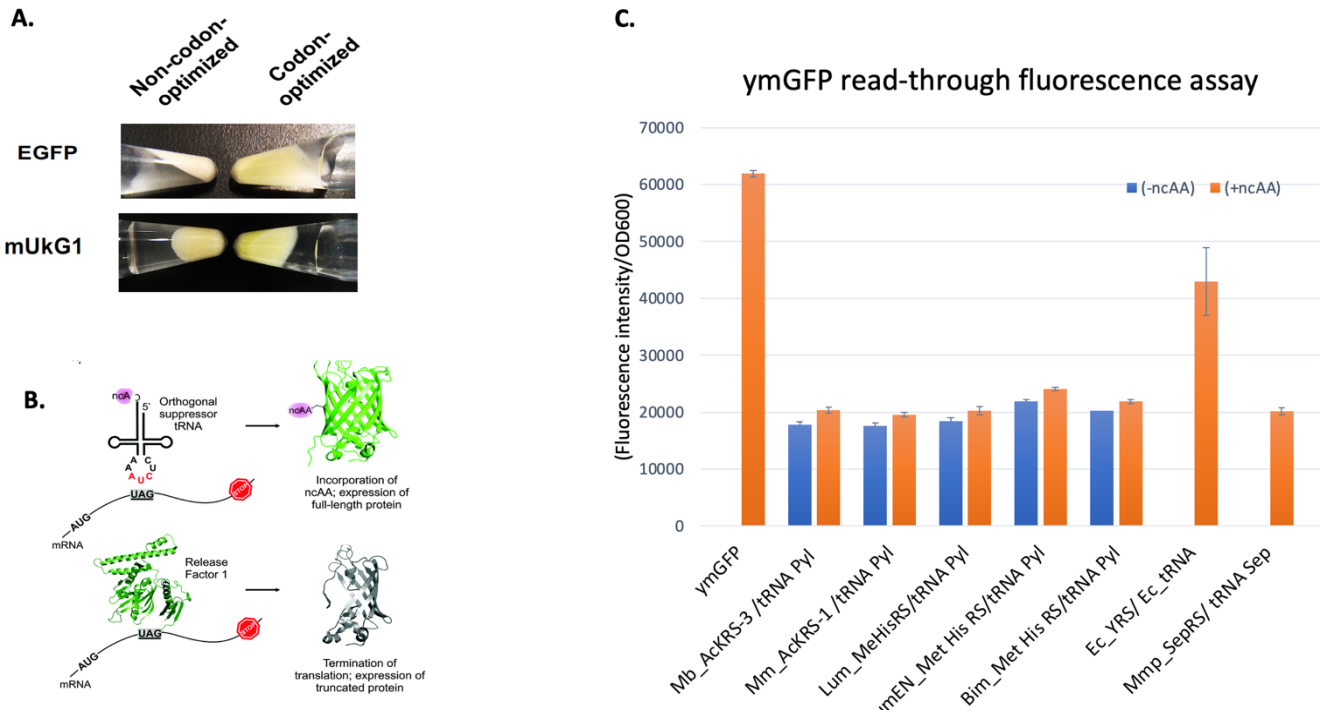


Figure 3.23. The Read-Through GFP Functional Assay shows that the aaRS/tRNA pairs tested have a suboptimal translational activity in the L40 yeast. A. Visual images of the yeast cells expressing GFPs under natural light. EGFP and ymUKG1 codon-optimized and non-codon-optimized GFPs were used in the evaluations. **B. Schematic representation of the GFP Read-through Assay.** It based on a GFP reporter with an in-frame TAG codon at position 39 (a site known to be permissive for diverse ncAAs). The PTM-AAs incorporation is verified with the generation of the read-through GFP fluorescence directly in intact cells: green fluorescence of GFP should be detected in cells only in presence of PTM-AAs. In the absence of the unnatural amino acid, no green fluorescence is detected **C. Translation of the ymGFP reporter (UAG codon at position 39) by the aaRS/tRNA pairs tested in this study was measured by fluorescence intensity.** Fluorescence signals are represented as bars, they are provided by the presence or absence of ncAAs and per each pair. Bar A1 is a positive control to detect the production signal of WT ymGFP (set as 100%). Bars B and C: fluorescence signal from incorporation of AcK by MbAcKRS-3/tRNA^{Pyl} and MmAcKRS-1/tRNA ^{Pyl} respectively. D, E, F : fluorescence signals from incorporation of Me-His by Me-His aaRS/tRNA^{Pyl} variants. G: fluorescence signal from incorporation Tyr by EcYRS/tRNA ^{yr}. Bar H: fluorescence signal from incorporation of p-Serine by SepRS/tRNA^{Sep}. The mean values (normalized with OD600) were calculated from three replicates. Data rappresented SD for at least three biological replicates.

3.2.7 tRNA EXtension (tREX) analysis to measure tRNA aminoacylation

In light of the previous results, I needed to determine whether the bottleneck in the application of the amber suppression systems in yeast was upstream or downstream the aminoacylation process. The former case would have involved too many different mechanisms (not easily separable confounding factors) to be studied and assessed. To name just one, the synthetases might be poorly folded due to the different physical parameters required for the yeast growth (e.g.: growth temperature, pH etc.). Additionally, non-canonical localization sequences in the archaea aaRSs might also cause their segregation from the cytosol. Also, the tRNAs might lack important modifications, preventing them from being correctly processed or aminoacylated. On the contrary, if the bottleneck was downstream, I would have been able to act on the system for instance by performing molecular evolution onto the translation machinery components [113,269]. To solve this knot, I used a rapid and scalable approach called tRNA extension (tREX) [214] to empirically determine the *in vivo* expression and aminoacylation status of the o-tRNAs. The protocol involves three steps: oxidation, hybridization and extension. The fate of the tRNA of interest differs depending on whether it is aminoacylated or not **[Figure 3.24 A].** Total tRNA extracted from cells expressing the pairs was treated with $NaIO_4$ at pH 5 to selectively oxidize the diol on the ribose at the 3'-end of non-aminoacylated tRNAs. Conversely, aminoacylation is maintained at this pH and aminoacylated tRNAs are not oxidized (Oxidation step I). Fluorescently labeled DNA probes that anneal to the 3'-end of their target tRNA were designed such that annealing to tRNA would produce a 5'-overhang of ssDNA. The probe annealing occurs at pH 7.9, at this higher pH tRNAs are deacylated (Hybridization step II). The exo- Klenow fragment *E. coli* DNA polymerase I was added to the hybridization solution (Enzymatic extension III). tRNAs which where aminoacylated before oxidation are extended by the exo- Klenow fragment *E. coli* DNA polymerase I using the probe as template, whereas oxidized tRNAs, that were never aminoacylated, cannot be extended **[Figure 3.24 A]**. The extension leads to a decreased electrophoretic mobility in native PAGE, it is therefore possible to read out the aminoacylation status via the ratio of extended to non-extended DNA/tRNA hybrid. As a control of the mobility of the non-aminoacylated tRNA, the extract is treated with NaOH (which hydrolyses ester bonds between the amino acid and the tRNA) before the tREX protocol. **Figure 3.24 B** shows tREX experiment onto the o-tRNAs used in this thesis project (*Mb*-tRNA^{Pyl}, *Mm*-tRNA^{Pyl}, *Ml*-tRNA$^{Pyl\ 13C}$, *Ma*-tRNA$^{Pyl\ 6C10}$, *Ma*-tRNA$^{Pyl\ 6B03}$, *Mmp*-tRNA^{Sep}**)** .

Lane 1 shows a control for probe specificity with respect to the native *Sc*-tRNAs. Lane 2 and 3 show references for the electrophoretic mobility of the extended or native tRNA obtained by

not oxidizing the extract or by oxidation after total de-acylation with NaOH. Lanes 4 shows the actual tREX experiment (Steps I, II, III) **[Figure 3.24 B].** On the grounds that the cognate probe of our *Ec*-tRNA^{Tyr} displays non-specific binding to *Sc.* total tRNA extract, our positive control for tREX experiment in *S.cerevisiae* cannot be interpreted. However, since each experiments had its complete internal control groups (lanes 1-2-3), the pattern in the lanes 4 unequivocally indicated that none of the tRNAs tested was aminoacylated or that only a minor and not detectable portion of the o-tRNAs was aminoacylated. As a matter of fact, the mobility is the same as the de-acetylatd sample.

Importantly, the tREX experiment highlighted that the bottleneck in the system in yeast is upstream the aminoacylation step. Further investigation on the possible reasons for this phenomenon, however, was not part of my current research (being mainly focused on the PISA 2.0 platform).

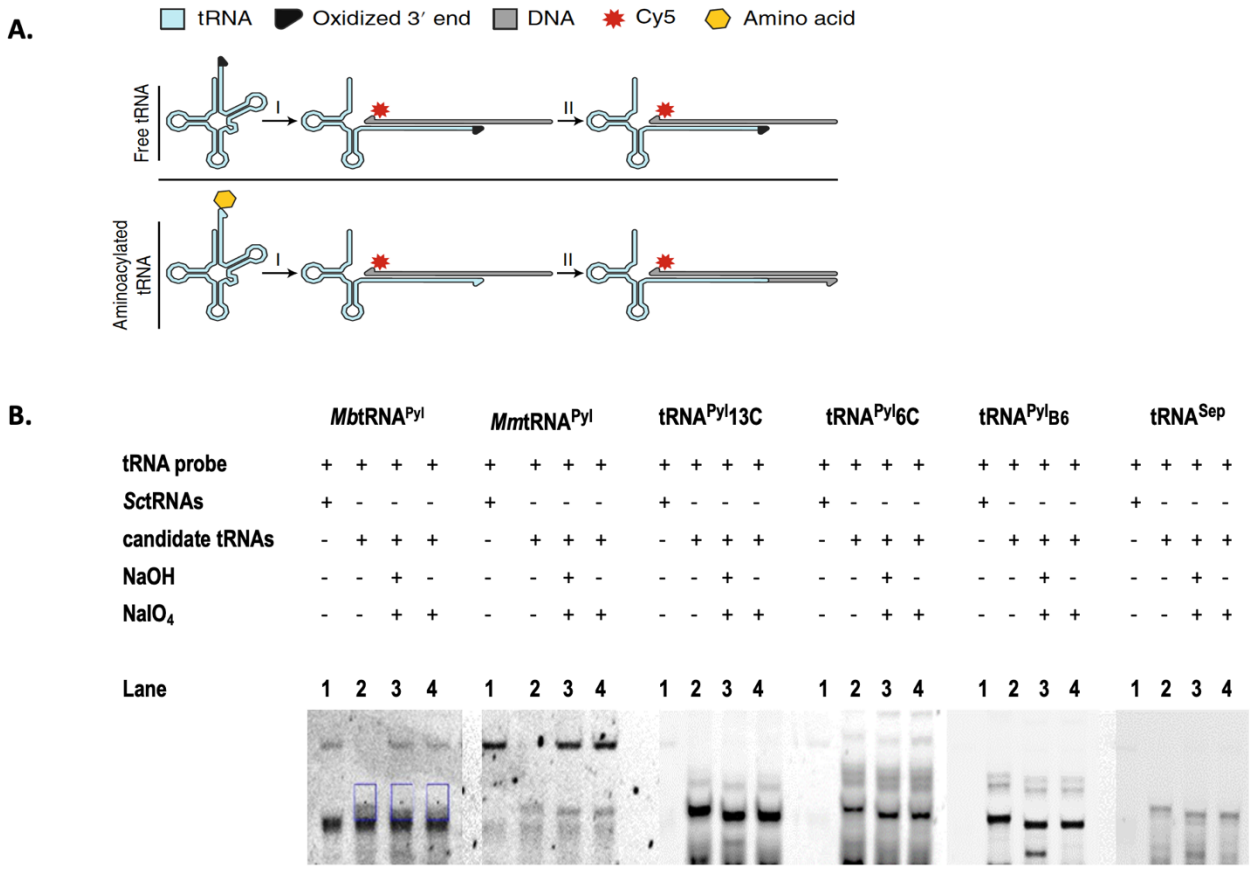


Figure 3.24. tREX Experiment on Mb-tRNA^{Pyl}, Mm-tRNA^{Pyl}, tRNA$^{Pyl\ 13C}$, tRNA^{Pyl6C10}, tRNA$^{Pyl\ 6B03}$, tRNA^{Sep}. None of the tRNAs tested was abundantly aminoacylated. A) Schematic of tREX Experiment Steps: A total tRNA extract from cells contains a tRNA of interest (light blue). The fate of the tRNA of interest differs depending on whether it is aminoacylated with an amino acid (yellow hexagon) or not. Total tRNA is treated with NaIO4 at pH 5 to selectively oxidize the diol on the ribose at the 3′ end of non-aminoacylated tRNAs (black bar). Aminoacylation is maintained at this pH, and aminoacylated tRNAs are not oxidized. A DNA probe specific for the tRNA of interest (gray) with a fluorophore attached (red star) anneals to the tRNA at pH 7.9 (I: probe annealing); at this higher pH, tRNAs are deacylated. Next, the exo(–) Klenow fragment of E. coli DNA polymerase I is added to the hybridization solution (II: enzymatic extension). tRNAs that were aminoacylated before oxidation are extended by the polymerase using the probe as a template, whereas oxidized tRNAs, which were never aminoacylated, cannot be extended. **B) tREX Experiment on Mb-tRNA^{Pyl}, Mm-tRNA^{Pyl}, tRNA$^{Pyl\ 13C}$, tRNA^{Pyl6C10}, tRNA$^{Pyl\ 6B03}$, tRNA^{Sep}.** Specific tRNA probe-Cy5 labelled was added to all reaction. First lane: Sc. tRNA extract without heterologous tRNA. In all the other lane for each probe, the eterologous tRNA its cognate aaRS and the ncAAs are present. $NaIO_4$ indicates whether the oxidation was performed. Lane 1 shows a control for probe specificity with respect to Sc. tRNAs. Lane 2 and 3 show references for the electrophoretic mobility of the extended or native tRNA obtained by not oxidizing the extract or by oxidation after total deacylation with NaOH. Lanes 4 shows tREX experiment. The gel is native PAGE stained with SYBR Gold (green).The experiment was repeated twice with similar results.

Appendix I.

The PISA 2.0 bait vector: reversed LexA polarity pMICBD1 vector

Regarding the PISA 2.0 selection, I opted for modifying the bait plasmid pMICBD1 for a technical reason. Specifically, I constructed a new pMICBD1-MCS-LexA bait plasmid, in which the LexA DBD polarity (i.e., N-bait-LexA-C) was reversed compared to the canonical PISA pMICBD1 bait polarity (N-LexA-bait-C) **[Figure 3.25 A].** Most Y2H bait plasmid used N-terminally fused LexA DBD [262], on the contrary because here LexA is fused to the target C-terminus, the identification of PTM-antigen binders would occur only if the PTM-AA is incorporated, greatly accelerating the PISA 2.0 selection.

The suitability of pMICBD1-MCS-LexA vector for Y2H interaction was tested performing the IACT screening with a couple of bait-prey interactors previously tested in the lab (exploiting the canonical pMICD1). *α*-synuclein (*α*Syn) (a 140-residue protein whose aggregation is associated with Parkinson's disease) and both the Nanobodies NbSyn2 [270] and the NbSyn87 [271,272] were chosen. Their interactions, in fact, have been well characterized using a range of biophysical techniques: the NbSyn87 and NbSyn2 bind with nanomolar affinity to distinctive epitopes within the C-terminal domain of soluble *α*Syn, comprising approximately amino acids 118-131 and 137-140, respectively [270–272]. First, the transformation expression levels of the N- terminal fusion of LexA to *α*Syn was assessed in L40 by WB **[Figure 3.25 B-C].** No difference could be detected compared to the C-terminal fusion in the pMICBD1 version **[Figure 3.25 C].** So, the L40 expressing the fusion protein Syn-LexA was co-transformed using LiOAc method with either the Nb Syn2, the NbSyn87, or the Nb242 α-LexA (as positive control). Screenings were monitored up to 6d. No difference in transformation efficiency was observed. As shown in **Figure 3.25 D**, L40-Syn-LexA was able to grow onto –WHL media when co transformed with NbSyn2 and NbLexA242 thus confirming the functionality of the new construct.

These results showed that the bait–LexA fusions are suitable for two-hybrid studies and display positive and specific responses **[Figure 3.25].** Furthermore, the laboratory will benefit from such a construct since it might lead to the identification of partners missed during classical two-hybrid screens: most Y2H vector systems use N-terminally fused test domains, impairing any interactions which involve regions around the N-terminus of these proteins.

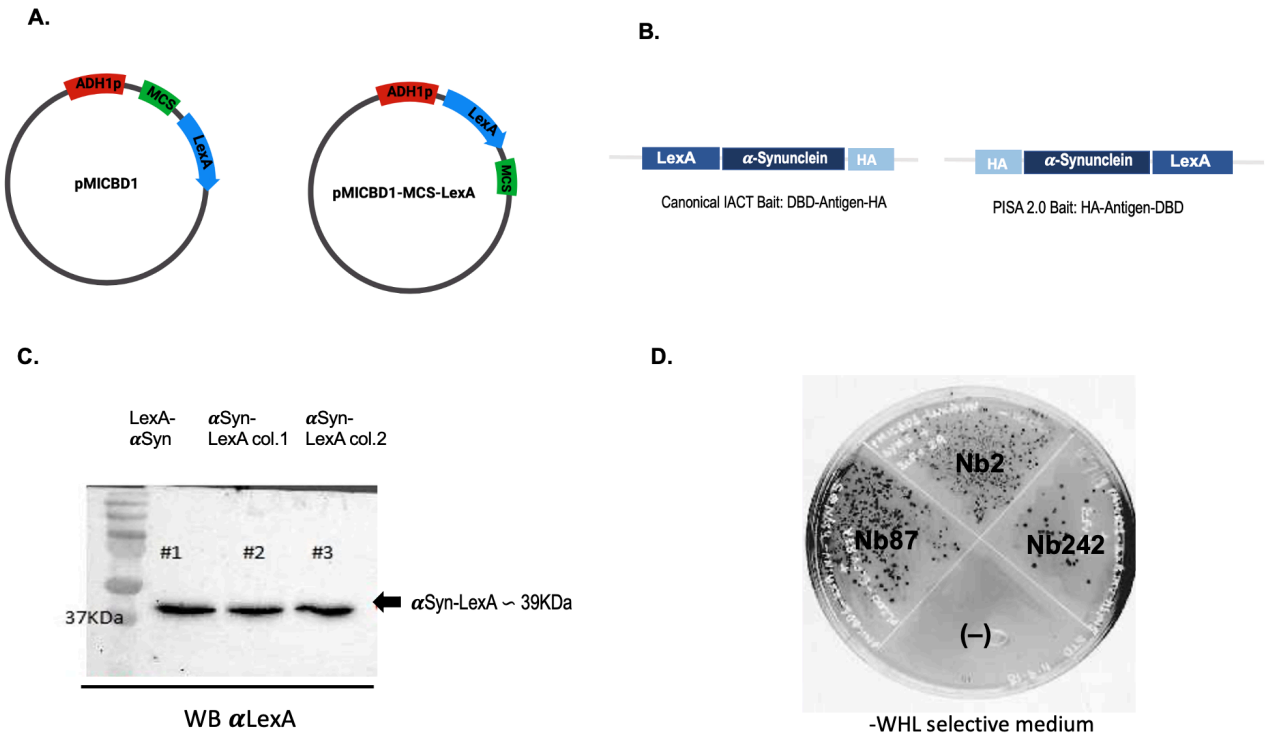


Figure 3.25. The novel pMICBD1-MCS-LexA vector: The Bait-LexA fusions are suitable for two-hybrid studies and display positive and specific responses. A) Main features of the pMICBD1-MCS-LexA compared to the pMICBD1 vector. In the new pMICBD1-MCS-LexA bait plasmid, the LexA DBD polarity (i.e., N-bait-LexA-C) was reversed compared to the canonical PISA pMICBD1 bait polarity (N-LexA-bait-C). ADH1p= ADH1 promoter; MCS= multi cloning site. **B)** Schematic of the IACT bait compared to the PISA 2.0 bait. In this example, the target antigen α-Synuclein is fused either to C-term of LexA DBD (IACT bait) or at the N-term of LexA (PISA 2.0 bait). The baits were expressed in yeast and their expression was checked by Western Blot. DBD= DNA Binding Domain; HA= HA tag **C)** WB analysis using anti-LexA antibody on protein extracts of L40 cells expressing either pMICBD1-LexA-**α**Syn or pMICBD1- **α**Syn-LexA bait (two clones were used). No difference in **α**Syn-LexA fusion proteins expression was detected between the N-terminal o C-terminal constructs. **(D) (Y2H)-IACT screening results**. Picture shows that **α**-Synuclein-LexA L40 colonies were able to grow onto selective medium –WHL plates when co-transformed with either Nbsyn2 or Nbsyn87, confirming the functionality of the new PISA 2.0 vector, pMICBD1-MCS-LexA. NbLexA242was used as positive control. Legend. (-) = ssDNA carrier DNA only.

Appendix II

Duplication of ADE2 gene in L40 strain: BLAST Alignment Result

Key: ade2 genomic locus

plasmidic ade2 marker

vector sequence

NdeI palindromic sequence

ADE2 promoter

Query: Frw ade#1

Sbjct: genomic ade#1

```
Query  1    TATATTGGGAGGGGGACAATTGGGACGTATGATTGTTGAGGCAGCTAACAGGCTCAACAT  60
            ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Sbjct  21   TATATTGGAGGGGGACAATTGGGACGTATGATTGTTGAGGCAGCAAACAGGCTCAACAT  80

Query  61   TAAGACGGTAATACTAGATGCTGAAAATTCTCCTGCCAAACAAATAAGCAACTCCAATGA  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  81   TAAGACGGTAATACTAGATGCTGAAAATTCTCCTGCCAAACAAATAAGCAACTCCAATGA  140

Query  121  CCACGTTAATGGCTCCTTTTCCAATCCTCTTGATATCGAAAAACTAGCTTAAAAATGTGA  180
            |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Sbjct  141  CCACGTTAATGGCTCCTTTTCCAATCCTCTTGATATCGAAAAACTAGCTGAAAAATGTGA  200

Query  181  TGTGCTAACGATTGAGATTGAGCATGTTGATGTTCCTACACTAAAGAATCTTCAAGTAAA  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  201  TGTGCTAACGATTGAGATTGAGCATGTTGATGTTCCTACACTAAAGAATCTTCAAGTAAA  260

Query  241  ACATCCCAAATTAAAAATTTACCCTTCTCCAGAAACAATCGGATTGATACAAGACAAATA  300
            |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Sbjct  261  ACATCCCAAATTAAAAATTTACCCTTCTCCAGAAACAATCAGATTGATACAAGACAAATA  320

Query  301  TATTCAAAAAGAGCATTTAATCAAAAATGGTATAGCAGTTACCCAAAGTGTCCCTGTGGA  360
            |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Sbjct  321  TATTCAAAAAGAGCATTTAATCAAAAATGGTATAGCAGTTACCCAAAGTGTTCCTGTGGA  380

Query  361  ACAAGCCAGTGAGACGTCCCTATTGAATGTTGGAAGAGATTTGGGTTTTCCATTCGTCTT  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  381  ACAAGCCAGTGAGACGTCCCTATTGAATGTTGGAAGAGATTTGGGTTTTCCATTCGTCTT  440

Query  421  GAAGTCGAGGACTTTGGCATACGATGGAAGAGGTAACTTCGTTGTAAAGAATAAGGAAAT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  441  GAAGTCGAGGACTTTGGCATACGATGGAAGAGGTAACTTCGTTGTAAAGAATAAGGAAAT  500

Query  481  GATTCCGGAAGCTTTGGAAGTACTGAAGGATCGTCCTTTGTACGCCGAAAAATGGGCACC  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  501  GATTCCGGAAGCTTTGGAAGTACTGAAGGATCGTCCTTTGTACGCCGAAAAATGGGCACC  560

Query  541  ATTTACTAAAGAATTAGCAGTCATGATTGTGAGATCTGTTAACGGTTTAGTGTTTTCTTA  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  561  ATTTACTAAAGAATTAGCAGTCATGATTGTGAGATCTGTTAACGGTTTAGTGTTTTCTTA  620

Query  601  CCCAATTGTAGAGACTATCCACAAGGACAATATTTGTGACTTATGTTATGCGCCTGCTAG  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  621  CCCAATTGTAGAGACTATCCACAAGGACAATATTTGTGACTTATGTTATGCGCCTGCTAG  680

Query  661  AGTTCCGGACTCCGTTCAACTTAAGGCGAAGTTGTTGGCAGAAAATGCAATCAAATCTTT  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  681  AGTTCCGGACTCCGTTCAACTTAAGGCGAAGTTGTTGGCAGAAAATGCAATCAAATCTTT  740

Query  721  TCCCGGTTGTGGTATATTTGGTGTGGAAATGTTCTATTTAGAAACAGGGGAATTGCTTAT  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  741  TCCCGGTTGTGGTATATTTGGTGTGGAAATGTTCTATTTAGAAACAGGGGAATTGCTTAT  800

Query  781  TAACGAAATTGCCCCAAGGCCTCACAACTCTGGACATTATACCATTGATGCTTGCGTCAC  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  801  TAACGAAATTGCCCCAAGGCCTCACAACTCTGGACATTATACCATTGATGCTTGCGTCAC  860

Query  841  TTCTCAATTTGAAGCTCATTTGAGATCAATATTGGATTTGCCAATGCCAAAGAATTTCAC  900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  861  TTCTCAATTTGAAGCTCATTTGAGATCAATATTGGATTTGCCAATGCCAAAGAATTTCAC  920

Query  901  ATCTTTCTCCACCATTACAACGAACGCCATTATGCTAAAT  940
             ||||||||||||||||||||||||||||||||||||||||
Sbjct  921  ATCTTTCTCCACCATTACAACGAACGCCATTATGCTAAAT  960
```

Query: Rv complement primer Rv ade#1

Subject: 3′ plasmidic ade#1

```
Query  1    TGTCAGCATATGCTATTTCCGCAAGCAAGCGTGGAATTAAAACAATTATCGCTGGAGCTG  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  682  TGTCAGCATATGCTATTTCCGCAAGCAAGCGTGGAATTAAAACAATTATCGCTGGAGCTG  623
```

```
Query  61   GTGGGGCTGCTCACTTGCCAGGTATGGTGGCTGCAATGACACCACTTCCTGTCATCGGTG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  622  GTGGGGCTGCTCACTTGCCAGGTATGGTGGCTGCAATGACACCACTTCCTGTCATCGGTG  563

Query  121  TGCCCGTAAAAGGTTCTTGTCTTGATGGAGTAGATTCTTTACATTCAATTGTGCAAATGC  180
            |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  562  TGCCCGTAAAAGGTTCTTGTCTAGATGGAGTAGATTCTTTACATTCAATTGTGCAAATGC  503

Query  181  CTAGAGGTGTTCCAGTAGCTACCGTCGCTATTAATAATAGTACGAACGCTGCGCTGTTGG  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  502  CTAGAGGTGTTCCAGTAGCTACCGTCGCTATTAATAATAGTACGAACGCTGCGCTGTTGG  443

Query  241  CTGTCAGACTGCTTGGCGCTTATGATTCAAGTTATACAACGAAAATGGAACAGTTTTTAT  300
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Sbjct  442  CTGTCAGACTGCTTGGCGCTTATGATTCAAGTTATACAACAAAAATGGAACAGTTTTTAT  383

Query  301  TAAAGCAAGAAGAAGAAGTTCTTGTCAAAGCACAAAAGTTAGAAACTGTCGGTTACGAAG  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  382  TAAAGCAAGAAGAAGAAGTTCTTGTCAAAGCACAAAAGTTAGAAACTGTCGGTTACGAAG  323

Query  361  CATATCTTGAAAACAAGTAATATATAAGTTTATTGATATACTTGTACAGCAAATAATTAT  420
            | ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  322  CTTATCTAGAAAACAAGTAATATATAAGTTTATTGATATACTTGTACAGCAAATAATTAT  263

Query  421  AAAATGATATACCTATTTTTTAGGCTTTGTTATGATTACATCAAATGTGGACTTCATACA  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  262  AAAATGATATACCTATTTTTTAGGCTTTGTTATGATTACATCAAATGTGGACTTCATACA  203

Query  481  TAGAAATCAACGCTTACAGGTGTCCTTTTTTAAGAATTTCATACATAAGATCTATGCGGT  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  202  TAGAAATCAACGCTTACAGGTGTCCTTTTTTAAGAATTTCATACATAAGATCTATGCGGT  143

Query  541  GTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAA  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  142  GTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAA  83

Query  601  TATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGC  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  82   TATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGC  23

Query  661  CGAAATCGGCAAA  673
            |||||||||||||
Sbjct  22   CGAAATCGGCAAA  10
```

```
Query: Frw plasmidic ade#2
Sbjct: plasmidic ade#2

Query  7     TAANTTTNNTTAAAAGAATCAAAGACAGATAAAATTTAAGAGATATTAAATATTAGTGAG  66
             ||| |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  55    TAACTTTTCTTAAAACAATCAAACACACATAAAATTTAACACATATTAAATATTAGTGAG  114

Query  67    AAGCCGAGAATTTTGTAACACCAACATAACACTGACATCTTTAACAACTTTTAATTATGA  126
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  115   AAGCCGAGAATTTTGTAACACCAACATAACACTGACATCTTTAACAACTTTTAATTATGA  174

Query  127   TACATTTCTTACGTCATGATTGATTATTACAGCTATGCTGACAAATGACTCTTGTTGCAT  186
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  175   TACATTTCTTACGTCATGATTGATTATTACAGCTATGCTGACAAATGACTCTTGTTGCAT  234

Query  187   GGCTACGAACCGGGTAATACTAAGTGATTGACTCTTGCTGACCTTTTATTAAGAACTAAA  246
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  235   GGCTACGAACCGGGTAATACTAAGTGATTGACTCTTGCTGACCTTTTATTAAGAACTAAA  294

Query  247   TGGACAATATTATGGAGCATTTCATGTATAAATTGGTGCGTAAAATCGTTGGATCTCTCT  306
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  295   TGGACAATATTATGGAGCATTTCATGTATAAATTGGTGCGTAAAATCGTTGGATCTCTCT  354

Query  307   TCTAAGTACATCCTACTATAACAATCAAGAAAAACAAGAAAATCGGACAAAACAATCAAG  366
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  355   TCTAAGTACATCCTACTATAACAATCAAGAAAAACAAGAAAATCGGACAAAACAATCAAG  414

Query  367   TATGGATTCAAGAACAGTTGGTATATTAGGAGGGGGACAATTGGGACGTATGATTGTTGA  426
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  415   TATGGATTCAAGAACAGTTGGTATATTAGGAGGGGGACAATTGGGACGTATGATTGTTGA  474

Query  427   GGCAGCAAACAGGCTCAACATTAAGACGGTAATACTAGATGCTGAAAATTCTCCTGCCAA  486
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  475   GGCAGCAAACAGGCTCAACATTAAGACGGTAATACTAGATGCTGAAAATTCTCCTGCCAA  534

Query  487   ACAAATAAGCAACTCCAATGACCACGTTAATGGCTCCTTTTCCAATCCTCTTGATATCGA  546
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  535   ACAAATAAGCAACTCCAATGACCACGTTAATGGCTCCTTTTCCAATCCTCTTGATATCGA  594

Query  547   AAAACTAGCTGAAAAATGTGATGTGCTAACGATTGAGATTGAGCATGTTGATGTTCCTAC  606
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  595   AAAACTAGCTGAAAAATGTGATGTGCTAACGATTGAGATTGAGCATGTTGATGTTCCTAC  654

Query  607   ACTAAAGAATCTTCAAGTAAAACATCCCAAATTAAAAATTTACCCTTCTCCAGAAACAAT  666
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  655   ACTAAAGAATCTTCAAGTAAAACATCCCAAATTAAAAATTTACCCTTCTCCAGAAACAAT  714
```

```
Query  667   CAGATTGATACAAGACAAATATATTCAAAAAGAGCATTTAATCAAAAATGGTATAGCAGT  726
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  715   CAGATTGATACAAGACAAATATATTCAAAAAGAGCATTTAATCAAAAATGGTATAGCAGT  774

Query  727   TACCCAAAGTGTTCCTGTGGAACAAGCCAGTGAGACGTCCCTATTGAATGTTGGAAGAGA  786
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  775   TACCCAAAGTGTTCCTGTGGAACAAGCCAGTGAGACGTCCCTATTGAATGTTGGAAGAGA  834

Query  787   TTTGGGTTTTCCATTCGTCTTGAAGTCGAGGACTTTGGCATACGATGGAAGAGGTAACTT  846
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  835   TTTGGGTTTTCCATTCGTCTTGAAGTCGAGGACTTTGGCATACGATGGAAGAGGTAACTT  894

Query  847   CGTTGTAAAGAATAAGGAAATGATTCCGGAAGCATTGGAAGTACTGAAGGATCGTCCTTT  906
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  895   CGTTGTAAAGAATAAGGAAATGATTCCGGAAGCATTGGAAGTACTGAAGGATCGTCCTTT  954

Query  907   NTACGCCGAAAAATGGGCACCATTTACTAAAGAATTAGCAGTCNNGATT  955
              |||||||||||||||||||||||||||||||||||||||||  ||||
Sbjct  955   GTACGCCGAAAAATGGGCACCATTTACTAAAGAATTAGCAGTCATGATT  1003
```

Query: Rv ext. ade#2
Sbjct: genomic ade2

```
Query  46    TACTTGTTTTCTAGATAAGCTTCGTAACCGACAGTTTCTAACTTTTGTGCTTTGACAAGA  105
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     TACTTGTTTTCTAGATAAGCTTCGTAACCGACAGTTTCTAACTTTTGTGCTTTGACAAGA  60

Query  106   ACTTCTTCTTCTTGCTTTAATAAAAACTGTTCCATTTTTGTTGTATAACTTGAATCATAA  165
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Sbjct  61    ACTTCTTCTTCTTGCTTTAATAAAAACTGTTCCATTTTCGTTGTATAACTTGAATCATAA  120

Query  166   GCGCCAAGCAGTCTGACAGCCAACAGCGCAGCGTTCGTACTATTATTAATAGCGACGGTA  225
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121   GCGCCAAGCAGTCTGACAGCCAACAGCGCAGCGTTCGTACTATTATTAATAGCGACGGTA  180

Query  226   GCTACTGGAACACCTCTAGGCATTTGCACAATTGAATGTAAAGAATCTACTCCATCTAGA  285
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181   GCTACTGGAACACCTCTAGGCATTTGCACAATTGAATGTAAAGAATCTACTCCATCTAGA  240

Query  286   CAAGAACCTTTTACGGGCACACCGATGACAGGAAGTGGTGTCATTGCAGCCACCATACCT  345
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241   CAAGAACCTTTTACGGGCACACCGATGACAGGAAGTGGTGTCATTGCAGCCACCATACCT  300

Query  346   GGCAAGTGAGCAGCCCCACCAGCTCCAGCGATAATTGTTTTAATTCCACGCTTGCTTGCG  405
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  GGCAAGTGAGCAGCCCCACCAGCTCCAGCGATAATTGTTTTAATTCCACGCTTGCTTGCG  360

Query  406  GAAATAGCATATGCTGACATCCTATGTGGAGTTCTATGAGCAGAGACTATTGTCACTTCA  465
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  GAAATAGCATATGCTGACATCCTATGTGGAGTTCTATGAGCAGAGACTATTGTCACTTCA  420

Query  466  AATGGAACGCCAAAATCTTTTAAAACCGCACATGCGGCAGACATTACCGGCAAGTCAGAG  525
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  AATGGAACGCCAAAATCTTTTAAAACCGCACATGCGGCAGACATTACCGGCAAGTCAGAG  480

Query  526  TCTGATCCCATGATGATTCCAACCAATGGTTTGACCATTGCTTCCAAGTCCAACTTTTGA  585
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  TCTGATCCCATGATGATTCCAACCAATGGTTTGACCATTGCTTCCAAGTCCAACTTTTGA  540

Query  586  GCGACAGAGATTTTGATTGGAATATCAGTTCTACCTGTAATGTAGTTCAGCCTTTGTTCA  645
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  541  GCGACAGAGATTTTGATTGGAATATCAGTTCTACCTGTAATGTAGTTCAGCCTTTGTTCA  600

Query  646  CATTCCGCCATACTGGAGGCAATAATATTTATGTGACCTACTTTTCTGTTAGGTCTTGAC  705
            |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Sbjct  601  CATTCCGCCATACTGGAGGCAATAATATTTATGTGACCTACTTTTCTGTTAGGTCTAGAC  660

Query  706  TCTTTTCCATATAAGTACACTGAGGAACCTGGAGTCGCCAATGCTCTTTCGCAAGTTTCT  765
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  661  TCTTTTCCATATAAGTACACTGAGGAACCTGGAGTCGCCAATGCTCTTTCGCAAGTTTCT  720

Query  766  AGCTCTTTATCTTTTGTATGTTTGTCTCCAAGAACATTTAGCATAATGGCGTTCGTTGTA  825
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  721  AGCTCTTTATCTTTTGTATGTTTGTCTCCAAGAACATTTAGCATAATGGCGTTCGTTGTA  780

Query  826  ATGGTGGAGAAAGATGTGAAATTCTTTGGCATTGGCAAATCCAATATTGATCTCAAATGA  885
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  781  ATGGTGGAGAAAGATGTGAAATTCTTTGGCATTGGCAAATCCAATATTGATCTCAAATGA  840

Query  886  GCTTCAAATTGNNANGTGACGCAAGCATCAATG  918
            |||||||||||  | ||||||||||||||||||
Sbjct  841  GCTTCAAATTGAGAAGTGACGCAAGCATCAATG  873
```

Key: ade 2 genomic locus

plasmidic ade2 marker

vector sequence

NdeI palindromic sequence

ADE2 promoter

3.3 The PISA 2.0 Project in yeast Discussion

3.3.1 Summary of main findings of the PISA 2.0 Project in yeast

My independent PhD work major advances have been reached in moving the first steps toward the implementation of the current PISA platform by exploiting the Genetic Code Expansion in order to accelerate and automatize the isolation of anti-individual PTM intrabodies in our yeast two-hybrid system. In summary, this part of my thesis work led to the following achievements, which include both new methodological findings and new biological insights.

- First, the Yeast-Two Hybrid (Y2H-) cells (the original host of PISA selection) have been successfully tailored *to genetically encoding N-acetyl-Lysine (AcK)* in the protein of interest via the stable integration of the corresponding orthogonal (o-)aaRS/tRNA pair. Also, the yeast strain was furtherly engineered to allow the incorporation of O-Phospho-serine (work in progress).
- Then, having reported low AcK incorporation efficiency, all the state-of-the-art technological procedures to improve the read-through protein yield have been systematically applied to the "acetylator" strain: a good PT-modified protein level is in fact paramount for the PISA 2.0 selection.
- Furthermore, the translational activity of recently discovered archaea aaRS/tRNA couples have been evaluated to find a fully active one in the yeast PISA 2.0 platform: a systematic comparative study of those OTSs was performed using a High-sensitive Fluorescence assay.
- Notably, also the cutting-edge tRNA-Extension Technology has been used for the first time in yeast cells to shed light onto the low read-through protein yields reported in this model organism.

At the end, since many interrelated factors turned out to impair the ncAAs incorporation in the PISA yeast strain, the proposed solution was to take advantage from a well stated bacterial-two-hybrid (B2H) system and the wide range of functional and orthogonal aaRS/tRNA pairs available in *E.coli* to set up the new 2.0 version of PISA (the main objective of this Project).

- Accordingly, the first proof of concept of the method in a Bacterial-Two-Hybrid strain in lieu of the Y2H original one has been achieved. A high translational activity of the aaRS/tRNA to encode N-acetyl-Lysine was in fact obtained in the B2H cells, which will be used for intracellular antibody domain selections.

In this section I will comment on obtained data and technical improvements. Moreover I will discuss the limitations encountered while applying the GCE technology in *L40 S.cerevisiae*, furnishing perspectives about possible solutions. I will further describe the first step toward the finalization of the PISA 2.0 project in *E.coli.*

3.3.2 Paving the way to the PISA 2.0 Project in yeast: The acetylator yeast strain data discussion

Having established an "acetylator" yeast strain, stably expressing the *Mb*AcKRS/tRNA$_{CUA}^{Pyl}$ pair to encode N-acetyl-Lysine, is one of the major achievements to pave the way to the PISA 2.0 project. On one hand, for the first time another letter (namely the AcK) was added to the alphabet of the L40 yeast strain. On the other hand, I set up the host organism of PISA 2.0 capable of expressing the acetylated targets at a desired position residue, although with low protein yield. Current methods for incorporating ncAAs in yeast in fact are based on transient expression. Because transient transfection experiments led to heterogenous expression levels, current approaches would have severely limited the possibility of implementing PISA 2.0.
Here, I described a method to integrate the tRNA/aaRS pairs in the yeast genome exploiting new generation non replicative yeast plasmid. Initially, I set the system for genetically encoding the non-canonical amino acid AcK, a widely found PTM with a regulatory role that rivals that of phosphorylation in eukaryotes. First, the *Methanosarcina barkeri (Mb)* AckRS/tRNA$_{CUA}^{Pyl}$ pair, which has emerged as a particularly versatile and orthogonal system in yeast *Saccharomyces cerevisiae*, was chosen and integrated via Homologous Recombination (HR) in the genome of the L40 yeast strain (referred to as tSL40). This pair was kindly provided by Doctor Chin at the UK Medical Research Council Laboratory of Molecular Biology, world class leader in chemical and synthetic biology. Chin and coworkers reported the activity of the *Mb*AckRS/tRNA$_{CUA}^{Pyl}$ in the *S.cerevisiae* MAV203 strain showing the incorporation of five ncAAs, including AcK. Stable transformation and correct expression of the amber suppression system components were checked through the following procedures: i) any integrated construct was sequenced to verify that the recombination has taken place ii) a specific PCR analysis and Northern Blot to detect tRNA$_{CUA}^{Pyl}$ was performed iii) the level of synthetase expression was evaluated by WB analysis. Thereafter, the amber suppressor activity and potential orthogonality of the *Mb*AckRS/tRNA$_{CUA}^{Pyl}$ pair was investigated in the PISA tSL40 strain. Robust preliminary results were obtained: WB analysis demonstrated the presence of the full-length protein only in presence of AcK; Additional WB analysis using an anti-H3K9Ac antibody will be addressed in future work. Mass Spectrometry-time-of-flight confirmed the acetylation at the genetically

encoded site. Nevertheless, data also showed very low readthrough protein yields, presumably due to a suboptimal catalytic activity of the amber suppression pair in L40 strain.
Still, this result suggested that it would be possible to implement the system, provided a higher read-through protein yield. As a matter of fact, the efficiency (ratio of full-length protein to truncated protein) and yield (total amount of full-length protein) of ncAA expression systems are a function of several factors, including the expression strain, the sequence context of the amber codon, the expression and activity of the synthetase and tRNA[183]. A variety of strategies have been reported to improve the efficiency, yield and specificity of ncAA incorporation, including optimization of orthogonal synthetase and tRNA expression level [114,188], and modifications to release factors (RFs)[16] [114,194]. Accordingly, I focused my effort on the systematic application of all the state-of-the-art technological improvement to boost the orthogonal translation system in yeast. First, I decided to minimize competition with eRF1, increasing the level of the suppressor tRNA (increasing the copy number of orthogonal suppressor o-tRNA). To begin with, I tried to double the o-tRNA gene cassette and the next step would be the creation of $tRNA_{CUA}^{Pyl}$ repeats to be integrated in the genome. I decided to proceed gradually because, even though creating repeats of orthogonal tRNA genes improves incorporation efficiency in some cases, it may also lead to unstable systems that lose tRNA genes via recombination. The strategy is still questioned and needs further testing. Moreover, in yeast it was reported that the increased activity was not a result of a higher tRNA transcription level, suggesting the importance of proper tRNA processing instead[116,192,193].

The other strategy, I adopted to improve the efficiency, was to create a Non Sense Mediated RNA Decay (NMD) deficient tSL40 strain (referred to as tSL40ΔUPF1) by knocking out the UPF1 gene (an essential component for the initiation of the NMD pathway) from the yeast genome. When stop codons are used to encode unnatural amino acids, NMD could result in a shorter lifetime for the target mRNA and thus a lower protein yield in yeast. However, I found that the tSL40ΔUPF1 strain did not increase the unnatural amino acid incorporation efficiency in comparison to the wild-type (WT) strain. This observation was in apparent discrepancy with previous reported findings obtained in a NMD deficient yeast by Wang and co-workers using the DanAlaRS/$tRNA_{CUA}^{Leu}$[199], indicating that the knot might lie in the activity of the *Mb*AcKRS synthetase variant itself. Nevertheless, disabling the NMD pathway in tSL40 added another building block toward the establishment of the PISA 2.0 host. The engineered tSL40ΔUPF1 strain is bound to facilitate the incorporation of ncAA in a manner that correlates to the UAG position. Provided a superior aaRS/tRNA couple, this strain would make it possible

[16] In eukaryotic cells, the release factor (eRF1) decodes all three stop codons and competes with ncAA incorporation systems at the amber codon

to reach a translational activity comparable to the one reported in *E.coli,* especially while using protein baits with the nonsense codon closer to the 5' end or 3' end of their mRNA.

Accordingly, I attempted to evolve a superior AcKRS[17]: a synthetase mutant library was amplified by using random mutagenesis, followed by a life-death assay selection. I reason that the random mutagenesis approach was likely to introduce mutations at positions outside of the aminoacylation active site, giving the possibility to explore a novel aaRS diversification strategy compared to commonly used ones [157]. Thanks to pilot experiments done by the lab[265], I decided to use a high rate mutagenic AckRS library in which approximately 15 amino acid mutations per ORF were expected (taking into account the expected number of mutations under similar conditions for a random mutation in a DNA stretch of similar size)[257]. A deep sequencing of the library capable of covering the full sequences of genes encoding aaRSs was beyond the scope of this work but it will be the objects of future study. The library was then ligated in the pRSII422 episomal vector containing the cognate $tRNA_{CUA}^{Pyl}$. As a first attempt, I co-transformed the mutagenized synthetase library with the pMICBD1 reporter plasmid carrying an amber stop in the Tryptophan (Trp) gene (essential for growth) in the yeast L40, so that it should be possible to link the amber suppression by the AcKRS3/$tRNA_{CUA}^{Pyl}$ pair to Trp production and cells survival. Trp was particularly attractive for two main reasons: i) L40 was auxotroph for Tryptophane, ii) the low rate of reversion mutation [266]. The selection was performed onto tryptophan deprived medium supplemented with AcK. Only a dozen revertant clones could be recovered probably due to the hard selection. It comes across that this kind of phenotype-based selection is sensitive to many factors and further improvements are needed. In particular, the development of an efficient selection method is a crucial step and the selection has to be tailored before repeating the experiment. In my design, I installed the TAG in the reporter plasmid Trp gene exploiting the autotrophy of my strain but I was not sure that codon position was a permissive site for amber reassignment. To ameliorate the selection approach, I could take advantage of previously used reporter or even better of transcriptional activated protein. I envision that placing the amber codon in a transcriptional activator coding sequence would link the suppression by the pair to the transcriptional activation. In this embodiment, by the choice of appropriate activated genes it would be possible to perform both positive and negative selection.

In conclusion, positive outcomes are expected merging the excellent plasticity and "evolvability" of this archea aaRs enzyme with careful adaptations of selection. Overall, this strategy was the most straightforward to implement in our laboratory, nevertheless more

[17] The o-aaRS derivatives in fact often showed suboptimal activity compared to the WT counterpart.

broadly, I consider that there remain significant questions regarding which selection or screening approaches are most effective for discovering and evolving aaRS in yeast.

3.3.3 Beyond acetylation: in yeast thorough comparison of off-the-shelves aaRS/tRNA pairs

Since all the collected results in the "acetylator" tSL40 strongly indicated a suboptimal catalytic activity of the *Mb*AcKRS in this strain, I decided to address this issue by changing the synthetase variant. The translational activity of recently evolved synthetase variants reported in literature were analyzed in order to compare them under the same experimental conditions and choose the actual working one in our strain.
First, the *M.mazei* AcKRS1/tRNA$_{CUA}^{Pyl}$ for encoding AcK was chosen to investigate if it would outperform the *M.barkeri* one used until now. Alongside the AcKRS pairs, I also tested the newly discovered single domain PylRS/tRNA$_{CUA}^{Pyl}$ pairs from Methanomassiliicocales that were shown to be very active and orthogonal in *Escherichia coli* [175]. These pairs lack the N-terminal domain, so they are smaller in size and more soluble than the AcKRS one I employed but share a homologous active site with it. Thus, if those N-terminus lacking enzymes had been found to display a significant activity also in yeast, I would have tried to evolve them to recognize the ncAAs of our interest (namely AcK), using selection methods previously used [116,143]. Despite the high translational activity of those PylRS in *E. coli* and mammalian systems, their activity has not been evaluated previously in yeast to the best of my knowledge. Additionally, to further explore the potential of the genetic code expansion technology in our yeast strains, I also analyzed the phosphoserynil-aaRS/tRNA pair (*Mmp*SepRS/tRNA$_{CUA}^{Sep}$) activity: it goes without saying that since serine phosphorilation is one of the most abundant PTM found in nature, developing an efficient PISA 2.0 system using phosphoserine is of great general interest.

All those amber suppression pairs were cloned in the same recipient vectors and compared under the same experimental condition using both i) a gold standard phenotypic assay and ii) a high-sensitive fluorescent assay. In this way all the potential confounding factors were level out allowing us to evaluate the translational activity solely.
First, I sought to replicate the phenotypic yeast assay performed by Chin and coworkers to validate the *Mb*AcKRS /tRNA$_{CUA}^{Pyl}$ variant in the same *Sc* MaV203 strain [120,149].
The method is based on activation of GAL4-responsive HIS3, URA3, or lacZ reporter genes by suppression of amber codons in GAL4 transcriptional activator protein; it was initially established to isolate new aminoacyl-tRNA synthetases in *Saccharomyces cerevisiae*.

Briefly, when a functional amber suppressing pair, such as the *Ec*TyrRS/tRNA$_{CUA}^{Tyr}$ pair, is transformed into this strain, full length GAL4 is produced, leading to activation of HIS3 genes. In my setting, the cells expressing either the AcKRS variants (the *Mb*AcKRS3 included) or the SepRS pair were almost unable to grow onto selective supplemented with the ncAAs, suggesting that these constructs were not fully functional. A better growth was observed for N-terminus lacking pairs, since individual clones were present at lower dilutions. Nevertheless, this growth phenotype was still too poor if compared to the *Ec*TyrRS/tRNA$_{CUA}^{Tyr}$ mediated amber suppression (the positive control).
In light of these results, I chose to perform a more sensitive assay: the GFP read-through assay. This highly sensitive fluorescence assay ensures a higher sensibility compared to other measurements. It is based on a GFP reporter with a TAG codon at position site known to be permissive for diverse ncAAs. And the ncAAs incorporation is verified by measuring fluorescence directly in intact cells expressing the Orthogonal Translation Systems under investigation. Even using the best performing and codon optimized green fluorescent protein from soft coral, none of the pairs produced a detectable level of GFP when used to suppress the TAG codon in the GFP gene.
While my observations resulted in apparent discrepancies with previously reported findings related to *Mb*AcKRS activities in yeast [149], I noted several differences between my experiments and the experimental conditions used in prior reports. In particular, changes in reporter system plasmid architecture, and orthogonal translation system plasmid architecture provide numerous potential explanations for the different experimental outcomes. Moreover, I would like to point out that the read-through level obtained was interpreted as not adequate only for the purpose of this project: the stable and robust expression of a PT-modified bait to be challenged with an antibody domain library. Furthermore, recent data in literature provide an independent validation of such a poor amber suppression phenotype in *S.cerevisiae* [162,193].

3.3.4 Reprogramming the yeast genetic code: Limitations

In this study some limitations in the GCE application in *S.cerevisiae* L40 strain, at least for the aaRS/tRNA pairs studied, have been highlighted. In fact, although all the amber suppressing pairs tested were well expressed and the intracellular concentrations of the ncAAs were in the working range, I was not able to observe any adequate protein read-through level, even using the highly sensitive fluorescence assay.

A similar conclusion was reached considering the weak growth phenotype obtained using the MaV203 selection strain by Chin et al., which enabled it to couple the amber suppression by the pairs to cell survival. These results clearly indicated that the engineered aaRSs have a suboptimal translational activity in our strain.

Additionally, the tRNA Extension experiments unequivocally indicated that only a minor and not detectable fraction of the o-tRNAs under investigation was aminoacylated in yeast cells. Northworthy, this is the first report using tREX technology to assess the in vivo aminoacylation status of o-tRNA in yeast models.

Most importantly, the tREX experiment observations highlighted that the bottleneck of the orthogonal translation systems in *S.cerevisiae* was upstream of the aminoacylation step. This phenomenon could be due to several reasons. As an example, while being correctly expressed, the synthetases may be poorly folded due to different physical parameters required for the yeast growth (e.g.: growth temperature etc.). Besides, I tested structurally different pairs: the SepRS has a quaternary structure, and it is bigger in size compared to the small and highly soluble *M. alvus* aaRSs lacking the N-terminal domain. Additionally, non-canonical localization sequences in the archea aaRSs may also cause their segregation from the cytosol to one of the eukaryotic cell compartments, preventing them from fulfilling their role in translation. For the sake of completeness, I made use of the DeepLoc prediction tool (a protein subcellular localization sequence prediction tool that used deep learning) to predict the presence of subcellular localization sequences (LSs) in the synthetase enzymes. I was not able to find any canonical LSs. However, many known LSs do not match any of the consensus sequences described.

On the other hand, the tRNAs may lack important modifications, preventing them from undergoing the correct maturation process or from being aminoacylated. Indeed, tRNAs undergo extensive modifications and the order of tRNAs processing events is not conserved [273]. As the tREX experiment suggested, I strongly believe that this is one of the most critical points to be analyzed. Further investigation of these phenomena was however outside the scope of my current research, which was focused on improving the PISA technology.

Taking everything into account, these findings strongly indicated that (unlike in *E. coli* and higher order eukaryotes), the use of these Orthogonal Translation Systems (OTSs) in yeast appeared to require higher ncAA concentrations while yielding low to moderate levels of ncAA [162]. And this inefficiency might be ascribed to many different mechanisms at the same time. Also, recent data in literature provided an independent validation of a poor amber suppression phenotype in *S.cerevisiae*[162,193].

When I started this project only few reports evaluating the PylRS activity in yeast were published. Specifically, Yokoyama and coworkers utilized two mutant Methanosarcina PylRSs to encode Boc-L-lysine and Nε-benzoyl-L-lysine, in S. cerevisiae MaV203 using 1 mM ncAA

concentrations during induction of protein synthesis, with low but detectable incorporation[191]. However, later, Chin et al. were able to demonstrate incorporation of five ncAAs with the engineered variant of the same PylRS (the AcKRS3 I used from the beginning) using concentrations during induction of protein synthesis ranging from 1.3–10 mM[149]. This evidence motivated me to explore this PylRS pair in the L40 platform used for the PISA selection.
Overall, our finding on the activity of a few specific synthetases, while relevant to our PISA 2.0 project, may not be applicable to other synthetases or yeast strains. Furthermore, these results are essential to draw a complete picture of GCE application and to achieve a better understanding of how would yeast, or more generally a eukaryotic host, respond to perturbations by these heterologous systems.

3.3.5 What happens to Legitimate UAG codon? Host cells response to amber suppression

Although UAG is the least-used stop codon, it still terminates ~7% of the total genes in *E.coli*[274] and ~20% in yeast and mammalian cells[142]. A long-standing question is how endogenous genes ending with the UAG are affected by code expansion. I therefore perceive it essential to report some scientific evidence on the subject and some considerations based on the data collected in this Thesis.

The amber suppression strategy indiscriminately increases suppression of all amber codons in the cell and therefore enhances the read-through of stop codons on chromosomal genes potentially disturbing cellular physiology[113,121,275]. However most laboratory strains of *E.coli* and yeast resulted to be tolerant of UAG readthrough[275]. Accordingly, I did not observe any undesirable effect or growth inhibition phenotype in the PISA 2.0 host strains (*E.coli* BIIM cells and *S.cerevisiae* L40).

Accordingly to available literature, after examining the expression of endogenous UAG-ending E. coli genes, Wang and coworker found that the amber suppressor Tyr-aaRS/$tRNA_{CUA}^{Tyr}$ does not efficiently incorporate its cognate Tyr at the TAG sites in the presence of RF1 [276]. However, upon RF1 knockout in *E. coli*, the UAG codon of endogenous genes was then efficiently suppressed by the orthogonal synthetase and tRNA pair, and such suppression led to a slower growth phenotype. Thus, it can be stated that amino acid incorporation at endogenous UAG codons is dependent on RF1, which can explain why *E. coli* tolerates

apparent global suppression of UAG [142,276] . To the best of my knowledge, no similar studies were instead performed in yeast.

Concerning this Thesis project, I did not observe any undesirable effect or growth inhibition phenotype in the PISA 2.0 host strains (*E.coli* BIIM cells and *S.cerevisiae* L40). All the tRNA/synthetase variants used in this thesis were orthogonal to the host organism (evolved from a two-step positive/negative selection scheme). They did not lead to observable readthrough of the nonsense amber codon in bait protein containing the TAG codon in the absence of ncAAs, suggesting that these aaRS/tRNA pairs do not efficiently suppress natural TAG stop codons in the absence of unnatural amino acids. Furthermore, the amber suppression by the pair was induced adding the ncAAs only for the time scale of a selection (around 7 days on plate), this may also explain why both *E.coli* and yeast can tolerate the system without notable adverse effect. Of course, a safer solution would be furtherly control the UAG read-through by a conditional production of the aaRS/tRNA$_{CUA}$ pair (especially in mammalian systems where occurrence of the TAG codon is higher compared to *E.coli*) [170,197].

Generally speaking, I considered that an important aspect to be addressed is the effect of long-term amber suppression on cells. For instance, I monitored the aaRS/tRNA transcription and expression over time, but it could be possible that the host cell explores ways to adapt to amber suppression in the long run. Moreover, Wang et al. [142] grew *E.coli* cells harboring one aaRS/tRNA$_{CUA}$ pair on plate and continued to pass cells several times. After ~500 generations, proteomic change of cells was characterized quantitatively. Over 30 proteins were downregulated, and 21 proteins up regulated. The most remarkable change identified was a hypothetical protein, YdiI, showing 16-fold increase. Intriguingly, this YdiI protein has an unexpected function of expelling plasmids from *E. coli*, helping cells to eliminate amber suppression pressure [148].

To conclude, the evidence collected in this Thesis work (and in other laboratories working on GCE) suggested that UAG read-through did not cause significant impairments; whether long term effect would cause problems awaits further study.

3.4 The PISA 2.0 Project in yeast: Future Perspective

3.4.1 Reprogramming the yeast genetic code: Future perspectives

In contrast to *E. coli*, efforts to engineer the OTSs in *S. cerevisiae* to date have used only a narrow range of approaches, despite this yeast's well established role as both a model biological organism and as a chassis for protein engineering[277]. As extensively described, archaeal PylRSs have proved invaluable to advance applications that use *E. coli* for protein expression [183,199]. On the contrary, PylRS variants were relatively underdeveloped in *S.cerevisiae*, suggesting that they may need to be evolved in yeast itself in order to improve their activities. Auspiciously, *the engineered Y2H strains generated during this work, will represent the ideal and robust yeast platform for engineering these enzymes*.

In addition, a substantial number of aaRS/tRNA pairs that exhibit translational activity in *E. coli* have been discovered in the last few years[113]. Evaluating the performance of these aaRSs in yeast may lead to the identification of OTS that exhibit the proper levels of translational activity in this model organism. For instance, recent work described a computational and experimental pipeline for the discovery of new pairs in *E.coli*[214]. This computational pipeline took advantage of prior work mapping identity elements, which are specific nucleotides within the tRNA sequence that mediate the specificity of cognate aaRS/tRNA recognition in *E. coli* [151,278]. The sequences of millions of tRNA genes, from genomic data, were scored for the absence of *E. coli* identity elements, using a metric that was benchmarked on known orthogonal tRNAs; the resulting candidate orthogonal tRNAs and their cognate aaRSs were tested for orthogonality in *E. coli* **[Figure 3.26].** Using this approach, several new orthogonal synthetases and tRNAs were discovered. Nevertheless, I reasoned that this approach might be easily applied to eukaryotic and so yeast cells.

In addition to the advancing discovery of highly active OTSs, I believe there remain significant questions regarding which selection (or screening) strategies are most effective for evolving prior aaRS/tRNA pairs that support genetic code expansion in organisms other than *E. coli*. For instance, to the best of my knowledge, all yeast-based campaigns to identify aaRSs have utilized the same set of positive and negative selections. These approaches have yielded a number of aaRSs, nonetheless they tend to lead to only moderately active variants that have unknown and uncontrolled specificity profiles[125,162,183]. On the contrary, in *E. coli*, careful adaptations of selection strategies using bioinformatics and other high throughput methods (such as phage-assisted continuous evolution and several flow cytometry-based screenings)

[157,279] have demonstrated to be very promising for the engineering of aaRSs. They remain not deeply explored in eukaryotic cells, though. To conclude, I reasoned that the newly discovered pairs combined with the availability of high-throughput and cutting-edge selection assays will promptly push the limits of our ability to engineer yeast. In this regard, the vector series to quickly integrate different pairs in the *S.cerevisiae* genome, the Y2H strains tailored to inactivate the NMD pathway and engineered to encode Phospho-serine, furnished in this thesis work, will provide an invaluable resource for this transition. Given the importance of yeast in basic biology and as a chassis for protein engineering and synthetic biology, a GCE larger chemical toolkit is expected to support advances ranging from a better understanding of eukaryotic biology to the identification of new therapeutic targets and discovery and validation of new therapeutic leads.

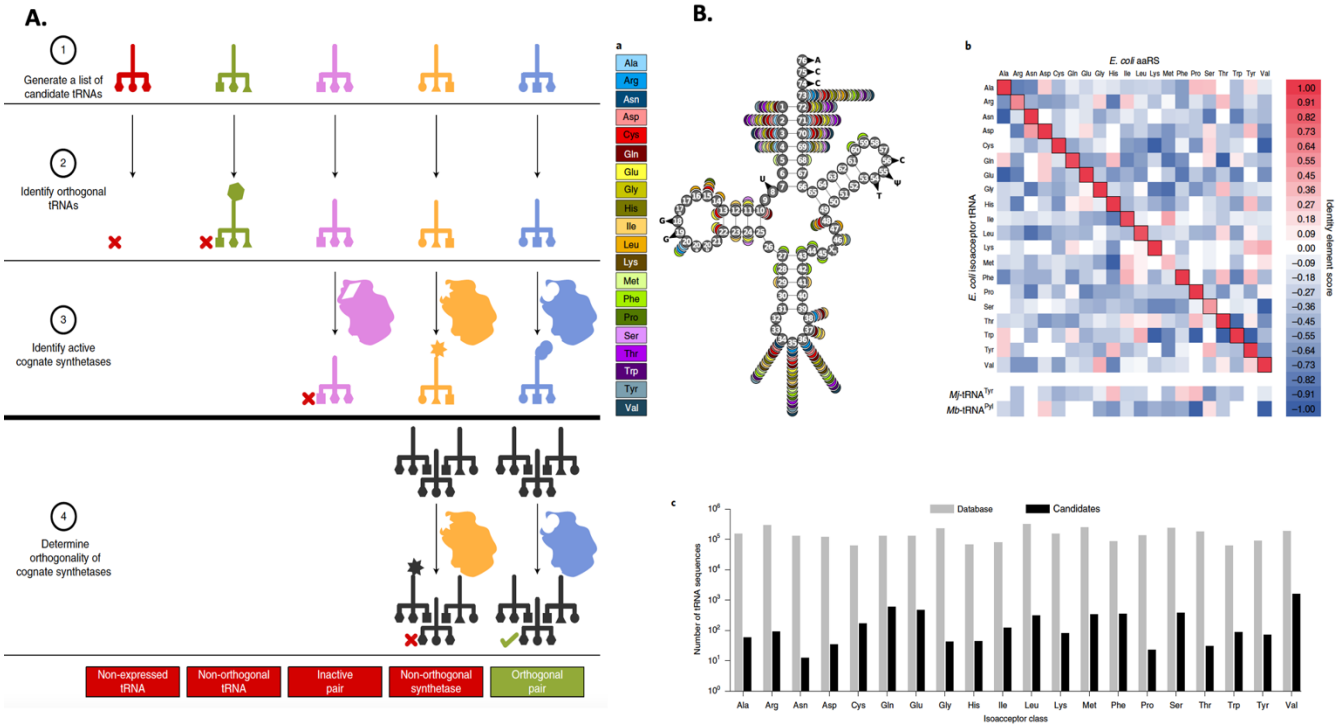


Figure 3.26. Pipeline for identifying orthogonal aaRS/tRNA pairs. The specificity of aaRS/ tRNA recognition is mediated, in part, by identity elements. Cervettini et al. scored the extent to which each class of E. coli isoacceptor tRNAs contain identity elements recognized by each E. coli synthetase by using a simple system. For each of the identity elements for a given synthetase, a tRNA received a score of +1 if its nucleotide matched the synthetase's substrate tRNA at that identity element and a score of −1 otherwise. The overall score for a tRNA with respect to each E. coli synthetase was the mean of the scores across all the identity elements for the synthetase; thus, all scores ranged between –1 and +1. This process generated 400 scores: the higher the score, the more similar a tRNA is to the endogenous substrates of the synthetase at the identity elements. **B) Developing and applying a computational filter to identify candidate orthogonal tRNAs. a,** Secondary structure diagram for tRNAs showing the canonical numbering scheme. Conserved nucleotide numbering is indicated. The identity elements for synthetases that recognize each amino acid are color coded as indicated to the left of the secondary structure diagram. **b,** Top, application of our scoring scheme to E. coli aaRSs and tRNAs. The score for each E. coli synthetase with each set of isoacceptor tRNAs is shown in the 20 Å~ 20 matrix. Bottom, the score for each E. coli synthetase with two commonly used orthogonal tRNAs**. c,** Filtering tRNAs from bacteria, archaea, mitochondria, chloroplasts and phage present in the tRNA-DB-CE database. The number of tRNA sequences for the indicated amino acid, as of March 2017, is represented by the gray bars. Removing tRNAs that scored more than 0.0 for the E. coli synthetase for the same amino acid reduced the number of candidates for experimental investigation by orders of magnitude (black bars). Adapted from Cervettini et al. 2021 Nature Biotechnology.

3.4.2 The proposed solution: the Bacterial-Two-Hybrid platform

The aim of the PISA 2.0 project is overcoming the limitation of the tethered catalysis method to accelerate and most importantly to automatize the isolation of antibody domains against a single specific PTM of a given protein. Initially I envisaged to stably integrate the o- aaRS/tRNA genes in the Y2H-PISA strain. The yeast cell would stably express the PT-modified protein target at a desired position residue. Having encountered several limitations in applying the GCE methodology in yeast, I decided to use the *E.coli* Bacterial-Two-Hybrid (B2H) as a host in lieu of the Y2H one to finalize the PISA 2.0 Project.

The novel B2H/GCE scheme will consist of a bacterial two-hybrid platform in which any orthogonal aminoacyl-tRNA synthetase/tRNA pair would be readily inserted to genetically encode PTM-UAA in the protein bait of interest. This would allow the bacterial-two hybrid system to select intrabodies against a specific PTM. I reasoned that the proposed scheme would remediate the shortcomings encountered installing the "UAA methodology"[115] in yeast.

As a matter of fact, most of the work on the expansion of genetic code to date has been conducted in *Escherichia coli* [113,115]. *Escherichia coli* was initially chosen as the host organism to develop the method for genetically encoding unnatural amino acids in vivo, because its translational machinery had been extensively studied and genetic manipulation was efficient and relatively straightforward[141,154]. Indeed, over the past two decades the several advances in GCE methodology have resulted in an impressive number of fully functional aaRSs/tRNAs in this model [214]. These pairs have been engineered to incorporate more than

200 uAAs in *E.coli* with a yield of several milligrams per L of cultures (the yield in yeast is only a few micrograms)[183].

On the other end, the B2H per se is also a vailable alternative to Y2H[280]. Both system work on similar principles and have succeeded in the identification of functional intrabodies, even if the yeast system has been more systematically and comprehensively exploited[35,80,91]. Different users may choose one over another depending on their needs (properties of the proteins to be screened, and ease of screening). The major advantages of bacteria vs. yeasts as host for two-hybrid-based antibody selections, comprise faster growth and higher transformation efficiency, allowing better coverage in library-based screenings. The introduction of the B2H in the lab would also make it possible for the selection of binders against homologous-eucaryotic proteins, or membrane associated proteins. The interactions detected by Y2H screens took place in the nucleus and thus proteins which have hydrophobic transmembrane region could not be able to reach the nucleus ultimately resulting in no interaction [281].

As the PISA 2.0 is concerned, I chose to use the B2H system by Muyldermans et al. [216] since it is based on transcriptional activation of two reporter genes as our Y2H, allowing the possibility to validate the interaction between bait and target with reduced background.

The PISA 2.0 Project in *E.coli* will be detailed in the (following) Chapter 4.

4 The PISA 2.0 Project in *E.coli*

Since the all results collected so far on Genetic Code Expansion in yeast highlighted poor non-canonical amino acid incorporation efficiency in *S.cerevisiae,* I decided to take advantage from i) the established bacterial-two-hybrid (B2H) system [215,216] as an intrabody selection platform and ii) the wide range of functional and orthogonal aaRS/tRNA pairs available in *E.coli* to set up the new 2.0 version of PISA in this model organism **(The PISA 2.0 Project *in E.coli*).** On one hand, *Escherichia coli* was initially chosen as the host organism to develop the method for genetically encoding unnatural amino acids in vivo and a plethora of fully functional tRNA/aaRSs are available in this model. On the other hand, the Bacterial Two Hybrid platformed is also a valuable alternative to Y2H: both systems have succeeded in the identification of functional intrabodies (even though the yeast two hybrid system has been the most widely used for the selection of intrabodies.).
Accordingly, during the last part of my thesis work, I accomplished the preliminary steps to adapt the PISA 2.0 project to bacteria. The project referred to as the PISA 2.0 Project in *E.coli* will be described in the following paragraphs.

4.1.1 The Bacterial-Two-Hybrid System used in this project

Bacterial two-hybrid platforms have also been used to select intrabodies [215,216], even though the yeast two hybrid system has been the most widely used so far. They are based on the principle of restoring protein activity upon non-covalent reconstitution of split protein fragments. There are multiple implementations of this technology, which use different biomolecular designs to achieve this behavior[98,99] . The bacterial-two- hybrid system developed by Dove, Joung, and Hochschild of Harvard Medical School and further refined by Joung and Pabo [96,97] of the Massachusetts Institute of Technology was adopted in this thesis, and it is therefore better detailed below.
This system detects protein-protein interactions based on transcriptional activation. A protein of interest (in our case the antibody domain, the prey) is fused to the full-length bacteriophage λ repressor protein (λcI, 237 amino acids), containing the amino-terminal DNA-binding domain and the carboxyl- terminal dimerization domain. The corresponding target protein (the bait) is fused to the N-terminal domain of the α-subunit of RNA polymerase (248 amino acids). The bait is tethered to the λ operator sequence upstream of the reporter promoter through the DNA-binding domain of λcI. When the bait and target interact, they recruit and stabilize the

binding of RNA polymerase at the promoter and activate the transcription of the *HIS3* reporter gene. A second reporter gene, *aadA*, encoding a protein that confers streptomycin resistance ($strp^R$), provides an additional mechanism to validate the bait and target interaction **[Figure 4.1].** When the reporter strain is co-transformed with hybrid bait and target proteins that interact, the RNA polymerase is recruited to the promoter, activating the transcription of *HIS3*. Growth of the reporter strain on media lacking histidine and containing 5 mM 3-AT occurs when transcriptional activation increases expression of the *HIS3* gene product to levels that are sufficient to overcome the competitive inhibition by 3-AT. This allows positive selection for plasmids encoding interacting proteins on media containing 5 mM 3-AT. Although both the B2H and Y2H work on similar principles, different users may choose one over another depending on their needs, available resources, properties of the proteins to be screened, and ease of screening.

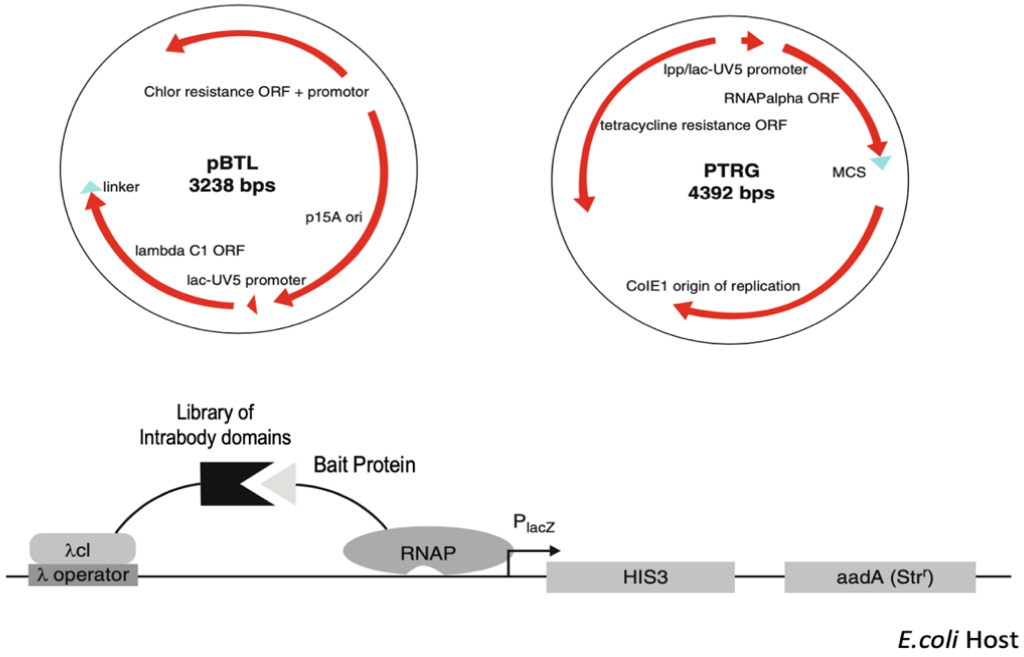


Figure 4.1. Schematic representation of the two plasmids used for the B2H selection screening (top) and the principle of the reporter gene activation (bottom). The pBTL vector at the left accommodates the VHH genes from camelid. The plasmid confers chloramphenicol resistance and the VHH is cloned in the multiple cloning site (MCS) at the 3′ end of the **λ**cI gene encoding a DNA binding domain recognizing the **λ** operator sequence. The pTRG plasmid (right) is used to express the protein target. The cDNA of the target antigen is cloned downstream the RNA polymerase (RNAPalpha) are indicated. Expression of the **λ** cI and RNAP alpha fusion genes are under the control of the lac UV5 promotor. Presence of both plasmids and expression of the fusion proteins inside the BMII reporter cells will lead to the binding of the **λ** cI DNA binding domain to the **λ** operator. The fused nanobody might interact with the target antigen fused to the RNAPalpha so that the reporter genes HIS3 and *aad*A are expressed (bottom), allowing the formation of bacterial colonies on plates in the presence of 3-AT and streptomycin. Adapted from Pellis et al 2012.

4.1.2 The novel Bacterial Two Hybrid & Genetic Code Expansion (B2H/GCE) Scheme

In the last part of my thesis project the first steps in order to adapt the Genetic Code Expansion method to the B2H platform will be accomplished:

- Construction and validation of the B2H bait and prey plasmids necessary to achieve the designed proof of principle;
- A trial B2H assay was assessed to test the system in our laboratory setting;
- The Expanded Genetic Code method was adapted to the B2H platform. Two orthogonal tRNA/synthetase pairs - one for the insertion of N-ε-Acetyl-L-Lysine [106] and the other for the insertion of O-Phospho-L-Serine [146] – were implemented in the Bacterial-Two-Hybrid system vectors.;
- Read-through Fluorescence assay was employed to test the functionality of the pair in the novel embodiment.

Suitably in the proposed B2H/GCE scheme, the intrabody selection will take place in the bacterial two- hybrid strain (in lieu of the Y2H) in which any orthogonal tRNA/aminoacyl-tRNA synthetase pair can be readily inserted to genetically encode PTMs in the protein bait of interest.

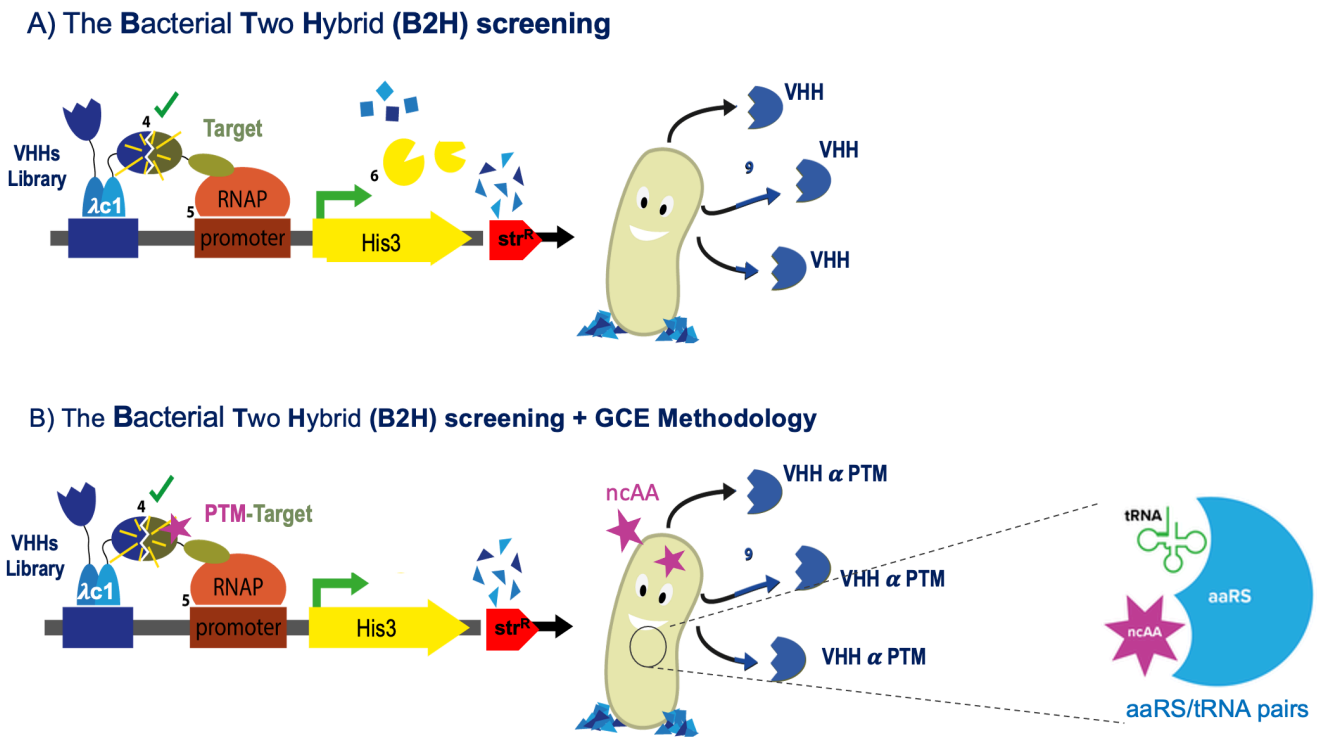


Figure 4.2. Illustration of the proposed Bacterial Two-Hybrid /Genetic Code Expansion system. A) The canonical B2H system. At first two fusion proteins were expressed. The first protein contains a DNA binding domain (***λ***cI) with one intrabody from our house-made VHHs library. The second fusion protein contains the activation domain (***α***-pol subunit) with the target protein. At first the DNA binding domain binds at the specific binding site upstream of the reporter. If the binding protein can interact with our target the RNA Polymerase I can be recruited and binds to the promoter. The result is the expression of the reporter gene, in example the His3 gene and the streptomycin resistance. No binding between the target and the binding protein leads to no expression of the reporter gene. At the end only the bacteria survive with the activated reporter. These bacteria you can use to produce a lot of the binding proteins against your specific target. Finally, tagged versions of the intrabodies can be directly purified. Interactions are thus confirmed also in vitro, by standard biochemical methods and in vivo. **B) The B2H-Genetic Code Expansion selection scheme.** In the proposed B2H/GCE scheme, the intrabody selection will take place in the same Bacterial Two-Hybrid platform in which, however, different orthogonal tRNA/aminoacyl-tRNA synthetase pairs are inserted to genetically encode PTM in the protein bait of interest. The amber suppression pairs are embedded in the bait plasmid without the need of engineering the bacterial strain and the TAG codon is used as unique codon in the bait construction. This would allow the bacterial-two hybrid system to select intrabodies against a specific PTM. As for the B2H, the putative anti- PTM target intrabodies can be purified directly from the bacterial cells. Interactions are thus confirmed also in vitro, by standard biochemical methods and in vivo.

4.2 The PISA 2.0 Project *in E.coli*: Material and Methods

In the following chapters the required steps for the realization of the PISA 2.0 project in *E.coli* will be detailed.

4.2.1 Plasmids and Constructs

In this section, the Bacterial- two-hybrid vector maps used in this this work are provided. Relevant features are shown including unique restriction sites that were useful in this project **[Figure 4.3].**

pTGR bait plasmid (Stratagene #200414). pTGR plasmid from [216] uses the lac UV5 promoter to express bait protein as fusions to the alpha the fragment of the RNA polymerase (RNAPalpha).pTRG is a low copy-number plasmid and present at about 20–30 copies per cell.

pBTL prey plasmid (Stratagene (#200414). pBTL from [216] uses the lac UV5 promoter to express the VHH as fusion to the ***λ***cI gene encoding a DNA binding domain recognizing the ***λ*** operator sequence. pBTL is present at 5–10 copies per cell.

pBTL-newMCS prey plasmid. The pBTL-new MCS prey plasmid was created starting from the original pBTL. This new version contains the *BssHII* and *NheI* restriction sites for cloning the VHHs library in frame with the λ-cI DNA-binding domain and was used for this specific thesis work. Briefly, in the original vector a 1374 bp *PstI–NheI* fragment of plasmid pBTL containis the MCS, the chloramphenicol resistance coding region. The *PstI-NheI* fragment was replaced by a fragment generated by PCR in which the *NheI* site was mutated into a SpeI and encoding a new MCS with the required *BssHII-NheI* sites with a spacer in between, followed by the chloramphenicol resistance coding region.

During purification of plasmid DNA, protocol modifications recommended by the purification kit manufacturer for low-copy-number plasmids were employed.

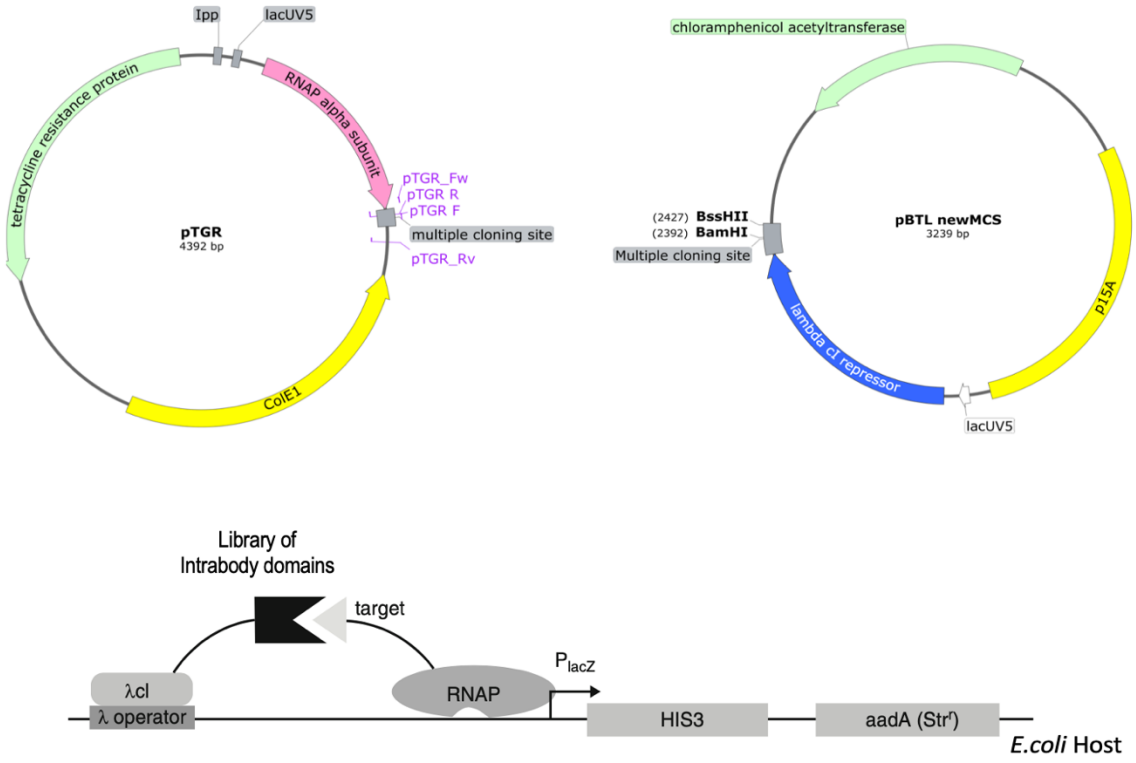


Figure 4.3. Maps of the two plasmids used for the B2H selection screening (top) and the principle of the reporter gene activation (bottom). (A) The pTRG plasmid (right) is taken from the Stratagene Bacterial Two Hybrid kit (#200414). The gene for tetracycline resistance, the colE1 origin of replication, and the gene for the fragment of the RNA polymerase (RNAPalpha) are indicated. The cDNA of the target antigen is cloned downstream the latter gene. **(B)**The pBTL new MCS vector at the right is an adapted pBT vector from Stratagene (#200414) containing a modified cloning site to accommodate the VHH genes from camelid. The plasmid confers chloramphenicol resistance and the VHH is cloned in the multiple cloning site (MCS) at the 3′ end of the ***λ*** cI gene encoding a DNA binding domain recognizing the ***λ*** operator sequence. Expressions of the ***λ*** cI and RNAP alpha fusion genes are under the control of the lac UV5 promotor. Presence of both plasmids and expression of the fusion proteins inside the BMII reporter cells will lead to the binding of the ***λ*** cI DNA binding domain to the ***λ*** operator. The fused nanobody might interact with the target antigen fused to the RNAPalpha so that the reporter genes HIS3 and aadA are expressed (bottom), allowing the formation of baterial colonies on plates in the presence of 3-AT and streptomycin.

Bait plasmid generation

pTGR-GFP-HA: the super folded GFP coding sequence was amplified using primer Forw 5′-ggatccgcggccgcaagaattcttGTTAGCAAAGGTGAAGAACTG-3′ and Rv restriction 5′-GTAATCTGGTACGTCGTATGGGTAGCTGCCTTTATACAGTTCATCCATACCG-3′ to add a HA tag at the C-terminus. The PCR resulting fragment was ligated into the MCS of pTGR using Gibson Assembly.The final vector was named pTGR-GFP- HA. This bait was chosen to perform the co-transformation assay.

Prey plasmid generation

pBTL-GFPNb-myc: the nanobody anti-GFP coding sequence was amplified from pCMV-pRAN-GFP26 (Addgene ##106406) using primer Forw with the *BssHII* restriction site 5′ ggatccgcggccgcaagaattcttGTTAGCAAAGGTGAAGAACTG 3′ and Rv with the *NheI* restriction site 5′-GTAATCTGGTACGTCGTATGGGTAGCTGCCTTTATACAGTTCATCCATACCG-3′. The PCR resulting fragment was digested using restriction enzymes *NheI*/*BssHII* and ligated into the MCS of pBTLnewMCS in frame with the λ-cI DNA-binding domain. The final vector was named pBTL-GFPNb-myc. The anti-GFP Nanobody is a synthetic nanobody isolated via phage display against GFP and its derivatives [282]. It was chosen to perform the B2H co-transformation assay since it has already been tested and employed in our lab.

4.2.2 Bacterial cells Media & Maintenance

Microbial growth Media

Luria Bertani (LB) rich media (Sigma MERK, L3522) was used for bacterial cells liquid growth coltures. For the B2H screening, the M9 minimal media was employed [216].

M9+ His-dropout Broth (500 ml)

380 ml of sterile, deionized H2O
50 ml of 10 x M9 salts
M9 Media Additives (67.5 ml, add sol II to sol I and mix well)
pH was adjusted to 7.0 by adding 10 N NaOH. Store at 4°C for up to one month

10X M9 salts (1 L):

74.76 g Na 2 HPO 4 .H 2 O;
30 g KH 2 PO 4 ;
5 5 g NaCl, 10 g NH 4 Cl; bring to 1 L in H 2 O and autoclave

Solution I:

10 ml of 20% glucose (filter-sterilized)
5 ml of 20 mM adenine HCl (filtersterilized)
50 ml of 10x His dropout amino acid
supplement (BD/Clontech, Cat. #630415)
Note: Sterilize the 10 X His dropout supplement by autoclaving at 121°C for 15 minutes prior to addition to Solution I. Do not exceed 15 minutes.

Solution II:

0.5 ml of 1 M MgSO4
0.5 ml of 1 M Thiamine HCl
0.5 ml of 10 mM ZnSO4
0.5 ml of 100 mM CaCl2
0.5 ml of 50 mM IPTG
Note: Filter-sterilize the thiamine-HCl andthe IPTG prior to use. Sterilize the remaining components of Solution II using the method of choice.

To prepare solid media 15g/L of Bacto-Agar (OXOID LP0011T) was added to the media before autoclaving. After autoclaving at 121°C for 20', antibiotics were added when the medium reached 50°C.
In all the experiments, bacterial BIIM cell growth is performed either on rich LB medium using a 30°C incubation temperature or in M9 minimal medium using a 37°C incubation temperature.

Non-selective screening medium (NSM)

500 ml of non-selective medium are made by adding chloramphenicol and tetracycline (final concentrations of 25 and 15 μg/ml respectively) to M9 minimal medium. Chloramphenicol is prepared as a 50 mg/ml stock solution in 100% EtOH, while tetracycline is prepared as a 5 mg/ml stock solution in 70% EtOH.

Selective Screening Medium (SSM)

SSM plates are prepared in the same way as NSM plates by adding 3- AminoTriazole (3-AT) (Sigma, A8056) to a final concentration of 5mM. 3-AT is prepared as a 1 M stock solution in 100% DMSO and filter-sterilized.

Double Selective Screening Medium (DSSM)

DSSM plates are prepared in the same way as SSM plates by adding streptomycin to a final concentration of 12.5 μg/ml. Streptomycin is prepared as a stock solution of 50 mg/ml streptomycin in water.

4.2.3 Bacterial strains:

XL1-Blue MRF´Kan strain is the host strain for propagating pBTL and pTRG recombinants. The BacterioMatch II Validation Reporter Competent Cells (Catalog #200192) -hereafter referred as BIIM-are the reporter selection strain. Genotype: Δ*(mcrA)183* Δ*(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F´*proAB* $lacI^{q}$Z*ΔM15* Tn*5 (*Kan^{r}*)]*.

Preparation of electro-competent BMII cells

The preparation of electrocompetent cells (~ 35 50−µl aliquots) follows this protocol (all steps performed at 4°C):

1. BMII cells are streaked on LB plates from glycerol stock (Sigma cat #G5516) and incubated at 30°C overnight
2. Pick a colony and inoculate it in a small culture overnight
3. Dilute the overnight culture to OD600 = 0.3 in a 300 ml culture
4. When the culture reaches an OD600 ~ 0.5 − 0.6, put it on ice in the cold room for 1 hour
5. Centrifuge cells (6 ml Falcon tubes) for 7 minutes at 2200 x g
6. Resuspend the pellet with 10 ml 1 mM Hepes pH=7.0, then add Hepes until you reach the initial volume (total=300ml)
7. Centrifuge cells for 7 min at 2200 x g
8. Resuspend pellets in 10 ml 10%w/v glycerol, and then add glycerol until you reach half of the initial volume (total=150 ml)
9. Centrifuge cells for 7 min at 2100 x g
10. Resuspend a pellet in 10 ml of 10% w/v glycerol. Then move the solution to the next tube until all tubes have been harvested. Repeat with another 10 ml
11. Centrifuge for 6 min at 2100 x g, keep the pellet and transfer the supernatant to another tube.
12. Centrifuge the supernatant from the previous step for 7 min at 3500 x g. Discard the supernatant.
13. Resuspend the pellets from the last two steps in a total final volume of 2 ml. Divide the volume in 50µl aliquots and put them in dry ice.
14. Transfer the aliquots to the -80°C fridge

4.2.4 The B2H selection protocol

Selection of binders by bacterial-two-hybrid

Aliquots of 75 µl freshly prepared electrocompetent BMII cells carrying pTRG-bait plasmid were transformed with pBTL-newMCS prey plasmid. The cells were brought into 1 ml SOC medium and allowed to recover at 37 °C for 1–2 h before washing and transferring into $M9^+$ His-dropout minimal medium for 1–2 h at 37 °C, and plating on Single Selective Screening Medium (SSM) (i.e., $M9^+$ His-dropout minimal medium-agar plates supplemented with 2.5 mM 3-AT). To test the co-transformation efficiency, cells were plated on LB-agar plates supplemented with 25 µg/ml chloramphenicol, 12.5 µg/ml tetracycline and 70 µg/ml kanamycin. The NMS-agar plates were used to monitor the total number of cells. After 24–30 h incubation at 37 °C, colonies on the SSSM plates were counted and manually transferred to DSSM plates (i.e. $M9^+$ His-drop- out minimal medium-agar plates supplemented with 2.5 mM 3-AT) 24 h before counting the surviving colonies..

Characterization of positive clones: Isolation of plasmid DNA from bacterial colonies

Colonies that grow when streaked on DSM plates are checked for the presence of a VHH by colony PCR (primers pBTL-Forw and pBTL-newR, Tann = 55°C, tel = 1'). Positive clones (producing a band at around 500 bp) are then inoculated in LB-Cmp and grown overnight for protein extraction and in vitro assay.

4.2.5 In vitro assays

Western blot

Protein extraction for western blot was performed as follows: an appropriate amount of cell culture (ranging from 1 to 3 ml) either from an o/n or from an exponential growth (OD600 from 0.3 to 0.5) was pelleted and resuspended in Laemmli buffer 2x (Sigma, S3401), boiled at 95 °C for 5' and then loaded on in house made 8-10% acrylamide gels for SDS-PAGE. For Western blots detection the following antibodies were used:

- antiHA-HRP, (Roche, 1:1000)
- antiMYC-HRP, (Cell Signalling, 1:1000)

For imaging, a Chemidoc XRS (Biorad) was used.

Co-Immunoprecipitation Assay

Protein extraction for Co-Ip was performed as follows: an appropriate amount of cell culture either from an o/n or from an exponential growth (OD600 from 0.3 to 0.5) was pelleted and resuspended in TrisHCl-buffer (pH 7.0) (4 volumes buffer for 1 volume cells). The solution was sonicated 60 times 30 seconds at 20% power while keeping the sample in ice water to prevent heating. The protein suspension was centriguged for 15 minutes at 1400 x g 4°C and resuspended in TBS 1x and rotease inhibitors (Roche cOmplete mini) + PMSF 1x 1X (buffer).
A total of 0,5 ug of extracts were used for each immunoprecipitation (IP). The IP was performed adding to each sample 1 ug of specific primary antibody in rotation at 4 °C O/N. Then, 35 uL of Protein G beads were added and samples were put to rotate at 4 C. Beads were recovered, sequentially washed in, and TBS 1x + tween 0.05%. Finally, the beads were resuspended in 30 uL of 2X laemmli buffer, boiled for 10 min, and the eluates were analyzed through % SDS-PAGE.

4.2.6 Protocol for ncAAs incorporation in Bacterial cells:

Plasmid and Construct:

p15A-sfGFP150* reporter plasmid. The reporter plasmid p15A GFP contained the p15A origin or replication and a constitutively expressed tetracycline resistance gene in addition to the fluorescent reporter super folded (sf) Green Fluorescent Protein (GFP) gene containing an amber stop codon at position 150 under the control of the l-arabinose-inducible PBAD promoter. This reporter plasmid was used to assess the functionality and orthogonality of the aaRS/tRNA couples under investigation **[Figure 4.4 A].**

orthogonal translation system expression plasmid pRST. The pRST expression plasmid was used to constitutively express the tRNA–synthetase pairs. It consists of the pMB1 origin of replication (High copy number) the constitutive *E. coli glnS* promoter controlling expression of the aminoacyl-tRNA synthetase of interest followed by a spacer, the strong constitutive *E. coli lpp* promoter controlling expression of the tRNA of interest; and a *rrnC* terminator following the tRNA. Specifically the pRST vectors bearing the WT PylRS/tRNA$_{CUA}^{Pyl}$ from either *M.barkeri (Mb)*, *M. mazei(Mm)* or *M.alvus (Ma)* and the *Mb*AcKRS/tRNA$_{CUA}^{Pyl}$ were gifted by J.Chin Lab (MRC-LMB Cambridge, UK) **[Figure 4.4 B].**

All plasmids containing the aaRS derivatives were constructed starting from the plasmid containing the respective wt PylRS and tRNA$_{CUA}^{Pyl}$ as follows:

pRST-*Mm*AcKRS1/tRNA$_{CUA}^{Pyl}$. The *Mm*PylRS that aminoacylate *Mm*tRNA$^{Pyl}_{CUA}$ with *N* -acetyl-L-lysine (AcKRS1) were created by transferring mutations (L301M/Y306L/L309A/C348F) identified in Umehara 2012 [172] into the *Mm*PylRS gene. The *Mm*PylRS gene was amplified using two pairs of site-directed mutagenesis primers. The resulting PCR fragments were cloned into the pRST-*Mm*PylRS/tRNA$^{Pyl}_{CUA}$ beckbone in a two-step Gibson.

pRST-*Ma*AcKRS3/tRNA$_{CUA}^{Pyl}$. The *Ma*AcKRS3 derivatives were created by transferring mutation (L266M/L270I/Y271F/L274A/C313F) identified in *MbAcKRS* into the *Ma*PylRS gene. The *Ma*PylRS gene was amplified using two pairs of site-directed mutagenesis primers. The resulting PCR fragments were cloned into the pRST-*Ma*PylRS/tRNA$_{CUA}^{Pyl}$ vector in a two-step Gibson.

pRST-*Ma*AcKRS1/tRNA$_{CUA}^{Pyl}$. The *Ma*AcKRS1 were created by transferring mutations (L301M/Y306L/L309A/C348F) identified in Umehara 2012[172] into the *Ma*PylRS gene The *Ma*PylRS gene was amplified using two pair of site-directed mutagenesis primers containing the mutations. The resulting PCR fragments were cloned into the pRST-*Ma*PylRS/tRNA$^{Pyl}_{CUA}$ backbone in a two-step Gibson.

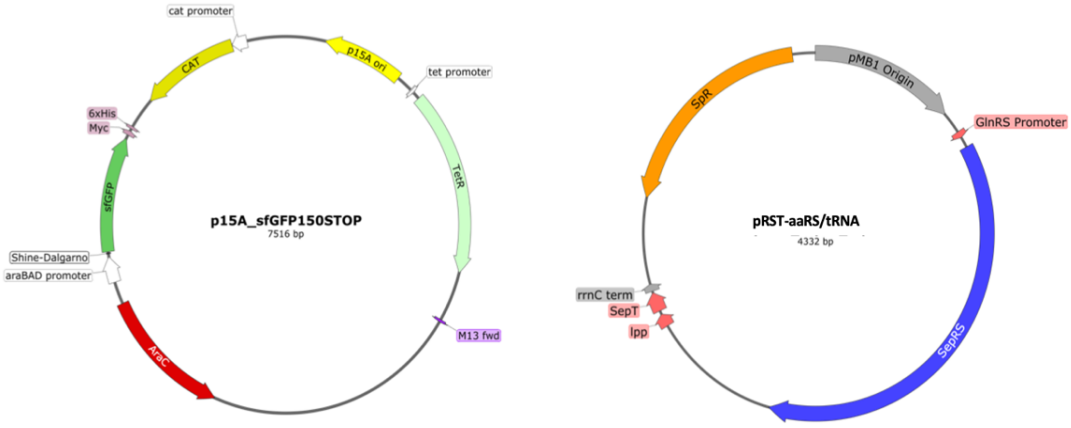


Figure 4.4. Maps of the orthogonal translation system expression plasmids adopted in this project (gifted by Dr. Jason Chin MRC-LMB, Cambridge). **(A) The p15A-sfGFP150* plasmid** (on the left) is the reporter plasmid used for the sfGFP read-through assay. It contained the p15A origin of replication and the fluorescence reporter GFP gene containaning an amber codon at position 150 under the control of PBAD promoter (B)**The pRST plasmid** (on the right) was used to constitutively express the aaRS/tRNA pairs. It consists of the pMB1 origin of replication (High copy number) the constitutive E. coli glnS promoter controlling expression of the aminoacyl-tRNA synthetase of interest followed by a spacer, the strong constitutive E. coli lpp promoter controlling expression of the tRNA of interest; and a rrnC terminator following the tRNA.

The $tRNA_{CUA}^{Pyl}$ optimized (opt). The $tRNA_{CUA}^{Pyl\text{-}opt}$ variant from [283] was created by inserting mutation U49-A65/G50-C64 in the $tRNA_{CUA}^{Pyl}$ gene **[Figure 4.5].**

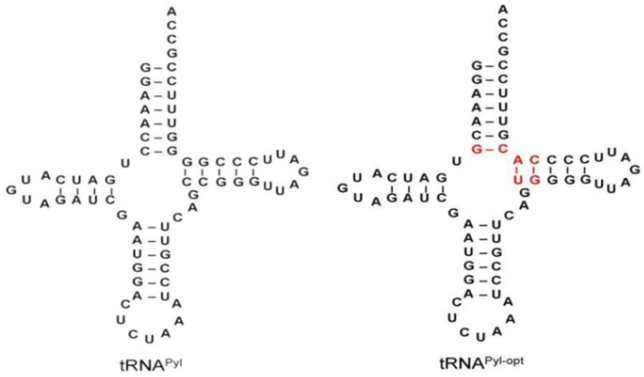


Figure 4.5. The secondary (cauliflower) structure of $tRNA_{CUA}^{Pyl}$ and $tRNA_{CUA}^{Pyl\text{-}opt}$. The based pairs different from the wild-type are marked with red color.

Table 4.1. List of the o-aaRS/tRNA components used in the PISA Project *in E.coli*

>*Mb*AcKRS3 MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKTCKR CRVSGEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVSAKASTNTSRSVPSPAKSTPNS SVPASAPAPSLTRSQLDRVEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDREDYLGKLERDITK FFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMMAPTIFNYARKLDRILPGPIKIFEVGP CYRKESDGKEHLEEFTMVNFFQMGSGCTRENLEALIKEFLDYLEIDFEIVGDSCMVYGDTLDIMHGDLELSS AVVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRSESYYNGISTNL*
>MmPylRS MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRHHKYRKTCKR CRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVST QESVSVPASVSTSISSISTGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFR ELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDK NFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLESIITDFL NHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDFKNIK RAARSESYYNGISTNL*
>MaPylRS MTVKYTDAQIQRLREYGNGTYEQKVFEDLASRDAAFSKEMSVASTDNEKKIKGMIANPSRHGLTQLMNDI ADALVAEGFIEVRTPIFISKDALARMTITEDKPLFKQVFWIDEKRALRPMLAPNLYSVMRDLRDHTDGPVKIF EMGSCFRKESHSGMHLEEFTMLNLVDMGPRGDATEVLKNYISVVMKAAGLPDYDLVQEESDVYKETIDVEI NGQEVCSAAVGPHYLDAAHDVHEPWSGAGFGLERLLTIREKYSTVKKGGASISYLNGAKIN*
>*Mb*tDNA$_{CUA}^{Pyl}$ ggaaacgtgatcatgtagatcgaatggactctaaatccgttcagtggggttagattccccacgtttccgcca
>*Mm*tDNA$_{CUA}^{Pyl}$ ggaaacgtgatcatgtagatcgaatggactctaaatccgttcagtggggttagattccccacgtttccgcca
>*Ma*tDNA$_{CUA}^{Pyl}$ gggggacgguccggcgaccagcgggucucuaaaaccuagccagcgggguucgacgccccggucucucgcca

4.2.7 Construction of pTGR-*Mm*AckRS1-/tRNA$_{CUA}^{Pyl}$ and pTGR-*Mm*pSepRS/tRNA$_{CUA}^{Sep}$ plasmids

The orthogonal tRNA/synthetase pairs to achieve either N-ε-Acetyl-L-Lysine or O-phospho-L-Serine insertion in the bait were cloned in the pTGR plasmid through Gibson assembly. pTGR plasmids containing only the cognate tRNAs were also constructed to be used as controls for tRNA orthogonality. Gibson assembly allowed cloning the genes of interest in the backbone region outside the multiple cloning site of pTGR, and at the same time, eliminating a stuffer portion of pTGR backbone to avoid inflating the final size of the plasmids.

pTGR-tRNA$_{CUA}^{Sep}$ and pTGR-tRNA$_{CUA}^{Pyl\text{-}opt}$. DNA fragments for both the tRNAs were amplified from pRST vectors received from Prof. Jason Chin's using primers Forw 5′

ctgctaaccagtaaggcaacccCAAAAATAAGTGGCGCCCCATC 3′ and Rv 5′ agccaatcaattcttgcggaCGTTGTAAAACGACGGCCAG 3′ and cloned into pTGR-bait plasmid using one-step Gibson.

pTGR-MmAcKRS1/tRNA$_{CUA}^{Pyl-opt}$ and pTGR-SepRS/tRNA$_{CUA}^{Sep}$. DNA fragments for either AcKRS3 and SepRS were amplified from pRST vectors using primers Forw 5′cactcagggtcaatgccagcgcTCGGGAGTTGTCAGCCTGTCCC3′ and Rv 5′GTTACAGGTCCTCCTCTGAGATCAGCTTCTGCTCCAGGTTCGTGCTAATGCCGTTATA G 3′. Primers were designed for cloning synthetase inserts with a MYC tag at their C-terminus. Amplicons were subsequently cloned in either pTGR-tRNA$_{CUA}^{Pyl}$ or pTGR-tRNA$_{CUA}^{Sep}$ respectively.

pTGR-sfGFP150*-HA-aaRS/tRNA. DNA fragments encoding the sfGFP reporter were amplified from p15A-sfGFP or p15A-sfGFP150* (namely, sfGFP harbouring a stop codon in position 150) using primers Forw 5′cggatccgcggccgcaagaattGTTAGCAAAGGTGAAGAACTG TTTACCG 3′and Rv 5′ GTAATCTGGTACGTCGTATGGGTAGCTGCCTTTATA CAGTTCATCCATACC 3′ to add a HA-tag at the C terminus and ligated into the MCS of either pTGR-aaRS/tRNA or pTGR-tRNA vectors via Gibson assembly.

Growth culture protocols for ncAAs incorporation

Accordingly with standard protocol [124]the appropriate selective medium supplemented with 2 mM of N-ε-Acetyl-Lysine (AcK in the following, Sigma, A4021) was inoculated with a stationary phase culture of the required bacterial strain in a round bottom tube. Parallel cultures with no ncAA were also prepared as negative controsl. Cultures were grown at 37°C in an orbital shaker for 24-48 h.

4.2.8 sfGFP read-through assay: GFP expression for fluorescence quantification.

The gene encoding sfGFP was in a plasmid with the p15A origin of replication under the control of the L-arabinose-inducible PBAD promoter. *E. coli* DH5alpha cells containing the gene encoding sfGFP with an in-frame amber stop codon at position 150 were transformed with a pRST plasmid by heat shock. Recovery medium was added (1 ml SOC medium), and the transformation was divided into three aliquots. After recovery (1 h, 37 °C, 850 r.p.m.), 50 µl of each transformation was added to a different well of a 24-well deep-well plate (Riplate SW 24, PP, 10 ml) containing 5 ml of 2xYT medium (Sigma Aldrich Y1003) supplemented with the appropriate antibiotics for selection, 0.2% arabinose to induce expression of GFP and 2 mM ncAA, as necessary. Cells were incubated at 37 °C (22 h, 220 r.p.m.), and 500 µl of medium was

transferred to clear-bottom 24-well plates. The GFP fluorescence in this standard volume was measured with the PHERAstar FS (BMG Labtech) plate reader using the optic module with an excitation wavelength of 485 nm and an emission wavelength of 520 nm. The gain was set to 8 (arbitrary units). Measurements were performed on each sample grown in individual wells. In each graph, individual data points are shown together with the SD.

4.2.9 Verify ncAA incorporation with WB analysis

To verify if N-ε-Acetyl-Lysine is correctly incorporated, BMII cells were transformed with the pTGR-*Mm*AcKRS-1/tRNA$_{CUA}^{Pyl}$ plasmid carrying a sfGFP with an amber codon in correspondence with amino acid 150, indicated as sfGFP150*. M9-His Drop-out medium supplemented with histidine (Sigma, H0750000) (0.2% w/v), N-ε-Acetyl-Lysine (2 mM), tetracycline (15 μg/ml), IPTG (100 mM) was inoculated with the stationary phase culture of the BIIM expressing sfGFP150* strain to yield an OD600 ≈ 0.2 in a 15mL round bottom tube. After 24h in an orbital shaker at 37 °C culture fluorescence was qualitatively assessed by illuminating the culture with UV light. Moreover, 1 ml of culture was pelleted, and used to prepare protein extract for western blot analysis. The sfGFP150* bait is HA-tagged at the C- terminus thus WB using the anti-HA antibody will detect the acetylated-protein (in the sample supplemented with AcK) at the expected molecular weight only if there has been a readthrough of the stop codon in the GFP protein.

4.3 The PISA 2.0 in *E.coli*: Results

In this Chapter, all the collected results obtained in order to adapt the Genetic Code Expansion method to the B2H platform will be described.

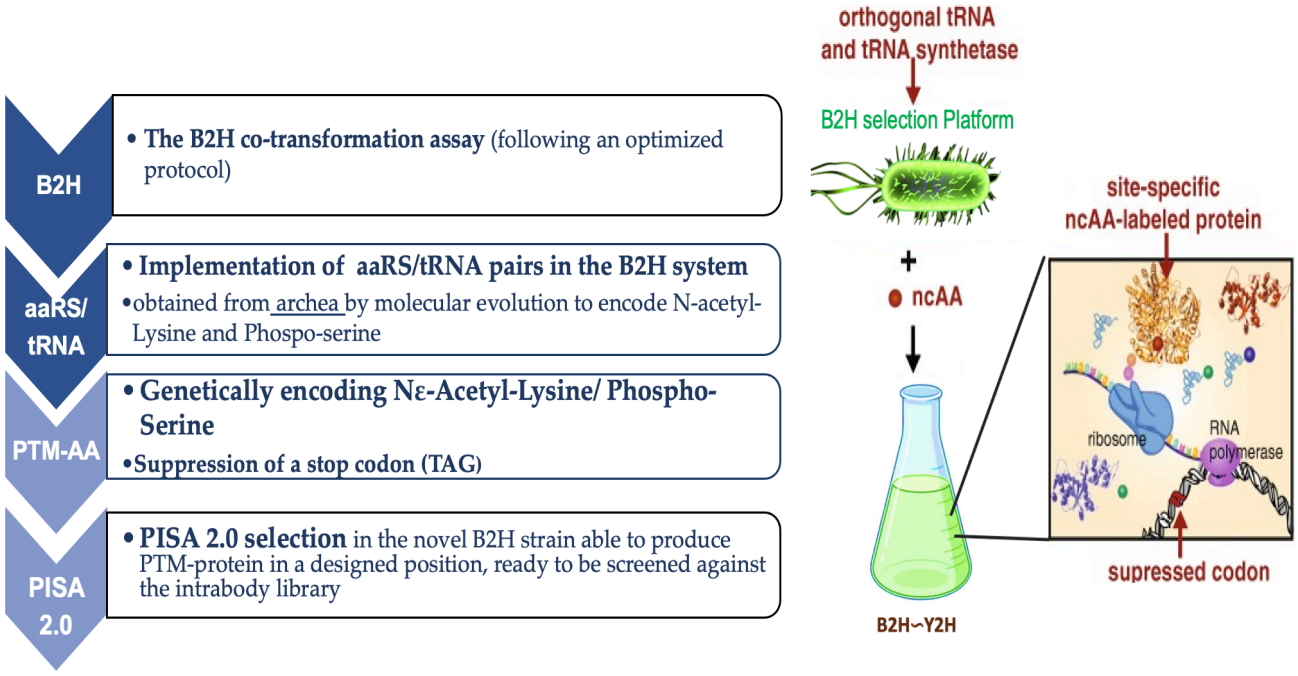


Figure 4.6. Schematic Pipeline of the PISA 2.0 Project in *E.coli*

4.3.1 The Bacterial-Two-Hybrid trial screening: *in vivo* co-transformation assay with the GFP/αGFP interactors

To set up the PISA 2.0 selection *in E.coli*, I chose to take advantage of an established B2H system **[Figure 4.7]** [216] [215]. Briefly **[Figure 4.7],** the VHH of interest is screened against a bait (target protein) in the B2H *E.coli* strain. If the intrabody (the prey) is able to interact with the bait, the HIS3-aad3 reporter cassette will be transcribed, allowing the survival of the cells on selective medium. Expression of the HIS3 reporter gene allows the cells to grow onto Single Selective Medium (SSM) plates in presence of 3-AT (3-amino-1,2,4-triazole), a competitive inhibitor of the His3 enzyme. Surviving cells are subsequently tested using the *aadA* gene as a secondary reporter, which confers streptomycin resistance onto Double Selective Medium (DSSM). Double positives are then further characterized by a secondary screening to confirm their true positivity. These intrabodies will be also used for in vitro and in cell assays to confirm interaction and functionality.

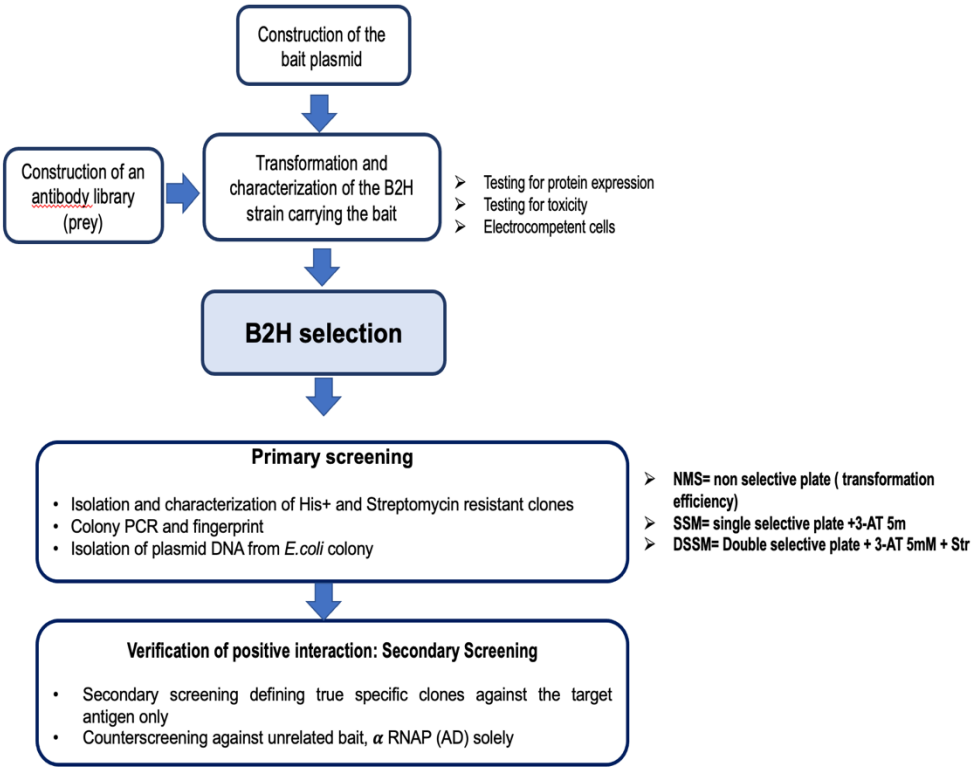


Figure 4.7 Schematic pipeline of the B2H selection protocol

Before proceeding with the application of the Expanded Genetic Code "UAA methodology" to the B2H, a B2H co-transformation assay was performed using a known bait/prey interactor couple [Figure 4.8]. This trial allowed us to set the conditions for intrabody selections using the B2H system. Specifically, we chose the sfGFP protein as bait and a synthetic nanobody anti GFP (αGFPNb) as prey. The latter was isolated to recognize GFP and its derivatives from [282]and it has been already tested for another research project in our laboratory.

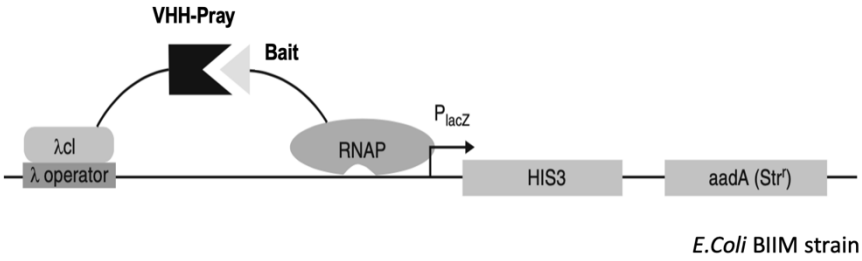


Figure 4.8. Schematic of Bacterial-Two-Hybrid selection against Arc protein. The DNA binding domain (***λ***cI) is fused with the VHH domain of interest. The activation domain the (***α***-pol subunit) is fused with the target protein. The expression of the fusion proteins inside the B2H reporter cells will lead to the binding of the ***λ*** cI DNA binding domain to the ***λ*** operator. The fused nanobody might interact with the bait fused to the RNAPalpha so that the reporter genes HIS3 and aadA (streptomycin resistance) are expressed (bottom), allowing the formation of baterial colonies on plates in the presence of 3-AT and streptomycin. Adapted from Pellis et al 2012.

Induction and stable expression of the B2H bait protein: the RNAP-sfGFP-HA

First, the correct bait expression and induction condition have to be assessed. The BMII cells strain transformed with the pTGR-sfGFP-HA plasmid were grown o/n in LB-Tetracycline (Tet). The next morning the culture was diluted to OD600 = 0.3 in LB-Tet and divided in 4 different aliquots that were treated with IPTG (0, 10, 100, 1000 μM) to induce the GFP-HA production. Western Blot of cells protein extracts revealed the correct expression of the bait protein with increasing concentrations of IPTG **[Figure 4.9].**

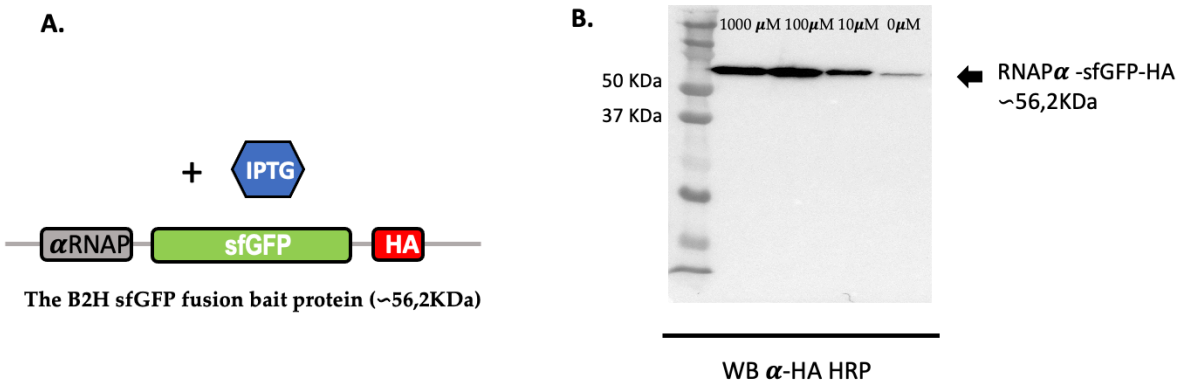


Figure 4.9. Induction of the B2H GFP bait. A) Schematic representation of the B2H sfGFP-HA bait. The target antigen is fused to C-term of the α-RNAP and at the N-term of an HA tag. **B)** WB analysis using anti-HA-HRP antibody on protein extracts of BMII cells transformed with pTGR-sfGFP-HA induced using increasing concentration of IPTG (0,10,100,1000 μM).

The B2H test co-transformation assay

Electrocompetent BMII cells containing plasmid pTGR- *sf*GFP-HA were transformed with either the pBTL-αGFPNb-myc prey or an unrelated one (an anti Microtubule Binding Domain of Tau protein (MTDB)) pBTL-αMTBDNb-myc. Furthermore, one electroporation was performed with the empty pBTL-newMCS (with no VHH) to control for possible toxic effects due to the expression of the VHH. From serial dilutions plated on LB-chloramphenicol (LB-cmp) the trasformation efficency on LB-chloramphenicol was estimated to be 4×10^8 CFU/μg. Some toxicity due to the αGFP was observed in the reduced (10^{-2}) number of colonies on Non-Selective Medium (NSM) (compared to LB-Cmp plates), as this reduction was not observed for the empty plasmid. The obtained efficiencies (in CFU/μg) are reported in the following **Table 4.2**.

Table 4.2 The Trial B2H selection: Estimated transformation efficiencies in CFU/μg.

DNA	Media	Efficiency (CFU/ug)
pBTL-GFPNb and pBTL	LB +Chloramphenicol	4×10^8
pBTL-GFPNb	NMS	4.2×10^6
Empty pBTL control vector	NMS	1.8×10^8

Legend.LB+Cmp= LB medium + Chloramphenicol; NSM= Non Selective Medium; SSM= Selective Screening Medium.

Following the protocol [216], cells were plated on single selective (SSM) plates and incubated at 37 °C for 48 hours. Colonies grown on SSM were considered putative interactors (CFU=3X10^4). As expected sfGFP colonies were able to grow onto selective plate only when transformed with αGFP Nb and not the unrelated prey. Random colonies were picked and streaked on double selective (DSSM) plates, to control for aspecific auxotrophy for histidine. After 48 hours at 37 °C all streaks had grown on DSSM**.**

Verification of positive interaction

Co-immunoprecipitation assay was chosen to detect and further confirm the bait/prey physical interaction from colonies grown on double selective medium.

Randomly picked colonies from DSSM plate were used to inoculate an overnight starter culture of 5 ml LB medium containing tetracycline and chloramphenicol and grown for 16-18 h at 30°C. The overnight culture was diluted to OD_{600} = 0.8 in 1 ml LB-tet-Cmp and added IPTG (1 mM final concentration). The cultures were incubated (30°C, 225 r.p.m) for 4 h and pelleted (14,000 r.p.m., 1 min) to perform protein extraction and Co-immunoprecipitation. The same procedure was followed for the BIIM cells expressing the GFP bait and the unrelated Nb as control (grown onto LB+ CMP+Tet plate). As expected, only the Nb αGFP pulled down RNAP-sfGFP-myc, as shown in the subsequent anti-myc blot (detecting the myc-tagged GFP) **[Figure 4.10A-B].**

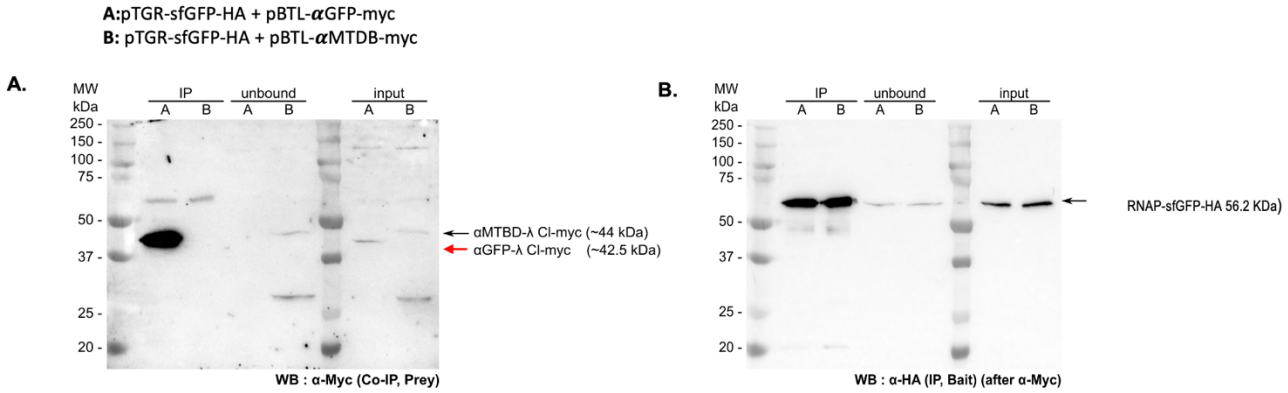


Figure 4.10. Verification of positive bait/prey interaction: Confirmatory Co-Ip assay. One randomly selected positive colony (A) from DSSM, expressing HA tagged RNAP-GFP and myc tagged nanobody (***α***GFP) and one colony (B) from LB +CMP+Tet, expressing HA tagged RNAP-GFP and myc tagged unrelated nanobody (***α***MTDB) were used in the experiment. Lysates were

prepared and subjected to immunoprecipitation (IP) with anti-HA antibody. Immunoprecipitated proteins were detected by western blot (WB) using the anti-myc antibody (left blot). Unbounds and Inputs were also shown as detailed in the picture. On the right the membrane blotted with anti-HA antibody.

In conclusion, the B2H screening assay was tested and optimized: it turned out to be suitable to correctly assess the interaction between a VHH domain and the protein target.

4.3.2 Implementation of the Genetic Code Expansion (GCE) technology in the B2H system

To assess the feasibility of the application of the GCE methodology to the B2Hplatform, I decided to start with testing the new scheme using the aaRS/tRNA pair to encode N-Acetyl-Lysine (AcKRS/tRNA). N-acetylation of lysine, in fact, is a reversible Post-Translational Modification with a regulatory role that rivals that of phosphorylation in eukaryotes. The implementation of the orthogonal translation system (OTS) to encode O-Phospho-L-serine was also planned however the BMII cells have to be previously engineered for correct O-Phospho-L-serine insertion. Indeed, the knockout of the *serB* gene (encoding a phosphoserine phosphatase) is required as phosphoserine is an intermediate in the biochemical pathway for the biosynthesis of Serine, and it would be therefore rapidly depleted in vivo in wild-type cells. This project has been started but not completed yet due to time constraints.

Finding a superior aaRS/tRNA pair to encode N-Acetyl-Lysine: sf-GFP read-through fluorescence assay

To begin with, I collected the most widely used AcKRS variants in literature, to compare them under the same experimental condition and to find the best working one in the B2H platform cells. Instead, the best performing SepRS variant in *E.coli*, namely the *Mmp*SepRS/tRNA$_{CUA}^{Sep}$, has been recently screened by Rogerson et al. [145] in Dr. Jason Chin Laboratory ((MRC-LMB, Cambridge UK) and chosen for this project.
The *Methanosarcina mazei Mm*AcKRS-1/*Mm*tRNA$_{CUA}^{Pyl}$, the *Methanosarcina barkeri Mb*AcKRS-3/*Mb*tRNA$_{CUA}^{Pyl}$ and an improved version of it bearing mutation in the N-terminus called *Mb*AcKRS3(IYPE)[18] were chosen [157]. At the same time, I sought to test a recently characterized new pyrrolysyl-tRNA synthetase PylRS/ tRNA$_{CUA}^{Pyl}$ pair (gently gifted from J.Chin lab) from *Methanomethylophilus alvus (Ma)* that was shown to outperform the

[18] Mutation from *Mb*AcKRS: V31I T56P H62Y A100E

Methanosarcina PylRSs in *Escherichia coli*. Since that *M.alvus* aaRS shared a homologous active site with *Mm*PylRS, I transplanted the mutations from the *Mm*AcKRS-1 active site into conserved position of *M.alvus* one to reprogram its specifity, devolping here the new variant referred as *Ma*AcKRS-1/tRNA$_{CUA}^{Pyl}$. Additionaly, I combined all of these synthetases in conjunction with either their cognate tRNA$_{CUA}^{Pyl}$ or a rationally optimized version of it tRNA$_{CUA}^{Pyl\text{-}opt}$ [283] with 6 nt changes in the stem loops; in the assay it was included the WT *Mb*PylRS[19]/tRNA$_{CUA}^{Pyl}$ as positive control.

For this comparative study the super-folded GFP (sfGFP) read through assay was chosen **[Figure 4.11 A].** The aaRS/tRNA cassettes have been subsequently cloned in the same recipient pRST plasmid (to iron out any possible confounding factors and test the activity of these amber suppression systems only) and transformed into TOP10 *E.coli* cells harboring the sfGFP150* reporter (the sfGFP bearing the TAG codon at position 150)[284]. The incorporation efficiency of AcK was so measured by the sfGFP fluorescence intensity. As it can be appreciated from **Figure 4.11 B**, the best performing synthetase turned out to be the *M.mazei* AcKRS-1 and the efficiency (~77% compared to the WT PylRS) was further improved when used together with tRNA$_{CUA}^{Pylopt}$; notably this AcKRS-1/tRNA$_{CUA}^{Pyl}$ was for the first-time tested here. Interestingly it is noteworthy that transplanting the mutations discovered in the *Mm*AcKRS-1 was sufficient to reprogram the specificity of *Ma* synthetase, although with low efficiency.

[19] The natural substrate of the WT PylRS is Pyrrol-Lysine. For this assay, the analog Nε-boc-l-lysine (BocK) was employed[304].

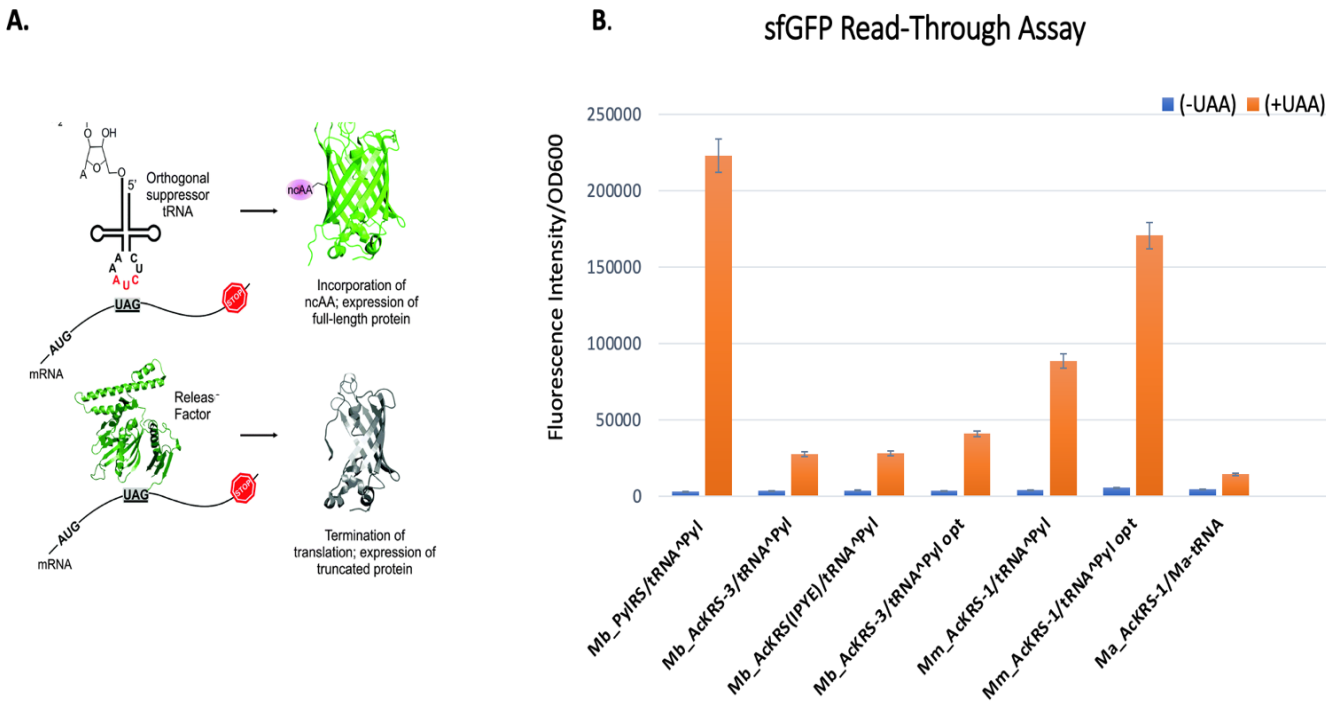


Figure 4.11. The MmAcKRS1/tRNA$_{CUA}^{Pyl}$ translational activity reached ~77 % compared to the WT PylRS counterpart. **A.** Schematic representation of the sfGFP Read-through Assay. It is based on a GFP reporter with an in-frame TAG codon at position 150 (a site known to be permissive for diverse ncAAs). The PTM-AAs incorporation is verified with the generation of the read-through GFP fluorescence directly in intact cells: green fluorescence of GFP should be detected in cells only in presence of ncAA. In the absence of the unnatural amino acid, no green fluorescence is detected **B. The read-through of amber codon in the sfGFP by the AcKRS/tRNA derivatives tested in this study was measured by fluorescence intensity.** Fluorescence signals are represented as bars, they are provided by the presence or absence of ncAAs and per each pair. The relative fluorescence intensities were calculated from the absolute fluorescence intensities read at 12 h normalized by the corresponding cell densities. The value for the sfGFP read-through by the WT PylRS is set as 100%. The mean values and standard errors were calculated from three replicates.

4.3.3 Testing the orthogonality of the aaRS/tRNA pairs in B2H recipient cells

Cloning the OTS component in the B2H bait plasmid

In order to combine the GCE technology with the B2H, the next step was to insert either the chosen *Mm*AcKRS-1/tRNA$^{Pyl\ opt}$ and the *Mmp*SepRS/tRNA^{Sep} into the B2H bait plasmid, without the need of engineering the bacterial strain. The cassettes were cloned in a stuffer portion of the pTGR plasmid using Gibson Assembly. The assembly allowed trimming away a few hundred base pairs from pTGR to limit its size increase when inserting the tRNA and

synthetase of about 20-25% (~400 bp removed on a ~2000 bp total insert size). The vectors were subsequently transformed in the BIIM cells.

Synthetase expression

Before the employment in a B2H screening, it must be verified that all the pair components were correctly expressed and functional in the B2H cell system. The expression of the Myc-tagged synthetases was assessed by SDS-PAGE followed by WB using an anti-MYC- HRP antibody. Both AcKRS and SepRS were correctly expressed **[Figure 4.12].** The cognate tRNAs were only evaluated using PCR and sequencing. tRNA transcription will also be verified with a Northern Blot (work in progress).

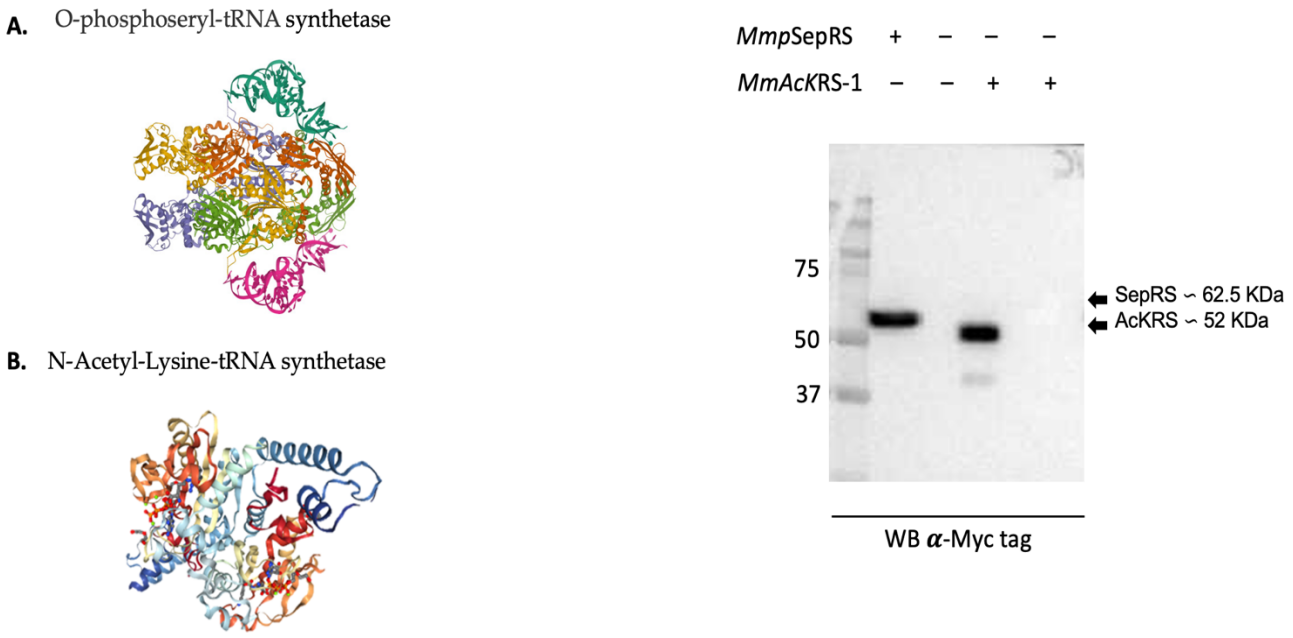


Figure 4.12. A) Crystal structure of the SepRS adapted from PDB 2DU5 file. **B)** Crystal structure of AcKRS adapted from PDB 2Q7H. **C)** WB using anti-MYC-HRP antibody on protein extracts of BIIM cells transformed with pTGR-SepRS/tRNA^{Sep} (lane 1), pTGR-tRNA^{Sep} (lane 2), pTGR-AckRS-1/tRNA^{Pyl} (lane 3) and pTGR-tRNA^{Pyl} (lane 4).

4.3.4 Encoding site specific B2H bait protein acetylation

In the meanwhile, I decide to directly test the orthogonality of the pair at least of the *Mm*AcKRS1/tRNA$_{CUA}^{Pyl\ opt}$ pair exploiting the sfGFP reporter. The rationale of the experiment

is that expression of sfGFP150*-HA bait and its detection using the a-HA antibody were strictly dependent on the incorporation of N-acetyl-lysine in response to the amber codon.

The BIIM cells expressing either the pTGR-*Mm*AcKRS1/tRNA$_{CUA}^{Pyl\ opt}$-sfGFP*-HA vector were cultured o/n in M9 complete medium with or without 2 mM N-ε-Acetyl-L- Lysine for 24 h at 37 °C. Production of sfGFP150*-HA reporter was induced by adding IPTG (100 μM) to the medium. Cells were cultured at 37°C for 2 hours 40' in a shaker, and then used to prepare protein cell extracts. Both the sample's fluorescence and the WB **[Figure 4.13]** showed that the sfGFP readthrough protein was abundantly present[20] in cells expressing the *Mm*AcKRS1/tRNA$^{Pyl\ opt}$ pair in presence of AcK confirming the correct setting of the entire system. Of course, Mass-Spectrometry confirmation should be performed to confirm the presence of AcK at the position 150 in the sfGFP purified protein (ongoing project). Importantly here, *the N-Acetyl-Lysine dependent amber suppression of a bait protein target was achieved in the B2H* platform, providing a rewarding validation of the novel B2H/GCE scheme.

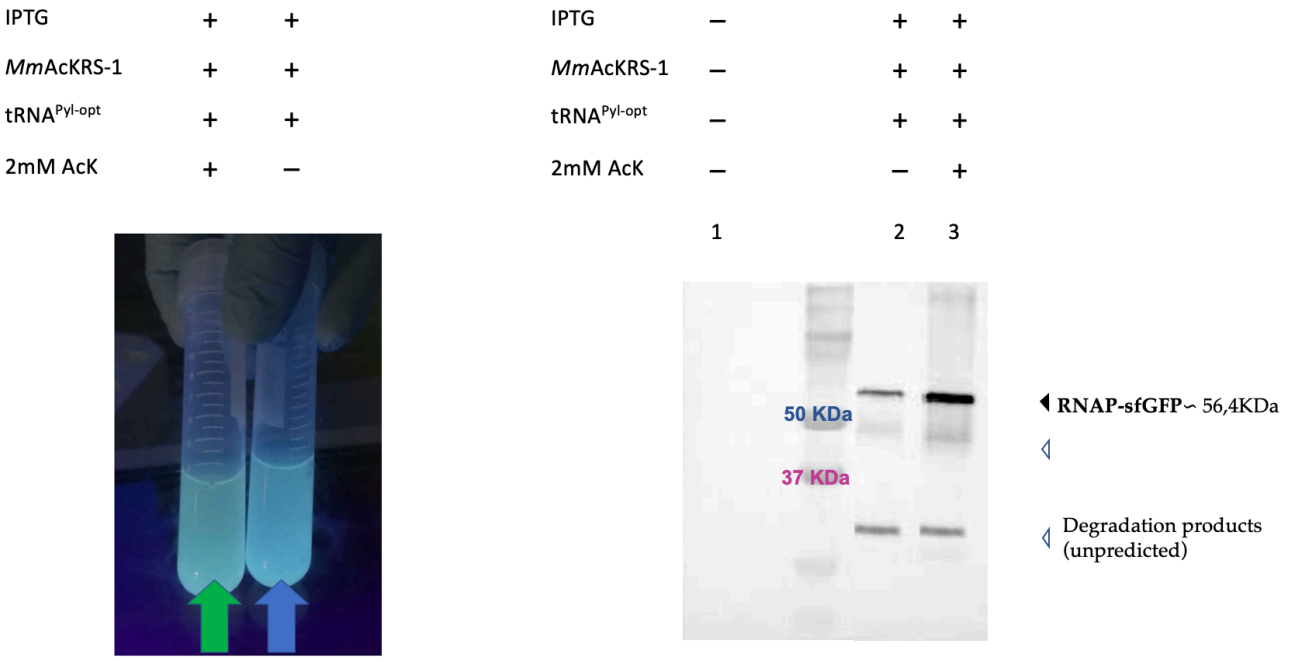


Figure 4.13. Amber suppression of the GFP bait by the MmAcKRS-1/tRNA in the B2H platform. (A) Detection of the full length sfGFP upon UV light exposure in BIIM *E.coli* culture supplemented with AcK. BIIM *E.coli* cells carrying the pTGR-MmAckRS-1/tRNA$^{Pyl-opt}$-sfGFP150*-HA plasmid were grown o/n for 24h at 37°C in M9 complete medium with (green arrow) or without (blue arrow) 2 mM AcK. The production of sfGFP150*-HA reporter was induced by adding IPTG (100 μM) to the medium. The fluorescent signal (green arrow) was detectable only in *E.coli* culture supplemented with N-Acetyl-Lysine. **(B)**

[20] Trace amounts of sfGFP-Ha were detected by Western Blot against the HA tag in the absence of N-acetyl-lysine. Unlikely, the BIIM reporter strain used for this experiment was found to contain the supE44 mutation, an amber suppressor mutation that allows some readthrough of TAG stop codons with insertion of glutamine. The presence of the supE44 was found to be in the duplicate copy of the *glnV* tRNA gene, *glnX*. This tRNA structural gene mutation, first isolated by Bachmann et al., 1976, causes the transition of the GUC glutamine anticodon to a AUC anticodon. However, for the purpose of the screening, we depleted this allele from the *E.coli* genome using the lambda red system. (Data not shown).

WB analysis to detect the read-through sfGFP protein (black arrowhead) using anti-HA-HRP antibody on protein extracts of BIIM cells transformed with empty bait plasmid pTGR (lane 1) and pTGR- MmAckRS-1/tRNA$^{Pyl\text{-}opt}$-sfGFP150*-HA vector grown in both presence (lane 2) or absence (lane 3) of N-Acetyl-Lysine. The full-length sfGFP protein (expected MW 56kDa) was abundantly present in culture expressing the amber suppression system in presence of AcK. Unpredicted degradation products were detected (white arrowheads) probably due to the cleavage of the linker between the RNAP and the GFP. Further optimization of the linker in the bait plasmid will be explored.

4.4 The PISA 2.0 Project *in E.coli:* Discussion

4.4.1 The proposed solution for PISA 2.0 application: integration of different orthogonal amber suppression aaRS/tRNA pairs into the Bacterial Two Hybrid platform

The aim of the PISA 2.0 project is to optimize the tethered catalysis method, overcoming its limitations (despite its successes), in order to accelerate and most importantly to streamline and automate the isolation of antibody domains against a single specific PTM of a given protein. Initially I envisaged to stably integrate the orthogonal (o-) o-aaRS/tRNA genes in the Y2H-PISA strain. The yeast cell would stably express the PT-modified protein target at a desired genetically encoded position residue. Having encountered several limitations in applying the GCE methodology in yeast, I decided to use the *E.coli* B2H asa selection host in lieu of the Y2H one, in order to finalize the PISA 2.0 Project. The novel B2H /Genetic Code Expansion (B2H/GCE) embodiment consists of a bacterial two-hybrid platform in which different orthogonal tRNA/aminoacyl-tRNA synthetase pair are readily integrated to genetically encode PTM-UAAs in the protein bait of interest. This would allow the bacterial-two hybrid platform to select intrabodies against a specific PTM.

Accordingly, during the last part of my thesis work, I accomplished the preliminary steps to adapt the PISA 2.0 project to bacteria (The PISA 2.0 Project in *E.coli)*, achieving the first proof of concept of the new scheme. In summary, this part of my thesis work led to the following achievements,

- Optimization and validation of the B2H bait and prey plasmids necessary to achieve the designed proof of principle;
- A trial B2H assay using known bait and prey interactors was assessed to test the system in our laboratory setting;
- After that, the Genetic Code Expansion methodology was adapted to the B2H platform. Two orthogonal tRNA/synthetase pairs - one for the insertion of N-ε-Acetyl-L-Lysine (AcKRS/tRNA) and the other for the insertion of O-Phospho-L-Serine (SepRS/SepT) – were incorporated in the B2H system vectors;
- Notably, an outperforming aaRS/tRNA pair to encode N-Acetyl-Lysine was first time tested here for the first time;

- The highly sensitive Read-through Fluorescence assay was employed to test the functionality of the aforementioned AcKRS/tRNA pair[21] in the novel embodiment. The amber suppression of a specific UAG codon in the sfGFP protein was successfully obtained in the *E.coli* "acetylator" platform, finally paving the way to the PISA 2.0 selection.

4.4.2 The Bacterial Two hybrid screening data discussion

The chosen B2H system

For this project, the B2H system by Muyldermans et al.[97,215] was employed since it is based on transcriptional activation of two reporter genes, similarly to our Y2H platform, allowing the possibility to validate the interaction between bait and target with reduced background. Moreover, it was previously proven by the successful recovery of antigen-specific Nanobodies of sub-micromolar affinity for HIV-1 integrase or nucleoside hydrolase [215]. Also, the sequence diversity, stability, antigen-specificity and affinity of these binders compared favorably to those that were retrieved in parallel by phage display pannings. Another characteristic that made this technique attractive was the low background level of 15–40% (or even less), which is even lower than some screenings of phage-displayed pannings [285].

Actually, another B2H variant, namely the protein fragments complementation assay (PCA) described by Mössner et al. [99] was available and would have been a possibility. In this version, an interaction between antibody and target, each fused to part of a split dihydrofolate reductase, reassembles the dissected enzyme to reconstitute its activity and to confer cell survival in presence of toxic trimethoprim. Although also this method was validated with a synthetic library of antibodies, it resulted in only a few poor binders recognizing their antigen [286]. Also, I assumed that the Achilles heel of such a system would be the spontaneous self-assembly of the protein parts contributing to a background, and the lack of a second reporter system to validate the interaction between bait and target.

[21] The orthogonal translational system (OTS) to encode phospho-serine has not been tested yet since further engineering is required.

4.4.3 Discussion of The Bacterial Two hybrid selection data

In this thesis work I provided the first steps to implement a new technology. Accordingly, we started testing the B2H assay via a cotransformation assay using a known bait a prey couple, the sfGFP and its binder αGFP Nb.
First, I optimized and validated the bait and prey plasmid. Bait and prey plasmid were modified in a way that any protein and VHH can be inserted in the vectors through the Gibson assembly method, leading to the insertion of an in-frame HA-tagged bait or myc-tagged VHH respectively.

To note, in the available format, the antibody domains are C-terminally fused -via their N-terminal end- to the bacteriophage λ repressor protein DBD. This configuration is the one exploited in the published paper; however, it does not appear to be ideal since in the antibody variable regions the N-terminal end extrudes from the antigen binding surface[287]. Besides, this format is opposite to the one in the Y2H SPLINT, in which the antibody domains have their N-terminal end free (they are fused as N-terminal extensions to LexA DNA binding domain). Accordingly, we are in the process of reverting the polarity of the prey construct fusing the antibody domain library via their C-terminal end to the DBD. The novel configuration will also provide the advantage of rapidly shifting the selected VHHs between yeast and bacterial system for further two-hybrid investigations.

4.4.4 The B2H/GCE selection scheme: a platform for the isolation of intrabodies against a single and *a priori* designed PTM.

In the proposed Bacterial-Two Hybrid/Genetic Code Expansion scheme, the orthogonal (o-) amynoacil-tRNA synthetase tRNA (aaRS/tRNA) pairs of interest were embedded in the B2H bait plasmid without the need of engineering the bacterial strain and the TAG codon was used as unique codon in the bait construction. Both the o-synthetase and o-tRNA gene cassettes were inserted into the B2H bait plasmid using Gibson Assembly method. Primers were designed so that any synthetase and any tRNA could be integrated in this plasmid with a single assembly reaction and limiting the vector size increase of about 20-25% (~400 bp removed on a ~2000 bp total insert size).
To demonstrate the feasibility of the new scheme, the orthogonal translation system (OTS) to encode the N-ε-Acetyl-L-Lysine (AcK) was chosen. The implementation of the OTS to encode O-Phospho-L-serine was also planned however the B2H cells have to be previously engineered

for O-Phospho-L-serine insertion. The knockout of the *serB* gene (encoding a phosphoserine phosphatase) is required since in *E.coli* phosphoserine is an intermediate in the biochemical pathway for the biosynthesis of Serine, and it would be therefore rapidly depleted in vivo in wild-type cells. This project (knockout of the *serB* phosphoserine phosphatase) has been started but not completed yet due to time constraints: the lambda red system [288] was used for *serB* deletion from the *E.coli* genome.

As far as the synthetase to encode AcK is concerned, different aaRS variants from archea were collected from literature and compared under the same experimental condition using the sfGFP read through assay in *E.coli* B2H platform cells. The best working synthetase turned out to be the *M.mazei* AcKRS-1, obtained in this work by transferring mutation from the active site of the *M. barkeri* AcKRS3 (kindly provided by Dr. Jason Chin. MRC-LMB.Cambridge,UK) into the conserved positions of *M.mazei* PylRS. The efficiency of this synthetase was further improved when used together with the optimized version of the cognate $tRNA_{CUA}^{Pyl}$, namely the $tRNA_{CUA}^{Pylopt}$ developed by [157]. Notably this AcKRS-1/tRNA combination was first time tested in this study.

The chosen AcKRS-1/tRNA was subcloned in the bait plasmid and tested for orthogonaity and functionality in the B2H setting using a sfGFP reporter. Both the sample's fluorescence and the WB **[Figure 4.13]** showed that the sfGFP readthrough protein was abundantly present in cells expressing the *Mm*AcKRS1/$tRNA^{Pyl\ opt}$ pair in presence of AcK confirming the correct setting of the entire system. Of course, Mass-Spectrometry confirmation should be performed to confirm the presence of AcK at the position 150 in the sfGFP purified protein (ongoing project). What I obtained here was a remarkable result: *I found an outperforming AcKRS-1/tRNA pair*[22] *to encode N-Acetyl-Lysine in a specific position residue of the bait protein in a 2H platform, finally paving the way to the PISA 2.0 selection. This is in fact the first proof of concept of the new B2H/GCE method*. Of course, further confirmations were needed and orthogonality of the expanded genetic code system in the B2H cells needed to be assessed using other protein targets and via MS.

Trace amounts of sfGFP-Ha were detected by Western Blot against the HA tag also in the absence of N-acetyl-lysine, resulting from mis-aminoacylation of the $tRNA_{CUA}$ by natural synthetase that are constitutively present in the cells [289]. This is consistent with previous reports [145,176]. Normally the background mis-acylation of the amber suppressor tRNA can be effectively outcompeted in the presence of an efficient cognate synthetase and its substrate [192]. In this regard, I reckon that using the pTGR low copy plasmid to express the amber suppressor pair does not allow us to reach a proper concentration of the tRNA and the AcKRS1 favoring the mischarging from other synthetase [289].

[22] The translational activity measure by the read-through GFP150* (the GFP bearing the TAG codon in correspondence with amino acid 150 (N)) fluorescence intensity was ~77% compared to the WT PylRS counterpart.

Furthermore, the BIIM reporter strain used for this experiment was found to contain the supE44 mutation, an amber suppressor mutation that allows some readthrough of TAG stop codons with insertion of glutamine. The presence of the supE44 was found to be in the duplicate copy of the *glnV* tRNA gene, *glnX*. This tRNA structural gene mutation, first isolated by Bachmann et al. [290], causes the transition of the GUC glutamine anticodon to a AUC anticodon. However, for the purpose of the screening, we have already planned to deplete this allele from the *E.coli* genome using the lambda red system [288].

After these improvements have been completed, the first selection against a PTM-containing bait in the new PISA2.0 system will be performed. Differently from the canonical B2H selection, here the VHH library will be challenged in the "acetylator" or "phosphorilator" stain able to recode the TAG stop codon in the target as Acetyl-Lysine or phosphor-Serine respectively. After that, the nanobody candidates will have to be counter-screened in three different strains. The VHH will be tested in the selection strain expressing either the empty bait plasmid (***α***-RNAP subunit only), or the full-length protein target to assess the anti-PTM specificity. At the same time, the candidates need to be challenged in a B2H strain expressing the amber-containing sequence of the bait in the absence of the aaRS/tRNA pair to control for the binding of the truncated version of the bait **[Figure 4.16].** I envisage that this new selection system will be quite fast and versatile. In fact, it can be readily adapted to the use of new orthogonal pairs and once a system for the co-translational insertion of a specific ncAA is implemented, it will be applicable to every protein. This is also a great advantage over the current PISA platform, which needs cloning of a specific enzyme for each new bait protein, even if the chemical nature of the PTM is the same.

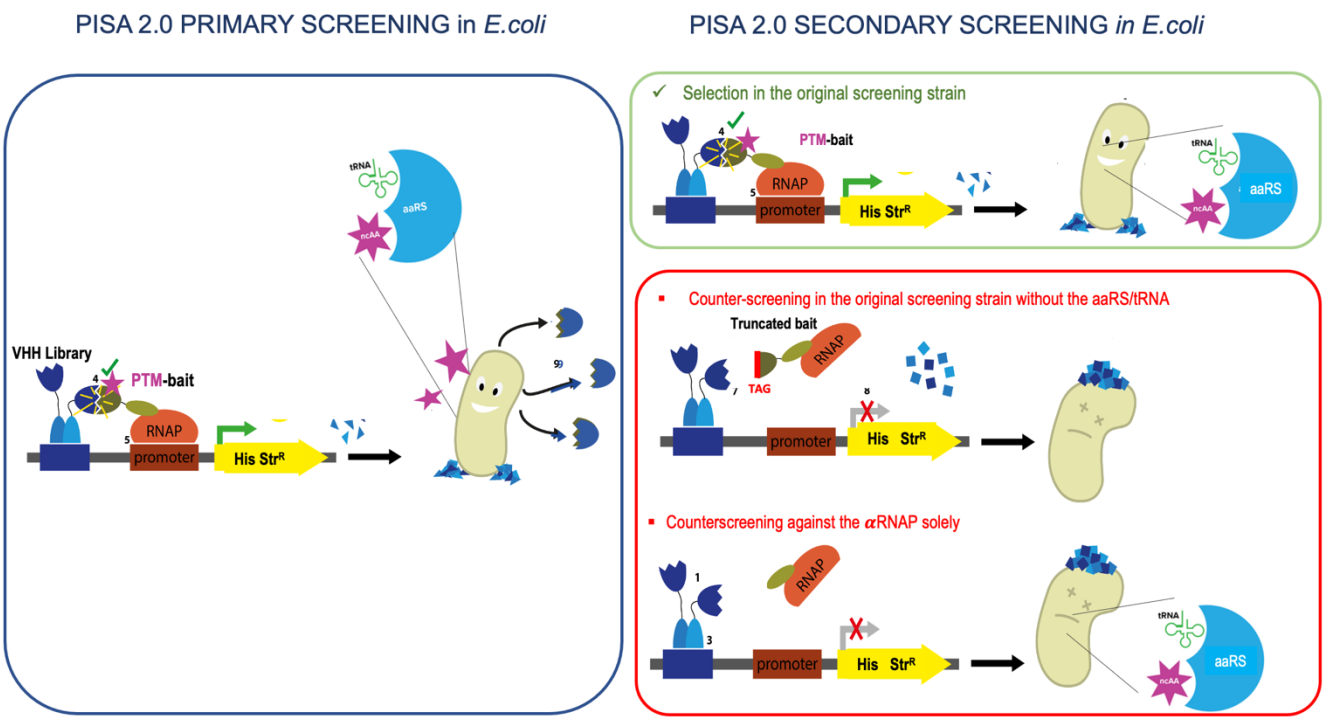


Figure 4.14. Schematic of the PISA 2.0 selection in *E.coli*. Differently from the canonical B2H selection, here the VHH library will be challenged in the "acetylator" or "phosphorilator" *E.coli* stain able to encode the TAG codon in the target as Acetyl-Lysine or phosphor-Serine respectively. After that the nanobody candidates will have to be counter-screened in three different strains. i) The VHH will be tested in the selection strain expressing either the empty bait plasmid (***α***-RNAP subunit only), ii) or the full-length protein target to assess the anti-PTM specificity. iii) At the same time, the candidates need to be challenged in a B2H strain expressing the bait (bearing the TAG) in the absence of the aaRS/tRNA pair to control for the binding of the truncated version of the bait.

4.5 The PISA 2.0 Project in *E.coli*: Future Perspective

4.5.1 The PISA 2.0 selections in the *E.coli* B2H strain genetically encoding Phospho-serine in proteins of interest

Unlimited bait construction and a plethora of possible interesting PTM targets make the future perspective of the PISA 2.0 technology enormous. However, some targets emerge more than others. In particular the first planned selection will be performed against the phosphorylated version of the Microtubule-associated protein Tau (Tau)[19] in B2H cells expressing the aaRS/tRNA pair to encode phosphor-Serine (SepRS/tRNA$_{CUA}^{Sep}$).

Tau pathology (the progressive deposition of Tau) correlates with cognitive decline in Alzheimer's disease (AD), a fatal neurodegenerative disease, for which no cure exists [291]. Thus, tau pathology is predictive of disease progression, supporting a central role of this protein [292], which makes it a prime target of AD diagnostic and therapeutic developments [293]. Pathologic Tau occurs in a hyperphosphorylated state **[Figure 4.16];** specifically, Tau iper-phosphorylation is linked to aggregation into multimers and fibrils [294] and modulates cognitive deficits induced by pathogenic tau or by Aβ [295]. Despite these clear associations of phosphorylated Tau with disease, phosphorylated tau also occurs in physiologic states [296], and site-specific tau phosphorylation mediates functions other than disease progression, including modulation of signaling complexes or lowering of microtubule binding affinity of tau itself [297].

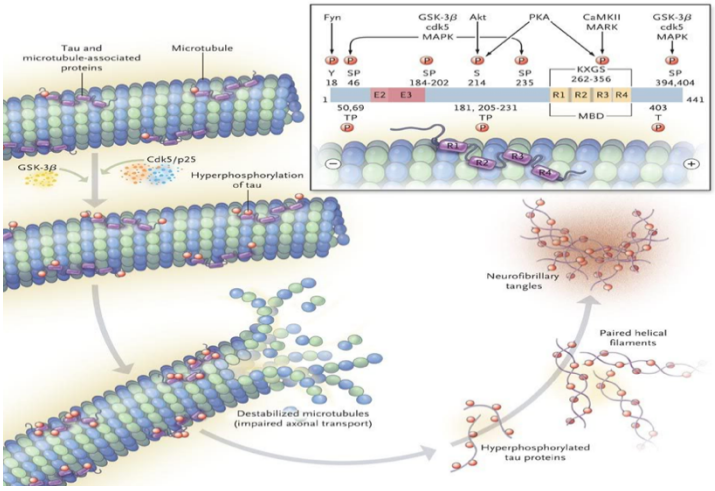


Figure 4.15. "Don't Phos Over Tau": Neurofibrillary degeneration of abnormally hyperphosphorylated Tau. Neurofibrillary degeneration of abnormally hyperphosphorylated Tau not only occurs in AD brain but is also seen in a family of related neurodegenerative diseases, called tauopathies. In every one of this tauopathies the neurofibrillary changes are made up of abnormally hyperphosphorylated tau and their occurrence in the neocortex is associated with dementia. Adapted from *flipper.diff.org*.

Interestingly, recent work [298] has suggested that a critical contribution to sequential tau phosphorylation arises from the initial phosphorylation at specific master sites (to name a few Ser422 and T181). Accordingly, the isolation of antibody domains specifically targeting the phosphorylated-Serine 422 (to name but one) of Tau (Phospho-TauSer422) achieved with PISA 2.0 platform would furnish an unprecedented tool to study the mechanistical link between site-specific and hyper- phosphorylation of Tau.

As a matter of course, before moving onto the main selection the novel "phosphorylator" *E.coli* (B2H cells expressing the SepRS/tRNA$_{CUA}^{Sep}$) platform will be tested taking advantage of an interaction test. The phosphorylator will be co-transformed with the Tau bait bearing the TAG codon in correspondence with Serine 235 and an ultra-specific Avia intrabody anti- Phospho-TauSer235 [299]. The latter has been extensively used in our laboratory and its functionality was proven by several Y2H screenings (unpublished data). After these improvements have been completed, the first PISA 2.0 selection against the Phospho- TauSer422 bait will be performed. Subsequently, the functional validation of the putative selected anti-phospho Tau intrabodies might be performed-for instance- in mammalian cell system stably expressing the cutting-edge FRET-based Conformational-Sensitive Tau sensor (CST)[300,301]. The CST developed at SNS allows to monitor Tau aggregation levels in live cells and at the same time to evaluate the effect exerted by the phospho-specific intrabody during this process[300]. Overall, the envisioned experiment may be a first hint to answer a central question in Alzheimer's research: Why and how does the neuronal tau protein get progressively hyperphosphorylated? This also raises the exciting possibility that therapeutic inhibition of master sites will block intrinsic augmenting mechanisms of tau phosphorylation.

Finally, it is worth mentioning that an alternative to specific phospo-serine inhibition or neutralization by binding competition is represented by protein degradation. As a matter of fact, the anti-phospho Tau antibodies can be seen as binding modules to which any degradation module can be appended. In this context, a pioneering approach exploited E3 ubiquitin ligases for targeted degradation, by fusing intrabodies to ubiquitin–proteasome pathway substrates to achieve ligand-inducible fast degradation of the target protein (suicide intrabody technology; SIT) [302]. This method was successfully used to conditionally and reversibly target the full length tau protein for proteasome degradation [302]. In light of this previous result, an

extension of SIT with new anti-phospho serine intrabodies fused to a more orthogonal degron (e.g. the LOV domain) would allow one to selectively deplete a cell of the PT-modified pathological Tau, without affecting the levels of the unmodified protein, and, thus, maintain healthy cellular function in AD and other tauopathies.

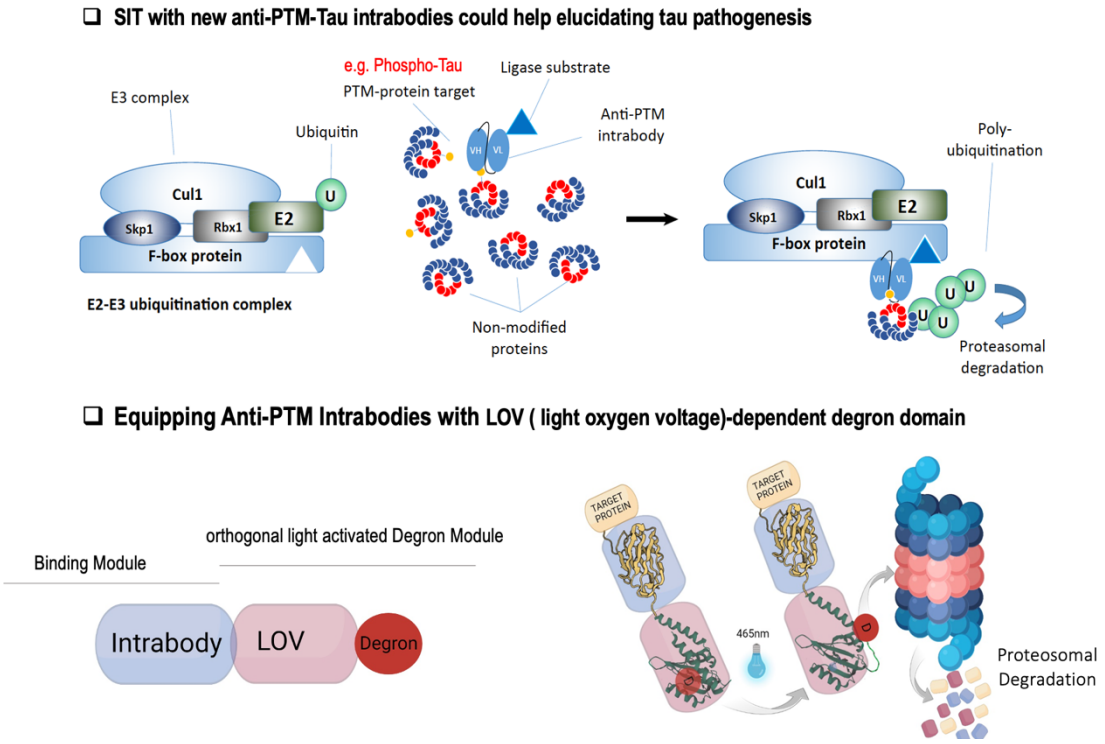


Figure 4.16.Equipping PISA 2.0 intrabodies with effector functions. **Top** with (Suicide Intrabody Technology) SIT, an intrabody can carry a target to degradation via the ubiquitin–proteasome pathway, in response to a ligand. The intrabody is fused to a E3 ligase substrate, which is bound and polyubiquitinated. The whole complex is finally degraded. SIT allows one to selectively deplete a cell of the PTM protein, without affecting the levels of the unmodified protein. Adapted from Chirichella and Cattaneo Trends in Biotechnology, 2017. **Bottom. Schematic of an orthogonal light activated nanobody tool**, called optobody, which is the first intrabody based optogenetic tool to modulate protein degradation (unpublished work). Optobodies are bifunctional and precisely controlled because on one side they bind the protein target, and on the other side they can escort the bound target to the proteasome dependent degradation upon blue light stimulation. (Author Angela Bitonti, PhD candidate at SNS).

4.5.2 The power of the PISA 2.0 Platform and what needs it will solve

Perturbing a biological system by selectively targeting a single PTM of a protein represents a huge (unsolved) objective of the proteomic era. The cell can be imagined as an ensemble of protein–protein interactions (the edges) mediated by distinct parts of a protein, including PTMs. that allow a precise cellular function. Indeed, ablation of a protein node will result in the disruption of all the edges, while a single PTM (edge)-selective interference will not affect all the other PTM edges, granting a new level of precision and specificity in the description of intra- and intercellular processes. Despite the large number of targets and the critical role of PTMs in connecting and branching protein networks, the validation of these targets in a research-oriented or in a therapeutic perspective has been lagging behind because of the lack of *in cell* PTM-specific interference methods. The PISA technology has been solving this gap in studying protein function, allowing the intracellular selection of antibody domains targeting native post-translationally modified proteins. However, targeting the post-translational modification of proteins implies the possibility of dissecting the individual contribution of each individual PTM. In the current protocol, the tethered catalysis-based PISA screening can provide only a limited solution to this problem, because chimeric baits are in principle modified at all the possible sites. In the proposed PISA 2.0 version, by exploiting the expanded genetic code technology, the systematic isolation of anti-individual-PTMs intrabodies will be greatly accelerated. Presently, PISA Technology is the only existing method that is able to isolate intrabodies against native PTM-proteins. Current methods to select PTM-selective antibodies, such as animal immunization or phage display[22,58] , use chemically modified linear peptides as antigens [65]. Pitfalls include using unstructured peptides with chemically synthesized PTMs instead of native PTM proteins, and generating antibody binders that are not adequate for their use in cells. Thus, the conformation of a linear unstructured PTM-containing peptide used as antigen will be different from that of the corresponding PTM-modified sequence in the context of the native protein. Therefore, these anti-PTM antibodies will not always recognize the native protein. Moreover, selecting antibody domains from phage display libraries that fold in the periplasm of Escherichia coli cells does not guarantee that the antibody will function as a neutralizing intracellular antibody within the protein network in the cell, because of the different redox conditions [51,86,303]. Thus, the antibody domains selected by monoclonal antibody or phage display technologies might be suitable for use as PTM detection probes, but not as intracellular function-neutralizing agents against native PTM proteins. On the contrary, the PISA 2.0 intrabody selection occurs in cells in which expanded genetic code technologies allow site specific genetic encoding of a single PTM of native proteins able to fold in their natural intracellular environment **[cfr. Table 4.4].**

Selection Method	Type of Selection	Manipulation/Purification of Antigen	PTM catalysis	Antibody Folding
Phage Display	In vitro	Yes	Post-purification in vitro	Periplasmic space of Bacteria
Yeast Display	In vitro	Yes	Post-purification in vitro	Biosynthetic-Secretory Pathway
Immunization of Animals	Extracellular	Yes	Post-purification in vitro	Biosynthetic-Secretory Pathway
PISA	In vivo	No	Native, in vivo	Cytoplasm
PISA 2.0	In vivo	No	Genetically encode and Site-specific, native, in vivo	Cytoplasm

Table 4.3. Comparison between different antibody selection technologies. PISA is the only platform that is able to select intrabodies *in vivo* without whatsoever manipulation of the PTM antigen. Protein antigen folds in the cell, in its native form. The PISA 2.0 developed during this thesis project allows to genetically encoded the PTM in a site-specific position residue : i)surpassing the limitation related to tethered-catalysis and ii) greatly accelerating the PISA selection protocol.

5. Conclusion

5 Conclusion

Post-translationally modified proteins represent a huge and untapped source of biologically important and disease-relevant targets, but their systematic biological elucidation and validation for therapeutic purposes is hampered by the lack of specific tools and experimental strategies. In our laboratory, we have recently shown the feasibility of an experimental strategy to directly target single Post Translational Modifications (PTMs) with PTM-specific intrabodies: the PISA Technology.

In the first part of this thesis project, the silencing activity of the acetylation-selective intrabody scFv- 58F, selected through PISA, has been successfully characterized.
In particular, for the first time it has been proved that the effect of the direct interference on a PTM-modified protein, achieved by a PTM-specific intrabody, is highly more selective than that obtained inhibiting the enzymes responsible for the modification itself. We provided here the first comparative evidence that the direct interference with a single acetylated residue, obtained by the ScFv-58F, results in a more specific effect on gene modulation, compared to that achieved by a broader interference using current HAT-targeting small molecule inhibitors *in S. cerevisiae*. Indeed, data obtained by NGS analysis revealed that the downstream effects of scFv-58F on global gene expression were highly selective and restricted to a small subset of genes. On the contrary, the two HAT inhibitors considered in this project displayed wider effects on gene expression. Also, gene ontology, gene-by-gene analysis as well as independent validation in the literature, encouraged us to go deeper in order to further characterize the specific activity of this intrabody: the comparison with known bromodomain inhibitors has been planned. Last but not least, our data indicate the ability of scFv-58F to induce specific transcriptional regulation of target genes in non-yeast cells. This finding opens novel avenues for future studies in mammalian systems in vivo and particularly in the nervous system, where neuroepigenetic studies would greatly benefit from this tool: an adeno-associated-virus (AAV) expressing the scFv-58F-NLS has been already generated to facilitate this application.
Importantly, all these results motivate us to move forward with the purpose to extend the anti-PTM approach to new biological applications. The PTM-binding antibody moiety can be further implemented and tailored by adding suitable effector functions allowing, for instance, PTM-specific targeted degradation of the acetylated protein versus the non-acetylated protein pool. Also, the PTM-specific intrabody can be engineered to recruit actuators for chromatin retargeting in living cells or for live imaging of the PTM pool of the protein. This concept represents a milestone for this thesis project.

Finally, in a therapeutic perspective, anti-PTM intrabodies could be delivered as genes or, alternatively, as protein macrodrugs delivered by cell penetrating peptides or by nanocarriers.
In conclusion, in this first thesis project we have validated the PTM Antibody Chromatin Retargeting approach, as a new strategy to address challenging questions in epigenetic research. At the same time, we have prospected the post-translational-specific intrabodies - selected by PISA- as novel specific pharmacological candidates for a variety of research and clinical applications (e.g. in cancer and neurological diseases).

Overall, these results highlighted the potential of the PISA Technology that, however, in its present form, based on tethered catalysis to install the PTM, may have some limitations. Tethered catalysis baits can be, in fact, modified at all possible sites hampering the possibility of dissecting the contribution of each individual PTM.
The second part of my thesis work was therefore aimed at implementing the current PISA platform exploiting the Genetic Code Expansion (GCE) (instead of tethered catalysis) developing the novel *PISA 2.0 version***.**

Initially, I provided a proof of concept of the method adapting the GCE methodology to PISA ***(The PISA 2.0 Project in yeast):*** the yeast-two hybrid platform was tailored leveraging on novel orthogonal aminoacyl-tRNA synthetase/tRNA (aaRS/tRNA) pairs, designed by molecular evolution to site-specifically incorporate Post Translationally Modified amino acids (PTMs) in target proteins. Specifically, an acetylator yeast strain was produced via the integration of the aaRS/tRNA pair *to* incorporate N-acetyl-Lysine (*Mb*AcKRS3/tRNA) in correspondence with the amber stop codon (TAG) in protein targets. The acetylator was therefore able to express the acetylated targets at a desired position residue, ready to be used as baits for intrabody selections (with no manipulation whatsoever of the PTM antigen). Robust preliminary results were obtained: the N-Acetyl Lysine dependent amber suppression of Histone H3 bait was successfully achieved. Nevertheless, data also showed rather low readthrough protein yields. Hence, all the state-of-the-art technological procedures to improve the read-through protein yield have been systematically applied, in a step-by-step approach, to the "acetylator" strain: a good PT-modified protein level is in fact paramount for the PISA 2.0 selection. First, the target mRNA stability was improved via the nonsense-mediated mRNA decay (NMD) pathway inactivation. In addition, a mutant library of the AcKRS synthetase was generated in order to select a more active variant. Nonetheless, these experiments still suggested a suboptimal activity of the AcKRS/tRNA pair at least in our strain.

Accordingly, the translational activity of recently discovered archaea aaRS/tRNA couples have been evaluated to find a fully active one: a thorough comparative study of these amber

suppression pairs was performed using a High-sensitive Fluorescence assay. Notably, also the cutting-edge tRNA-Extension Technology has been used for the first time in yeast cells to shed light onto the low read-through protein yields reported in this model organism.
In the end, all the results collected highlighted some limitation in the application of the GCE method in *S. cerevisiae:* although all the amber suppressing pairs tested were well expressed and the intracellular concentration of the PTM-AAs was in the working range, I was not able to observe any adequate protein read-through level, even using solid experimental design. Still, these findings suggested that it would be possible to implement the 2.0 system, provided a superior aaRS/tRNA pair. Archea synthetase derivatives are in fact relatively underdeveloped in *S.cerevisiae compared to E.coli*, suggesting that they may need to be evolved directly in yeast. Auspiciously, i) the vector series to quickly integrate different pairs in the *S. cerevisiae* genome, ii) the Y2H strains tailored to inactivate the NMD pathway and engineered to encode Acetyl-Lysine and P-serine iii) the randomized mutagenesis AcKRS mutant library provided in this thesis work, would be the ideal and robust yeast platform and toolkit for discovering and evolving the required outperforming aaRS/tRNA pairs directly in this essential model organism. On a final note, even though these findings did not turn out directly useful for the PISA 2.0 project, they are essential to draw a complete picture of GCE application and to achieve a better understanding of how would yeast, or more generally a eukaryotic host, respond to perturbations by these heterologous systems. The systematic results I obtained, as well as the tools/strains I developed, are foundational for the development of an improved platform for GCE in yeast.

Since, in contrast to *S.cerevisiae*, the vast majority of efforts on Genetic Code Expansion to date has been conducted *in E. coli*, I went on to finalize the PISA 2.0 project in bacteria taking advantage of the Bacterial-Two-Hybrid (B2H) system (in lieu of the Y2H) as intrabody selection platform and the wide range of functional and orthogonal aaRS/tRNA pairs available in this model organism. Accordingly, during the last part of my thesis work, I accomplished the preliminary steps to adapt the PISA 2.0 project to the *E.coli* platform.
First, the B2H platform was successfully proven by an *in vivo* co-transformation assay between a known bait/praey couple. This result provided a rewarding validation to move on implementing the GCE methodology in the B2H strain in lieu of yeast. To begin with, an acetylator B2H strain was set up by the incorporation of an outperforming and in-house screened pair to encode N-Acetyl Lysine (AcK) in the B2H vectors. Remarkably, the AcK-dependent amber suppression of a specific UAG codon in the target bait was successfully obtained, providing the first proof of concept of the PISA 2.0 also in bacteria.

A “phosphorylator” B2H strain was also generated, via the integration of a top-notch aaRS/tRNA pair to encode phospho-serine (developed by Chin et al). However, the platform has not been tested yet since further engineering of the *E. coli* strain was required.

After these improvements have been completed, the first selection against a PTM-containing bait in the new PISA 2.0 system will be performed. In particular the first selection has already been planned against the phosphorylated version of the Microtubule-associated protein Tau (Tau) in B2H cells expressing the aaRS/tRNA pair to encode Phospho-serine. There is a growing need for phospho-specific binding molecules for Tau proteins, given the steady increase in the number of functionally important phosphorylation sites that have been uncovered. The ability of the proposed 2.0 approach to rapidly furnish these intracellularly working anti-PTM tools is significant, especially compared to conventional methods such as animal immunization and phage display. Presently, PISA Technology is the only existing method that is able to isolate intrabodies against native PTM-proteins, including phosphoproteins (especially, in the 2.0 version).

In conclusion, the provided “acetylator” or “phosphorylator” two-hybrid strains will be available for use with standard PISA protocols. This novel platform will allow to accelerate and automatize the number of selections performed, generating a pipeline of anti-PTM binding domains for target validation, functional studies and potentially therapeutic purposes.

6 References

[1] L.M. Smith, J.N. Agar, J. Chamot-Rooke, P.O. Danis, Y. Ge, J.A. Loo, L. Paša-Toliæ, Y.O. Tsybin, N.L. Kelleher, The human proteoform project: Defining the human proteome, Sci. Adv. 7 (2021) 1–9. https://doi.org/10.1126/sciadv.abk0734.

[2] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, P. Walter, Molecular Biology of the Cell, 2014. https://doi.org/10.1017/CBO9781107415324.004.

[3] M.R. Wilkins, S.K. Kummerfeld, Sticking together? Falling apart? Exploring the dynamics of the interactome, Trends Biochem. Sci. 33 (2008) 195–200. https://doi.org/10.1016/j.tibs.2008.03.001.

[4] K.E. Krueger, S. Srivastava, Posttranslational protein modifications: current implications for cancer detection, prevention, and therapeutics, Mol. Cell. Proteomics. 5 (2006) 1799–1810. https://doi.org/10.1074/mcp.R600009-MCP200.

[5] B.D. Strahl, C.D. Allis, The language of covalent histone modifications., Nature. 403 (2000) 41–45. https://doi.org/10.1038/47412.

[6] A.J. Bannister, T. Kouzarides, Regulation of chromatin by histone modifications, Cell Res. 21 (2011) 381–395. https://doi.org/10.1038/cr.2011.22.

[7] L.A. Gates, J. Shi, A.D. Rohira, Q. Feng, B. Zhu, M.T. Bedford, C.A. Sagum, S.Y. Jung, J. Qin, M.-J. Tsai, S.Y. Tsai, W. Li, C.E. Foulds, B.W. O'Malley, Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation, J. Biol. Chem. 292 (2017) 14456–14472. https://doi.org/10.1074/jbc.M117.802074.

[8] I. Ali, R.J. Conrad, E. Verdin, M. Ott, Lysine Acetylation Goes Global: From Epigenetics to Metabolism and Therapeutics, Chem. Rev. 118 (2018) 1216–1252. https://doi.org/10.1021/acs.chemrev.7b00181.

[9] N. Ahuja, A.R. Sharma, S.B. Baylin, Epigenetic Therapeutics: A New Weapon in the War Against Cancer, Annu. Rev. Med. 67 (2016) 73–89. https://doi.org/10.1146/annurev-med-111314-035900.

[10] F.J. Stevens, P.R. Pokkuluri, M. Schiffer, Protein Conformation and Disease: Pathological Consequences of Analogous Mutations in Homologous Proteins, Biochemistry. 39 (2000) 15291–15296. https://doi.org/10.1021/bi001017x.

[11] S. Ramazi, J. Zahiri, Post-translational modifications in proteins: Resources, tools and prediction methods, Database. 2021 (2021) 1–20. https://doi.org/10.1093/database/baab012.

[12] P. Chi, C.D. Allis, G.G. Wang, Covalent histone modifications — miswritten, misinterpreted and mis-erased in human cancers, Nat. Rev. Cancer. 10 (2010) 457–469. https://doi.org/10.1038/nrc2876.

[13] D.J. Irwin, T.J. Cohen, M. Grossman, S.E. Arnold, S.X. Xie, V.M.-Y. Lee, J.Q. Trojanowski, Acetylated tau, a novel pathological signature in Alzheimer's disease and other tauopathies, Brain. 135 (2012) 807–818. https://doi.org/10.1093/brain/aws013.

[14] M.R. Larsen, M.B. Trelle, T.E. Thingholm, O.N. Jensen, Analysis of posttranslational modifications of proteins by tandem mass spectrometry, Biotechniques. 40 (2006) 790–798. https://doi.org/10.2144/000112201.

[15] J.-J. Vasquez, C. Wedel, R.O. Cosentino, T.N. Siegel, Exploiting CRISPR–Cas9 technology to investigate individual histone modifications, Nucleic Acids Res. 46 (2018) e106–e106. https://doi.org/10.1093/nar/gky517.

[16] D.Y. Kwon, Y.T. Zhao, J.M. Lamonica, Z. Zhou, Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC, Nat. Commun. 8 (2017). https://doi.org/10.1038/ncomms15315.

[17] F. Duclot, J. Meffre, C. Jacquet, C. Gongora, T. Maurice, Mice knock out for the histone acetyltransferase p300/CREB binding protein-associated factor develop a resistance to amyloid toxicity, Neuroscience. 167 (2010) 850–863. https://doi.org/10.1016/j.neuroscience.2010.02.055.

[18] G.J. Hannon, RNA interference, Nature. 418 (2002) 244–251. https://doi.org/10.1038/418244a [doi]\r418244a [pii].

[19] Y. Wang, E. Mandelkow, Tau in physiology and pathology., Nat. Rev. Neurosci. 17 (2015) 22–35. https://doi.org/10.1038/nrn.2015.1.

[20] C.H. Arrowsmith, C. Bountra, P. V. Fish, K. Lee, M. Schapira, Epigenetic protein families: A new frontier for drug discovery, Nat. Rev. Drug Discov. 11 (2012) 384–400. https://doi.org/10.1038/nrd3674.

[21] J.L. Dahlin, K.M. Nelson, J.M. Strasser, D. Barsyte-Lovejoy, M.M. Szewczyk, S. Organ, M. Cuellar, G. Singh, J.H. Shrimp, N. Nguyen, J.L. Meier, C.H. Arrowsmith, P.J. Brown, J.B. Baell, M.A. Walters, Assay interference and off-target liabilities of reported histone acetyltransferase inhibitors, Nat. Commun. 8 (2017) 1527. https://doi.org/10.1038/s41467-017-01657-3.

[22] A.R.M. Bradbury, S. Sidhu, S. Dübel, J. McCafferty, Beyond natural antibodies: the power of in vitro display technologies, Nat. Biotechnol. 29 (2011) 245–254. https://doi.org/10.1038/nbt.1791.

[23] A.F. Rudolf, T. Skovgaard, S. Knapp, L.J. Jensen, J. Berthelsen, A comparison of protein kinases inhibitor screening methods using both enzymatic activity and binding affinity determination, PLoS One. 9 (2014). https://doi.org/10.1371/journal.pone.0098800.

[24] K.-C. Hsu, C.-Y. Liu, T.E. Lin, J.-H. Hsieh, T.-Y. Sung, H.-J. Tseng, J.-M. Yang, W.-J. Huang, Novel Class IIa-Selective Histone Deacetylase Inhibitors Discovered Using an in Silico Virtual Screening Approach, Sci. Rep. 7 (2017) 3228. https://doi.org/10.1038/s41598-017-03417-1.

[25] F. Manzo, F.P. Tambaro, A. Mai, L. Altucci, Histone acetyltransferase inhibitors and preclinical studies, Expert Opin. Ther. Pat. 19 (2009) 761–774. https://doi.org/10.1517/13543770902895727.

[26] M. Faria Freitas, M. Cuendet, P. Bertrand, HDAC inhibitors: a 2013–2017 patent survey, Expert Opin. Ther. Pat. 28 (2018) 365–381. https://doi.org/10.1080/13543776.2018.1459568.

[27] P. Filippakopoulos, J. Qi, S. Picaud, Y. Shen, W.B. Smith, O. Fedorov, E.M. Morse, T. Keates, T.T. Hickman, I. Felletar, M. Philpott, S. Munro, M.R. McKeown, Y. Wang, A.L. Christie, N. West, M.J. Cameron, B. Schwartz, T.D. Heightman, N. La Thangue, C.A. French, O. Wiest, A.L. Kung, S. Knapp, J.E. Bradner, Selective inhibition of BET bromodomains, Nature. 468 (2010) 1067–1073. https://doi.org/10.1038/nature09504.

[28] P. Filippakopoulos, S. Knapp, Targeting bromodomains: Epigenetic readers of lysine acetylation, Nat. Rev. Drug Discov. 13 (2014) 337–356. https://doi.org/10.1038/nrd4286.

[29] S. Muller, P. Filippakopoulos, S. Knapp, Bromodomains as therapeutic targets, Expert Rev. Mol. Med. 13 (2011). https://doi.org/10.1017/s1462399411001992.

[30] C.D. Allis, S.L. Berger, J. Cote, S. Dent, T. Jenuwien, T. Kouzarides, L. Pillus, D. Reinberg, Y. Shi, R. Shiekhattar, A. Shilatifard, J. Workman, Y. Zhang, New Nomenclature for Chromatin-Modifying Enzymes, Cell. 131 (2007) 633–636. https://doi.org/10.1016/j.cell.2007.10.039.

[31] S. Biocca, A. Cattaneo, Intracellular immunization: antibody targeting to subcellular compartments, Trends Cell Biol. 5 (1995) 248–252. https://doi.org/10.1016/S0962-8924(00)89019-4.

[32] M. Chirichella, S. Lisi, M. Fantini, M. Goracci, M. Calvello, R. Brandi, I. Arisi, M. D'Onofrio, C. Di Primio, A. Cattaneo, Post-translational selective intracellular silencing of acetylated proteins with de novo selected intrabodies, Nat. Methods. 14 (2017) 279–282. https://doi.org/10.1038/nmeth.4144.

[33] A. Abbas, A. Lichtman, S. Pillai, Cellular and molecular immunology, 2015. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Cellular+and+Molecular+Immunology+Saunders#4%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Cellular+and+molecular+immunology+Saunders+Elsevier%231.

[34] C. Milstein, B. Frangione, Disulphide bridges of the heavy chain of human immunoglobulin G2., Biochem. J. 121 (1971) 217–225.

[35] T.H. Rabbitts, G. Matthyssens, P.H. Hamlyn, Contribution of immunoglobulin heavy-chain variable-region genes to antibody diversity, Nature. 284 (1980) 238–243.

https://doi.org/10.1038/284238a0.

[36] J.L. Xu, M.M. Davis, Diversity in the CDR3 Region of V H Is Sufficient for Most Antibody Specificities, Immunity. 13 (2000) 37–45. https://doi.org/10.1016/S1074-7613(00)00006-6.

[37] C. Milstein, The disulphide bridges of immunoglobulin $\varkappa$-chains, Biochem. J. 101 (1966) 338–351. https://doi.org/10.1042/bj1010338.

[38] The Croonian Lecture, 1989 Antibodies: a paradigm for the biology of molecular recognition, Proc. R. Soc. London. B. Biol. Sci. 239 (1990) 1–16. https://doi.org/10.1098/rspb.1990.0006.

[39] T.A. Yednock, C. Cannon, L.C. Fritz, F. Sanchez-Madrid, L. Steinman, N. Karin, Prevention of experimental autoimmune encephalomyelitis by antibodies against α4βl integrin, Nature. 356 (1992) 63–66. https://doi.org/10.1038/356063a0.

[40] R.-M. Lu, Y.-C. Hwang, I.-J. Liu, C.-C. Lee, H.-Z. Tsai, H.-J. Li, H.-C. Wu, Development of therapeutic antibodies for the treatment of diseases, J. Biomed. Sci. 27 (2020) 1. https://doi.org/10.1186/s12929-019-0592-z.

[41] M.R. Gaudinski, E.E. Coates, L. Novik, A. Widge, K. V Houser, E. Burch, L.A. Holman, I.J. Gordon, G.L. Chen, C. Carter, M. Nason, S. Sitar, G. Yamshchikov, N. Berkowitz, C. Andrews, S. Vazquez, C. Laurencot, J. Misasi, F. Arnold, K. Carlton, H. Lawlor, J. Gall, R.T. Bailer, A. McDermott, E. Capparelli, R.A. Koup, J.R. Mascola, B.S. Graham, N.J. Sullivan, J.E. Ledgerwood, C.S. Hendel, S.H. Plummer, P. Costner, J. Saunders, F. Mendoza, A.M. Eshun, J. Casazza, A. Ola, W. Whalen, X. Wang, J. Cunningham, O. Vasilenko, C.R. Boyd, O. Trofymenko, M.C. Burgos Florez, S. Hickman, R.S. Rothwell, I.R. Pittman, L.N. Le, B.D. Larkin, J.H. Cox, P.J. Apte, R.T. Hicks, C. Trelles Cartagena, P. V Williams, L. Requilman, T. Nguyen, C. Tran, S. Vazquez, M. Conan-Cibotti, J. Stein, T. Beresnev, Safety, tolerability, pharmacokinetics, and immunogenicity of the therapeutic monoclonal antibody mAb114 targeting Ebola virus glycoprotein (VRC 608): an open-label phase 1 study, Lancet. 393 (2019) 889–898. https://doi.org/10.1016/S0140-6736(19)30036-4.

[42] S. Vita, S. Rosati, T. Ascoli Bartoli, A. Beccacece, A. D'Abramo, A. Mariano, L. Scorzolini, D. Goletti, E. Nicastri, Monoclonal Antibodies for Pre- and Postexposure Prophylaxis of COVID-19: Review of the Literature, Pathogens. 11 (2022) 882. https://doi.org/10.3390/pathogens11080882.

[43] W. Towner, E. DeJesus, S. Schrader, J. De Vente, C. McGary, M. Zogheib, S. Weinheimer, P. Mesquita, 1027. Long-Term Efficacy, Safety, and Durability of Ibalizumab-Based Regimens in Subgroup of TMB-202 Participants, Open Forum Infect. Dis. 7 (2020) S542–S543. https://doi.org/10.1093/ofid/ofaa439.1213.

[44] G. Winter, C. Milstein, Man-made antibodies, Nature. 349 (1991) 293–299. https://doi.org/10.1038/350055a0.

[45] A. Plückthun, A. Skerra, [34] Expression of functional antibody Fv and Fab fragments in Escherichia coli, in: 1989: pp. 497–515. https://doi.org/10.1016/0076-6879(89)78036-8.

[46] J.S. Huston, D. Levinson, M. Mudgett-Hunter, M.S. Tai, J. Novotný, M.N. Margolies, R.J. Ridge, R.E. Bruccoleri, E. Haber, R. Crea, Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli., Proc. Natl. Acad. Sci. 85 (1988) 5879–5883. https://doi.org/10.1073/pnas.85.16.5879.

[47] M.-P. Lefranc, V. Giudicelli, C. Ginestoux, J. Jabado-Michaloud, G. Folch, F. Bellahcene, Y. Wu, E. Gemrot, X. Brochet, J. Lane, L. Regnier, F. Ehrenmann, G. Lefranc, P. Duroux, IMGT(R), the international ImMunoGeneTics information system(R), Nucleic Acids Res. 37 (2009) D1006–D1012. https://doi.org/10.1093/nar/gkn838.

[48] S.E. Ward, D. Gussow, A.D. Griffiths, J.P. T, W. Greg, Binding activities of a repertoir of single immunoglobulin variable domains secreted from Escherichia coli, Nature. 340 (1989) 301–303. https://doi.org/10.1038/340301a0.

[49] I. Jovčevska, S. Muyldermans, The Therapeutic Potential of Nanobodies, BioDrugs. 34 (2020) 11–26. https://doi.org/10.1007/s40259-019-00392-z.

[50] A. Bates, C.A. Power, David vs. Goliath: The Structure, Function, and Clinical Prospects of Antibody Fragments, Antibodies. 8 (2019) 28. https://doi.org/10.3390/antib8020028.

[51] A. Cattaneo, M. Chirichella, Targeting the Post-translational Proteome with Intrabodies, Trends Biotechnol. 37 (2019) 578–591. https://doi.org/10.1016/j.tibtech.2018.11.009.

[52] M. Gilodi, S. Lisi, E. F. Dudás, M. Fantini, R. Puglisi, A. Louka, P. Marcatili, A. Cattaneo, A. Pastore, Selection and Modelling of a New Single-Domain Intrabody Against TDP-43, Front. Mol. Biosci. 8 (2022). https://doi.org/10.3389/fmolb.2021.773234.

[53] M. Visintin, G.A. Meli, I. Cannistraci, A. Cattaneo, Intracellular antibodies for proteomics, J. Immunol. Methods. 290 (2004) 135–153. https://doi.org/10.1016/j.jim.2004.04.014.

[54] M. Visintin, G. Settanni, A. Maritan, S. Graziosi, J.D. Marks, A. Cattaneo, The intracellular antibody capture technology (IACT): towards a consensus sequence for intracellular antibodies, J Mol Biol. 317 (2002) 73-83. https://doi.org/10.1006/jmbi.2002.5392.

[55] J. McCafferty, A.D. Griffiths, G. Winter, D.J. Chiswell, Phage antibodies: filamentous phage displaying antibody variable domains, Nature. 348 (1990) 552–554. https://doi.org/10.1038/348552a0.

[56] S. Dübel, O. Stoevesandt, M.J. Taussig, M. Hust, Generating recombinant antibodies to the complete human proteome, Trends Biotechnol. 28 (2010) 333–339. https://doi.org/10.1016/j.tibtech.2010.05.001.

[57] Nobel work that galvanized an industry, Nat. Biotechnol. 36 (2018) 1023–1023.

https://doi.org/10.1038/nbt.4301.

[58] C.F. Barbas, A.S. Kang, R.A. Lerner, S.J. Benkovic, Assembly of combinatorial antibody libraries on phage surfaces: the gene III site., Proc. Natl. Acad. Sci. 88 (1991) 7978–7982. https://doi.org/10.1073/pnas.88.18.7978.

[59] P. Fuchs, F. Breitling, S. Dübel, T. Seehaus, M. Little, Targeting Recombinant Antibodies to the Surface of Escherichia coli: Fusion to a Peptidoglycan Associated Lipoprotein, Bio/Technology. 9 (1991) 1369–1372. https://doi.org/10.1038/nbt1291-1369.

[60] J.E. Butler, P. Navarro, J. Sun, Adsorption-Induced Antigenic Changes and Their Significance in Elisa and Immunological Disorders, Immunol. Invest. 26 (1997) 39–54. https://doi.org/10.3109/08820139709048914.

[61] F. Ferrara, L.A. Naranjo, S. D'Angelo, C. Kiss, A.R.M. Bradbury, Specific binder for Lightning-Link® biotinylated proteins from an antibody phage library, J. Immunol. Methods. 395 (2013) 83–87. https://doi.org/10.1016/j.jim.2013.06.010.

[62] P. Chames, D. Baty, Antibody engineering and its applications in tumor targeting and intracellular immunization, FEMS Microbiol. Lett. 189 (2000) 1–8. https://doi.org/10.1111/j.1574-6968.2000.tb09197.x.

[63] G.P. Smith, Phage Display: Simple Evolution in a Petri Dish (Nobel Lecture), Angew. Chemie Int. Ed. 58 (2019) 14428–14437. https://doi.org/10.1002/anie.201908308.

[64] A. Frenzel, T. Schirrmann, M. Hust, Phage display-derived human antibodies in clinical development and therapy, MAbs. 8 (2016) 1177–1194. https://doi.org/10.1080/19420862.2016.1212149.

[65] J.W. Kehoe, N. Velappan, M. Walbolt, J. Rasmussen, D. King, J. Lou, K. Knopp, P. Pavlik, J.D. Marks, C.R. Bertozzi, A.R.M. Bradbury, Using Phage Display to Select Antibodies Recognizing Post-translational Modifications Independently of Sequence Context, Mol. Cell. Proteomics. 5 (2006) 2350–2363. https://doi.org/10.1074/mcp.M600314-MCP200.

[66] S.A. Gai, K.D. Wittrup, Yeast surface display for protein engineering and characterization, Curr. Opin. Struct. Biol. 17 (2007) 467–473. https://doi.org/10.1016/j.sbi.2007.08.012.

[67] L. Benatuil, J.M. Perez, J. Belk, C.-M. Hsieh, An improved yeast transformation method for the generation of very large human antibody libraries, Protein Eng. Des. Sel. 23 (2010) 155–159. https://doi.org/10.1093/protein/gzq002.

[68] T.C. Scanlon, E.C. Gray, K.E. Griswold, Quantifying and resolving multiple vector transformants in S. cerevisiae plasmid libraries, BMC Biotechnol. 9 (2009) 95. https://doi.org/10.1186/1472-6750-9-95.

[69] N. Velappan, D. Sblattero, L. Chasteen, P. Pavlik, A.R.M. Bradbury, Plasmid incompatibility: more compatible than previously thought?, Protein Eng. Des. Sel. 20

(2007) 309–313. https://doi.org/10.1093/protein/gzm005.

[70] C. Zahnd, P. Amstutz, A. Plückthun, Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target, Nat. Methods. 4 (2007) 269–279. https://doi.org/10.1038/nmeth1003.

[71] P. Piccioli, A. Di Luzio, R. Amann, R. Schuligoi, M.A. Surani, J. Donnerer, A. Cattaneo, Neuroantibodies: Ectopic expression of a recombinant anti-substance P antibody in the central nervous system of transgenic mice, Neuron. 15 (1995) 373–384. https://doi.org/10.1016/0896-6273(95)90041-1.

[72] P. Piccioli, F. Ruberti, S. Biocca, A. Di Luzio, T.M. Werge, A. Bradbury, A. Cattaneo, Neuroantibodies: molecular cloning of a monoclonal antibody against substance P for expression in the central nervous system., Proc. Natl. Acad. Sci. 88 (1991) 5611–5615. https://doi.org/10.1073/pnas.88.13.5611.

[73] S. Capsoni, G. Ugolini, a Comparini, F. Ruberti, N. Berardi, a Cattaneo, Alzheimer-like neurodegeneration in aged antinerve growth factor transgenic mice., Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 6826–6831. https://doi.org/10.1073/pnas.97.12.6826.

[74] E.M. Parzych, J. Du, A.R. Ali, K. Schultheis, D. Frase, T.R.F. Smith, J. Cui, N. Chokkalingam, N.J. Tursi, V.M. Andrade, B.M. Warner, E.N. Gary, Y. Li, J. Choi, J. Eisenhauer, I. Maricic, A. Kulkarni, J.D. Chu, G. Villafana, K. Rosenthal, K. Ren, J.R. Francica, S.K. Wootton, P. Tebas, D. Kobasa, K.E. Broderick, J.D. Boyer, M.T. Esser, J. Pallesen, D.W. Kulp, A. Patel, D.B. Weiner, DNA-delivered antibody cocktail exhibits improved pharmacokinetics and confers prophylactic protection against SARS-CoV-2, Nat. Commun. 13 (2022) 5886. https://doi.org/10.1038/s41467-022-33309-6.

[75] A. Cattaneo, M.S. Neuberger, Polymeric immunoglobulin M is secreted by transfectants of non-lymphoid cells in the absence of immunoglobulin J chain., EMBO J. 6 (1987) 2753–8.
http://www.ncbi.nlm.nih.gov/pubmed/3119328%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC553699.

[76] T. Anelli, R. Sitia, Protein quality control in the early secretory pathway, EMBO J. 27 (2008) 315–327. https://doi.org/10.1038/sj.emboj.7601974.

[77] S. Biocca, M.S. Neuberger, A. Cattaneo, Expression and targeting of intracellular antibodies in mammalian cells., EMBO J. 9 (1990) 101–108. https://doi.org/10.1002/j.1460-2075.1990.tb08085.x.

[78] S. Biocca, P. Pierandrei-Amaldi, A. Cattaneo, Intracellular expression of anti-p21ras single chain Fv fragments inhibits meiotic maturation of xenopus oocytes., Biochem. Biophys. Res. Commun. 197 (1993) 422–7. https://doi.org/10.1006/bbrc.1993.2496.

[79] S. Biocca, P. Pierandrei-Amaldi, N. Campioni, A. Cattaneo, Intracellular immunization with cytosolic recombinant antibodies., Biotechnology. (N. Y). 12 (1994) 396–9.

https://doi.org/10.1038/nbt0494-396.

[80] J. Zhang, T.H. Rabbitts, Intracellular antibody capture: A molecular biology approach to inhibitors of protein–protein interactions, Biochim. Biophys. Acta - Proteins Proteomics. 1844 (2014) 1970–1976. https://doi.org/10.1016/j.bbapap.2014.05.009.

[81] S. Cattaneo, A. Biocca, Intracellular Antibodies, Springer Berlin Heidelberg, Berlin, Heidelberg, 1997. https://doi.org/10.1007/978-3-662-07992-8.

[82] K. Proba, A. Wörn, A. Honegger, A. Plückthun, Antibody scFv fragments without disulfide bonds, made by molecular evolution 1 1Edited by I. A. Wilson, J. Mol. Biol. 275 (1998) 245–253. https://doi.org/10.1006/jmbi.1997.1457.

[83] C.J. Bond, J.C. Marsters, S.S. Sidhu, Contributions of CDR3 to VHH Domain Stability and the Design of Monobody Scaffolds for Naive Antibody Libraries, J. Mol. Biol. 332 (2003) 643–655. https://doi.org/10.1016/S0022-2836(03)00967-7.

[84] S. Biocca, Intrabody Expression in Mammalian Cells, in: 2011: pp. 179–195. https://doi.org/10.1007/978-94-007-1257-7_9.

[85] S. Biocca, F. Ruberti, M. Tafani, P. Pierandrei-Amaldi, a Cattaneo, Redox state of single chain Fv fragments targeted to the endoplasmic reticulum, cytosol and mitochondria., Nat. Biotechnol. 13 (1995) 1110–5. https://doi.org/10.1038/nbt1095-1110.

[86] A. Cattaneo, S. Biocca, The selection of intracellular antibodies, Trends Biotechnol. 17 (1999) 115–121. https://doi.org/10.1016/S0167-7799(98)01268-2.

[87] S. Jung, A. Pluckthun, Improving in vivo folding and stability of a single-chain Fv antibody fragment by loop grafting, Protein Eng. Des. Sel. 10 (1997) 959–966. https://doi.org/10.1093/protein/10.8.959.

[88] W. Zhai, J. Glanville, M. Fuhrmann, L. Mei, I. Ni, P.D. Sundar, T. Van Blarcom, Y. Abdiche, K. Lindquist, R. Strohner, D. Telman, G. Cappuccilli, W.J.J. Finlay, J. Van den Brulle, D.R. Cox, J. Pons, A. Rajpal, Synthetic Antibodies Designed on Natural Sequence Landscapes, J. Mol. Biol. 412 (2011) 55–71. https://doi.org/10.1016/j.jmb.2011.07.018.

[89] H.R. Hoogenboom, G. Winter, By-passing immunisation, J. Mol. Biol. 227 (1992) 381–388. https://doi.org/10.1016/0022-2836(92)90894-P.

[90] A.D. Griffiths, S.C. Williams, O. Hartley, I.M. Tomlinson, P. Waterhouse, W.L. Crosby, R.E. Kontermann, P.T. Jones, N.M. Low, T.J. Allison, Isolation of high affinity human antibodies directly from large synthetic repertoires., EMBO J. 13 (1994) 3245–3260. https://doi.org/10.1002/j.1460-2075.1994.tb06626.x.

[91] M. Visintin, E. Tse, H. Axelson, T.H. Rabbitts, A. Cattaneo, Selection of antibodies for intracellular function using a two-hybrid in vivo system, Proc. Natl. Acad. Sci. 96 (1999) 11723–11728. https://doi.org/10.1073/pnas.96.21.11723.

[92] G. Settanni, J.G. Mainz, M. Visintin, G. Settanni, A. Maritan, S. Graziosi, The intracellular

antibody capture technology (IACT): Towards a consensus sequence for intracellular antibodies The Intracellular Antibody Capture Technology (IACT): Towards a Consensus Sequence for Intracellular Antibodies, (2002). https://doi.org/10.1006/jmbi.2002.5392.

[93] P. Dufner, L. Jermutus, R.R. Minter, Harnessing phage and ribosome display for antibody optimisation, Trends Biotechnol. 24 (2006) 523–529. https://doi.org/10.1016/j.tibtech.2006.09.004.

[94] S. Fields, O. Song, A novel genetic system to detect protein-protein interactions, Nature. 340 (1989) 245–246. https://doi.org/10.1038/340245a0.

[95] M. Visintin, M. Quondam, A. Cattaneo, The intracellular antibody capture technology: Towards the high-throughput selection of functional intracellular antibodies for target validation, Methods. 34 (2004) 200–214. https://doi.org/10.1016/j.ymeth.2004.04.008.

[96] S.L. Dove, A. Hochschild, Conversion of the ω subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target, Genes Dev. 12 (1998) 745–754. https://doi.org/10.1101/gad.12.5.745.

[97] S.L. Dove, J.K. Joung, A. Hochschild, Activation of prokaryotic transcription through arbitrary protein–protein contacts, Nature. 386 (1997) 627–630. https://doi.org/10.1038/386627a0.

[98] P. Secco, E. D'Agostini, R. Marzari, M. Licciulli, R. Di Niro, S. D'Angelo, A.R.M. Bradbury, U. Dianzani, C. Santoro, D. Sblattero, Antibody library selection by the β-lactamase protein fragment complementation assay, Protein Eng. Des. Sel. 22 (2009) 149–158. https://doi.org/10.1093/protein/gzn053.

[99] E. Mössner, H. Koch, A. Plückthun, Fast selection of antibodies without antigen purification: adaptation of the protein fragment complementation assay to select antigen-antibody pairs1 1Edited by I. A. Wilson, J. Mol. Biol. 308 (2001) 115–122. https://doi.org/10.1006/jmbi.2001.4575.

[100] T.K. Kerppola, Bimolecular Fluorescence Complementation (BiFC) Analysis as a Probe of Protein Interactions in Living Cells, Annu. Rev. Biophys. 37 (2008) 465–487. https://doi.org/10.1146/annurev.biophys.37.032807.125842.

[101] A.F. Rudolf, T. Skovgaard, S. Knapp, L.J. Jensen, J. Berthelsen, A Comparison of Protein Kinases Inhibitor Screening Methods Using Both Enzymatic Activity and Binding Affinity Determination, PLoS One. 9 (2014) e98800. https://doi.org/10.1371/journal.pone.0098800.

[102] J. Mehla, J.H. Caufield, P. Uetz, Mapping Protein–Protein Interactions Using Yeast Two-Hybrid Assays, Cold Spring Harb. Protoc. 2015 (2015) pdb.prot086157. https://doi.org/10.1101/pdb.prot086157.

[103] Y. Wang, T. Cui, C. Zhang, M. Yang, Y. Huang, W. Li, L. Zhang, C. Gao, Y. He, Y. Li, F.

Huang, J. Zeng, C. Huang, Q. Yang, Y. Tian, C. Zhao, H. Chen, H. Zhang, Z.-G. He, Global Protein−Protein Interaction Network in the Human Pathogen Mycobacterium tuberculosis H37Rv, J. Proteome Res. 9 (2010) 6665–6677. https://doi.org/10.1021/pr100808n.

[104] B. Valldorf, S.C. Hinz, G. Russo, L. Pekar, L. Mohr, J. Klemm, A. Doerner, S. Krah, M. Hust, S. Zielonka, Antibody display technologies: selecting the cream of the crop, Biol. Chem. 403 (2022) 455–477. https://doi.org/10.1515/hsz-2020-0377.

[105] D. Guo, T.R. Hazbun, X.-J. Xu, S.-L. Ng, S. Fields, M.-H. Kuo, A tethered catalysis, two-hybrid system to identify protein-protein interactions requiring post-translational modifications., Nat. Biotechnol. 22 (2004) 888–892. https://doi.org/10.1038/nbt985.

[106] J.T. Koerber, N.D. Thomsen, B.T. Hannigan, W.F. Degrado, J.A. Wells, Nature-inspired design of motif-specific antibody scaffolds, Nat. Biotechnol. 31 (2013) 916–921. https://doi.org/10.1038/nbt.2672.

[107] K.-L. Guan, W. Yu, Y. Lin, Y. Xiong, S. Zhao, Generation of acetyllysine antibodies and affinity enrichment of acetylated peptides., Nat. Protoc. 5 (2010) 1583–1595. https://doi.org/10.1038/nprot.2010.117.

[108] T. Hattori, J.M. Taft, K.M. Swist, H. Luo, H. Witt, M. Slattery, A. Koide, A.J. Ruthenburg, K. Krajewski, B.D. Strahl, K.P. White, P.J. Farnham, Y. Zhao, S. Koide, Recombinant antibodies to histone post-translational modifications, Nat. Methods. 10 (2013) 992–995. https://doi.org/10.1038/nmeth.2605.

[109] E. Contreras-Leal, A. Hernández-Oliveras, L. Flores-Peredo, Á. Zarain-Herzberg, J. Santiago-García, Histone deacetylase inhibitors promote the expression of ATP2A3 gene in breast cancer cell lines, Mol. Carcinog. 55 (2016) 1477–1485. https://doi.org/10.1002/mc.22402.

[110] O.Y. Dmitriev, S. Lutsenko, S. Muyldermans, Nanobodies as Probes for Protein Dynamics in Vitro and in Cells, J. Biol. Chem. 291 (2016) 3767–3775. https://doi.org/10.1074/jbc.R115.679811.

[111] C. Schwab, A. Twardek, H.R. Bosshard, T.P. Lo, G.D. Brayer, Mapping antibody binding sites on cytochrome c with synthetic peptides: Are results representative of the antigenic structure of proteins?, Protein Sci. 2 (1993) 175–182. https://doi.org/10.1002/pro.5560020206.

[112] S. Lisi, M. Trovato, O. Vitaloni, M. Fantini, M. Chirichella, P. Tognini, S. Cornuti, M. Costa, M. Groth, A. Cattaneo, Acetylation-Specific Interference by Anti-Histone H3K9ac Intrabody Results in Precise Modulation of Gene Expression, Int. J. Mol. Sci. 23 (2022) 8892. https://doi.org/10.3390/ijms23168892.

[113] D. de la Torre, J.W. Chin, Reprogramming the genetic code, Nat. Rev. Genet. 22 (2021) 169–184. https://doi.org/10.1038/s41576-020-00307-7.

[114] J.W. Chin, Expanding and reprogramming the genetic code, Nature. 550 (2017) 53–60. https://doi.org/10.1038/nature24031.

[115] M.A. Shandell, Z. Tan, V.W. Cornish, Genetic Code Expansion: A Brief History and Perspective, (2021). https://doi.org/10.1021/acs.biochem.1c00286.

[116] L. Wang, J. Xie, P.G. Schultz, Expanding the Genetic Code, Annu. Rev. Biophys. Biomol. Struct. 35 (2006) 225–249. https://doi.org/10.1146/annurev.biophys.35.101105.121507.

[117] J.W. Chin, T.A. Cropp, J.C. Anderson, M. Mukherji, Z. Zhang, P.G. Schultz, An expanded eukaryotic genetic code, Science (80-.). 301 (2003) 964–967. https://doi.org/10.1126/science.1084772.

[118] D.D. Young, P.G. Schultz, Playing with the molecules of life, ACS Chem. Biol. 13 (2018) 854–870. https://doi.org/10.1021/acschembio.7b00974.

[119] C.C. Liu, P.G. Schultz, Adding New Chemistries to the Genetic Code, Annu. Rev. Biochem. 79 (2010) 413–444. https://doi.org/10.1146/annurev.biochem.052308.105824.

[120] J.W. Chin, T.A. Cropp, S. Chu, E. Meggers, P.G. Schultz, Progress Toward an Expanded Eukaryotic Genetic Code, Chem. Biol. 10 (2003) 511–519. https://doi.org/10.1016/S1074-5521(03)00123-6.

[121] J.W. Chin, Expanding and Reprogramming the Genetic Code of Cells and Animals, Annu. Rev. Biochem. 83 (2014) 379–408. https://doi.org/10.1146/annurev-biochem-060713-035737.

[122] W. Liu, A. Brock, S. Chen, P.G. Schultz, Genetic incorporation of unnatural amino acids into proteins in mammalian cells, Nat Methods. 4 (2007) 239–244. https://doi.org/10.1038/nmeth1016.

[123] H. Neumann, S.Y. Peak-Chew, J.W. Chin, Genetically encoding N(epsilon)-acetyllysine in recombinant proteins., Nat. Chem. Biol. 4 (2008) 232–234. https://doi.org/10.1038/nchembio.73.

[124] H. Neumann, S.M. Hancock, R. Buning, A. Routh, L. Chapman, J. Somers, T. Owen-Hughes, J. van Noort, D. Rhodes, J.W. Chin, A Method for Genetically Installing Site-Specific Acetylation in Recombinant Histones Defines the Effects of H3 K56 Acetylation, Mol. Cell. 36 (2009) 153–163. https://doi.org/10.1016/j.molcel.2009.07.027.

[125] S. Hancock, Expanding the Genetic Code of Yeast via a Pyrrolysl- tRNA synthetase/tRNA Pair, 426 (n.d.) 1–10.

[126] J.W.B. Hershey, N. Sonenberg, M.B. Mathews, Principles of Translational Control: An Overview, Cold Spring Harb. Perspect. Biol. 4 (2012) a011528–a011528. https://doi.org/10.1101/cshperspect.a011528.

[127] A. Böck, K. Forchhammer, J. Heider, W. Leinfelder, G. Sawers, B. Veprek, F. Zinoni, Selenocysteine: the 21st amino acid, Mol. Microbiol. 5 (1991) 515–520.

https://doi.org/10.1111/j.1365-2958.1991.tb00722.x.

[128] G. Srinivasan, C.M. James, J.A. Krzycki, Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA, Science (80-.). 296 (2002) 1459–1462. https://doi.org/10.1126/science.1069588.

[129] C. Polycarpo, A. Ambrogelly, A. Bérubé, S.M. Winbush, J.A. McCloskey, P.F. Crain, J.L. Wood, D. Söll, An aminoacyl-tRNA synthetase that specifically activates pyrrolysine., Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 12450–4. https://doi.org/10.1073/pnas.0405362101.

[130] A. Ambrogelly, S. Palioura, D. Söll, Natural expansion of the genetic code, Nat. Chem. Biol. 3 (2007) 29–35. https://doi.org/10.1038/nchembio847.

[131] R.D. Knight, S.J. Freeland, L.F. Landweber, Rewiring the keyboard: evolvability of the genetic code, Nat. Rev. Genet. 2 (2001) 49–58. https://doi.org/10.1038/35047500.

[132] E. V. Koonin, A.S. Novozhilov, Origin and evolution of the genetic code: The universal enigma, IUBMB Life. 61 (2009) 99–111. https://doi.org/10.1002/iub.146.

[133] B. Hao, W. Gong, T.K. Ferguson, C.M. James, J.A. Krzycki, M.K. Chan, A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase, Science (80-.). 296 (2002) 1462–1466. https://doi.org/10.1126/science.1069556.

[134] L. Prat, I.U. Heinemann, H.R. Aerni, J. Rinehart, P. O'Donoghue, D. Söll, Carbon source-dependent expansion of the genetic code in bacteria, Proc. Natl. Acad. Sci. 109 (2012) 21070–21075. https://doi.org/10.1073/pnas.1218613110.

[135] S.M. Heaphy, M. Mariotti, V.N. Gladyshev, J.F. Atkins, P. V. Baranov, Novel Ciliate Genetic Code Variants Including the Reassignment of All Three Stop Codons to Sense Codons in Condylostoma magnum, Mol. Biol. Evol. 33 (2016) 2885–2889. https://doi.org/10.1093/molbev/msw166.

[136] V. Beránek, C.D. Reinkemeier, M.S. Zhang, A.D. Liang, G. Kym, J.W. Chin, Genetically Encoded Protein Phosphorylation in Mammalian Cells, Cell Chem. Biol. 25 (2018) 1067-1074.e5. https://doi.org/10.1016/j.chembiol.2018.05.013.

[137] Y. Chen, J. Ma, W. Lu, M. Tian, M. Thauvin, C. Yuan, M. Volovitch, Q. Wang, J. Holst, M. Liu, S. Vriz, S. Ye, L. Wang, D. Li, Heritable expansion of the genetic code in mouse and zebrafish, Cell Res. 27 (2017) 294–297. https://doi.org/10.1038/cr.2016.145.

[138] H. Liu, L. Wang, A. Brock, C.-H. Wong, P.G. Schultz, A Method for the Generation of Glycoprotein Mimetics, J. Am. Chem. Soc. 125 (2003) 1702–1703. https://doi.org/10.1021/ja029433n.

[139] J. Xie, P.G. Schultz, A chemical toolkit for proteins--an expanded genetic code., Nat. Rev. Mol. Cell Biol. 7 (2006) 775–782. https://doi.org/10.1038/nrm2005.

[140] L. Davis, J.W. Chin, Designer proteins: applications of genetic code expansion in cell

biology, Nat. Rev. Mol. Cell Biol. 13 (2012) 168–182. https://doi.org/10.1038/nrm3286.

[141] L. Wang, a Brock, B. Herberich, P.G. Schultz, Expanding the genetic code of Escherichia coli., Science. 292 (2001) 498–500. https://doi.org/10.1126/science.1060077.

[142] L. Wang, Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts, Acc. Chem. Res. 50 (2017) 2767–2775. https://doi.org/10.1021/acs.accounts.7b00376.

[143] A. Deiters, T.A. Cropp, M. Mukherji, J.W. Chin, J.C. Anderson, P.G. Schultz, Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae, J. Am. Chem. Soc. 125 (2003) 11782–11783. https://doi.org/10.1021/ja0370037.

[144] J. Fredens, K. Wang, D. de la Torre, L.F.H. Funke, W.E. Robertson, Y. Christova, T. Chia, W.H. Schmied, D.L. Dunkelmann, V. Beránek, C. Uttamapinant, A.G. Llamazares, T.S. Elliott, J.W. Chin, Total synthesis of Escherichia coli with a recoded genome, Nature. 569 (2019) 514–518. https://doi.org/10.1038/s41586-019-1192-5.

[145] D.T. Rogerson, A. Sachdeva, K. Wang, T. Haq, A. Kazlauskaite, S.M. Hancock, N. Huguenin-Dezot, M.M.K. Muqit, A.M. Fry, R. Bayliss, J.W. Chin, Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog, Nat. Chem. Biol. 11 (2015). https://doi.org/10.1038/nchembio.1823.

[146] T.M. Wannier, A.M. Kunjapur, D.P. Rice, M.J. McDonald, M.M. Desai, G.M. Church, Adaptive evolution of genomically recoded Escherichia coli, Proc. Natl. Acad. Sci. 115 (2018) 3090–3095. https://doi.org/10.1073/pnas.1715530115.

[147] J. Normanly, L.G. Kleina, J.-M. Masson, J. Abelson, J.H. Miller, Construction of Escherichia coli amber suppressor tRNA genes, J. Mol. Biol. 213 (1990) 719–726. https://doi.org/10.1016/S0022-2836(05)80258-X.

[148] Q. Wang, T. Sun, J. Xu, Z. Shen, S.P. Briggs, D. Zhou, L. Wang, Response and Adaptation of Escherichia coli to Suppression of the Amber Stop Codon, ChemBioChem. 15 (2014) 1744–1749. https://doi.org/10.1002/cbic.201402235.

[149] S.M. Hancock, R. Uprety, A. Deiters, J.W. Chin, 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 (2010) 14819–14824. https://doi.org/10.1021/ja104609m.

[150] O. Vargas-Rodriguez, A. Sevostyanova, D. Söll, A. Crnković, Upgrading aminoacyl-tRNA synthetases for genetic code expansion, Curr. Opin. Chem. Biol. 46 (2018) 115–122. https://doi.org/10.1016/j.cbpa.2018.07.014.

[151] R. Giege, M. Sissler, C. Florentz, Universal rules and idiosyncratic features in tRNA identity, Nucleic Acids Res. 26 (1998) 5017–5035. https://doi.org/10.1093/nar/26.22.5017.

[152] K. Nozawa, P. O'Donoghue, S. Gundllapalli, Y. Araiso, R. Ishitani, T. Umehara, D. Söll, O. Nureki, Pyrrolysyl-tRNA synthetase–tRNAPyl structure reveals the molecular basis

of orthogonality, Nature. 457 (2009) 1163–1167. https://doi.org/10.1038/nature07611.

[153] T. Suzuki, C. Miller, L.-T. Guo, J.M.L. Ho, D.I. Bryson, Y.-S. Wang, D.R. Liu, D. Söll, Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase, Nat. Chem. Biol. 13 (2017) 1261–1266. https://doi.org/10.1038/nchembio.2497.

[154] J. Xie, P.G. Schultz, An expanding genetic code, Methods. 36 (2005) 227–238. https://doi.org/10.1016/j.ymeth.2005.04.010.

[155] M.S. Packer, D.R. Liu, Methods for the directed evolution of proteins, Nat. Rev. Genet. 16 (2015) 379–394. https://doi.org/10.1038/nrg3927.

[156] D.R. Liu, P.G. Schultz, Progress toward the evolution of an organism with an expanded genetic code, Proc. Natl. Acad. Sci. 96 (1999) 4780–4785. https://doi.org/10.1073/pnas.96.9.4780.

[157] D.I. Bryson, C. Fan, L.-T. Guo, C. Miller, D. Söll, D.R. Liu, Continuous directed evolution of aminoacyl-tRNA synthetases, Nat. Chem. Biol. 13 (2017) 1253–1260. https://doi.org/10.1038/nchembio.2474.

[158] A.D. Haimovich, P. Muir, F.J. Isaacs, Genomes by design, Nat. Rev. Genet. 16 (2015) 501–516. https://doi.org/10.1038/nrg3956.

[159] A.M. Kunjapur, D.A. Stork, E. Kuru, O. Vargas-Rodriguez, M. Landon, D. Söll, G.M. Church, Engineering posttranslational proofreading to discriminate nonstandard amino acids, Proc. Natl. Acad. Sci. 115 (2018) 619–624. https://doi.org/10.1073/pnas.1715137115.

[160] K. Haruna, M.H. Alkazemi, Y. Liu, D. Söll, M. Englert, Engineering the elongation factor Tu for efficient selenoprotein synthesis, Nucleic Acids Res. 42 (2014) 9976–9983. https://doi.org/10.1093/nar/gku691.

[161] N. Manickam, K. Joshi, M.J. Bhatt, P.J. Farabaugh, Effects of tRNA modification on translational accuracy depend on intrinsic codon–anticodon strength, Nucleic Acids Res. 44 (2016) 1871–1881. https://doi.org/10.1093/nar/gkv1506.

[162] J.T. Stieglitz, J.A. Van Deventer, High-Throughput Aminoacyl-tRNA Synthetase Engineering for Genetic Code Expansion in Yeast, ACS Synth. Biol. 11 (2022) 2284–2299. https://doi.org/10.1021/acssynbio.1c00626.

[163] J. Nielsen, J.D. Keasling, Engineering Cellular Metabolism, Cell. 164 (2016) 1185–1197. https://doi.org/10.1016/j.cell.2016.02.004.

[164] M.S. Siddiqui, K. Thodey, I. Trenchard, C.D. Smolke, Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools, FEMS Yeast Res. 12 (2012) 144–170. https://doi.org/10.1111/j.1567-1364.2011.00774.x.

[165] C. Fan, K. Ip, D. Söll, Expanding the genetic code of Escherichia coli with phosphotyrosine, FEBS Lett. (2016) 3040–3047. https://doi.org/10.1002/1873-3468.12333.

[166] T. Mukai, M.J. Lajoie, M. Englert, D. Söll, Rewriting the Genetic Code, Annu. Rev. Microbiol. 71 (2017) 557–577. https://doi.org/10.1146/annurev-micro-090816-093247.

[167] L. Wang, P.G. Schultz, A general approach for the generation of orthogonal tRNAs, Chem. Biol. 8 (2001) 883–890. https://doi.org/10.1016/S1074-5521(01)00063-1.

[168] F. Li, H. Zhang, Y. Sun, Y. Pan, J. Zhou, J. Wang, Expanding the Genetic Code for Photoclick Chemistry in E. coli , Mammalian Cells, and A. thaliana, Angew. Chemie Int. Ed. 52 (2013) 9700–9704. https://doi.org/10.1002/anie.201303477.

[169] Q. Gan, B.P. Lehman, T.A. Bobik, C. Fan, Expanding the genetic code of Salmonella with non-canonical amino acids, Sci. Rep. 6 (2016) 39920. https://doi.org/10.1038/srep39920.

[170] S.J. Elsässer, R.J. Ernst, O.S. Walker, J.W. Chin, Genetic code expansion in stable cell lines enables encoded chromatin modification., Nat. Methods. 13 (2016) 158–64. https://doi.org/10.1038/nmeth.3701.

[171] R.J. Ernst, T.P. Krogager, E.S. Maywood, R. Zanchi, V. Beránek, T.S. Elliott, N.P. Barry, M.H. Hastings, J.W. Chin, Genetic code expansion in the mouse brain, Nat. Chem. Biol. 12 (2016) 776–778. https://doi.org/10.1038/nchembio.2160.

[172] T. Umehara, J. Kim, S. Lee, L.-T. Guo, D. Söll, H.-S. Park, N -Acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N -acetyl lysine specificity in vitro and in vivo, FEBS Lett. 586 (2012) 729–733. https://doi.org/10.1016/j.febslet.2012.01.029.

[173] J.C.W. Willis, J.W. Chin, Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs, Nat. Chem. 10 (2018). https://doi.org/10.1038/s41557-018-0052-5.

[174] W. Wan, J.M. Tharp, W.R. Liu, Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool, Biochim. Biophys. Acta - Proteins Proteomics. 1844 (2014) 1059–1070. https://doi.org/10.1016/j.bbapap.2014.03.002.

[175] D.L. Dunkelmann, J.C.W. Willis, A.T. Beattie, J.W. Chin, Engineered triply orthogonal pyrrolysyl–tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids, Nat. Chem. 12 (2020) 535–544. https://doi.org/10.1038/s41557-020-0472-x.

[176] H.-S. Park, M.J. Hohn, T. Umehara, L.-T. Guo, E.M. Osborne, J. Benner, C.J. Noren, J. Rinehart, D. Söll, Expanding the Genetic Code of Escherichia coli with Phosphoserine, Science (80-.). 333 (2011) 1151–1154. https://doi.org/10.1126/science.1207203.

[177] A. Sauerwald, W. Zhu, T.A. Major, H. Roy, S. Palioura, D. Jahn, W.B. Whitman, J.R. Yates, M. Ibba, D. Söll, RNA-Dependent Cysteine Biosynthesis in Archaea, Science (80-.). 307 (2005) 1969–1972. https://doi.org/10.1126/science.1108329.

[178] R. Fukunaga, S. Yokoyama, Structural insights into the first step of RNA-dependent cysteine biosynthesis in archaea, Nat. Struct. Mol. Biol. 14 (2007) 272–279. https://doi.org/10.1038/nsmb1219.

[179] H. Edwards, P. Schimmel, A bacterial amber suppressor in Saccharomyces cerevisiae is selectively recognized by a bacterial aminoacyl-tRNA synthetase, Mol. Cell. Biol. 10 (1990) 1633–1641. https://doi.org/10.1128/mcb.10.4.1633-1641.1990.

[180] N. Shao, N.S. Singh, S.E. Slade, A.M.E. Jones, M.K. Balasubramanian, Site Specific Genetic Incorporation of Azidophenylalanine in Schizosaccharomyces pombe, Sci. Rep. 5 (2015) 17196. https://doi.org/10.1038/srep17196.

[181] T.S. Young, I. Ahmad, J.A. Yin, P.G. Schultz, An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli, J. Mol. Biol. 395 (2010) 361–374. https://doi.org/10.1016/j.jmb.2009.10.030.

[182] M. Pott, M.J. Schmidt, D. Summerer, Evolved Sequence Contexts for Highly Efficient Amber Suppression with Noncanonical Amino Acids, ACS Chem. Biol. 9 (2014) 2815–2822. https://doi.org/10.1021/cb5006273.

[183] Q. Wang, A.R. Parrish, L. Wang, Review Expanding the Genetic Code for Biological Studies, Chem. Biol. 16 (2009) 323–336. https://doi.org/10.1016/j.chembiol.2009.03.001.

[184] Y. Chemla, E. Ozer, I. Algov, L. Alfonta, Context effects of genetic code expansion by stop codon suppression, Curr. Opin. Chem. Biol. 46 (2018) 146–155. https://doi.org/10.1016/j.cbpa.2018.07.012.

[185] M.Y. Pavlov, R.E. Watts, Z. Tan, V.W. Cornish, M. Ehrenberg, A.C. Forster, Slow peptide bond formation by proline and other N -alkylamino acids in translation, Proc. Natl. Acad. Sci. 106 (2009) 50–54. https://doi.org/10.1073/pnas.0809211106.

[186] M.A. Sørensen, S. Pedersen, Absolute in vivo translation rates of individual codons in Escherichia coli, J. Mol. Biol. 222 (1991) 265–280. https://doi.org/10.1016/0022-2836(91)90211-N.

[187] I. Wohlgemuth, S. Brenner, M. Beringer, M. V. Rodnina, Modulation of the Rate of Peptidyl Transfer on the Ribosome by the Nature of Substrates, J. Biol. Chem. 283 (2008) 32229–32235. https://doi.org/10.1074/jbc.M805316200.

[188] R. Gan, J.G. Perez, E.D. Carlson, I. Ntai, F.J. Isaacs, N.L. Kelleher, M.C. Jewett, Translation system engineering in Escherichia coli enhances non-canonical amino acid incorporation into proteins, Biotechnol. Bioeng. 114 (2017) 1074–1086. https://doi.org/10.1002/bit.26239.

[189] A. Chatterjee, S.B. Sun, J.L. Furman, H. Xiao, P.G. Schultz, A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli, Biochemistry. 52 (2013) 1828–1837. https://doi.org/10.1021/bi4000244.

[190] Y. Zheng, T.L. Lewis, P. Igo, F. Polleux, A. Chatterjee, Virus-Enabled Optimization and Delivery of the Genetic Machinery for Efficient Unnatural Amino Acid Mutagenesis in Mammalian Cells and Tissues, ACS Synth. Biol. 6 (2017) 13–18. https://doi.org/10.1021/acssynbio.6b00092.

[191] T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, S. Yokoyama, Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases, Biochem. Biophys. Res. Commun. 371 (2008) 818–822. https://doi.org/10.1016/j.bbrc.2008.04.164.

[192] K. Wang, H. Neumann, S.Y. Peak-Chew, J.W. Chin, Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion, Nat. Biotechnol. 25 (2007) 770–777. https://doi.org/10.1038/nbt1314.

[193] X. Fu, Y. Huang, Y. Shen, Improving the Efficiency and Orthogonality of Genetic Code Expansion, BioDesign Res. 2022 (2022) 1–13. https://doi.org/10.34133/2022/9896125.

[194] K. Mangano, T. Florin, X. Shao, D. Klepacki, I. Chelysheva, Z. Ignatova, Y. Gao, A.S. Mankin, N. Vázquez-Laslop, Genome-wide effects of the antimicrobial peptide apidaecin on translation termination in bacteria, Elife. 9 (2020). https://doi.org/10.7554/eLife.62655.

[195] M.J. Lajoie, A.J. Rovner, D.B. Goodman, H.-R. Aerni, A.D. Haimovich, G. Kuznetsov, J.A. Mercer, H.H. Wang, P.A. Carr, J.A. Mosberg, N. Rohland, P.G. Schultz, J.M. Jacobson, J. Rinehart, G.M. Church, F.J. Isaacs, Genomically Recoded Organisms Expand Biological Functions, Science (80-.). 342 (2013) 357–360. https://doi.org/10.1126/science.1241459.

[196] I.-L. Wu, M.A. Patterson, H.E. Carpenter Desai, R.A. Mehl, G. Giorgi, V.P. Conticello, Multiple Site-Selective Insertions of Noncanonical Amino Acids into Sequence-Repetitive Polypeptides, ChemBioChem. 14 (2013) 968–978. https://doi.org/10.1002/cbic.201300069.

[197] W.H. Schmied, S.J. Elsässer, C. Uttamapinant, J.W. Chin, Efficient Multisite Unnatural Amino Acid Incorporation in Mammalian Cells via Optimized Pyrrolysyl tRNA Synthetase/tRNA Expression and Engineered eRF1, J. Am. Chem. Soc. 136 (2014) 15577–15583. https://doi.org/10.1021/ja5069728.

[198] Q. Wang, L. Wang, Genetic Incorporation of Unnatural Amino Acids into Proteins in Yeast, in: 2012: pp. 199–213. https://doi.org/10.1007/978-1-61779-331-8_12.

[199] Q. Wang, L. Wang, New Methods Enabling Efficient Incorporation of Unnatural Amino Acids in Yeast, J. Am. Chem. Soc. 130 (2008) 6066–6067. https://doi.org/10.1021/ja800894n.

[200] G. Galli, H. Hofstetter, M.L. Birnstiel, Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements, Nature. 294 (1981) 626–631. https://doi.org/10.1038/294626a0.

[201] K. Sakamoto, Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells, Nucleic Acids Res. 30 (2002) 4692–4699. https://doi.org/10.1093/nar/gkf589.

[202] T. Zhang, J. Lei, H. Yang, K. Xu, R. Wang, Z. Zhang, An improved method for whole protein extraction from yeast Saccharomyces cerevisiae, Yeast. 28 (2011) 795–798. https://doi.org/10.1002/yea.1905.

[203] S. Chen, P.G. Schultz, A. Brock, An Improved System for the Generation and Analysis of Mutant Proteins Containing Unnatural Amino Acids in Saccharomyces cerevisiae, J. Mol. Biol. 371 (2007) 112–122. https://doi.org/10.1016/j.jmb.2007.05.017.

[204] M.A. Francis, U.L. Rajbhandary, Expression and function of a human initiator tRNA gene in the yeast Saccharomyces cerevisiae., Mol. Cell. Biol. 10 (1990) 4486–4494. https://doi.org/10.1128/MCB.10.9.4486.

[205] N. Amrani, M.S. Sachs, A. Jacobson, Early nonsense: mRNA decay solves a translational problem, Nat. Rev. Mol. Cell Biol. 7 (2006) 415–425. https://doi.org/10.1038/nrm1942.

[206] S. Meaux, A. van Hoof, K.E. Baker, Nonsense-Mediated mRNA Decay in Yeast Does Not Require PAB1 or a Poly(A) Tail, Mol. Cell. 29 (2008) 134–140. https://doi.org/10.1016/j.molcel.2007.10.031.

[207] C.I. GONZALEZ, W. WANG, S.W. PELTZ, Nonsense-mediated mRNA Decay in Saccharomyces cerevisiae: A Quality Control Mechanism That Degrades Transcripts Harboring Premature Termination Codons, Cold Spring Harb. Symp. Quant. Biol. 66 (2001) 321–328. https://doi.org/10.1101/sqb.2001.66.321.

[208] M.R. Culbertson, P.F. Leeds, Looking at mRNA decay pathways through the window of molecular evolution, Curr. Opin. Genet. Dev. 13 (2003) 207–214. https://doi.org/10.1016/S0959-437X(03)00014-5.

[209] K. Czaplinski, M.J. Ruiz-Echevarria, C.I. González, S.W. Peltz, Should we kill the messenger? The role of the surveillance complex in translation termination and mRNA turnover, BioEssays. 21 (1999) 685–696. https://doi.org/10.1002/(SICI)1521-1878(199908)21:8<685::AID-BIES8>3.0.CO;2-4.

[210] D. MUHLRAD, R. PARKER, Aberrant mRNAs with extended 3′ UTRs are substrates for rapid degradation by mRNA surveillance, RNA. 5 (1999) S1355838299990829. https://doi.org/10.1017/S1355838299990829.

[211] F. He, S.W. Peltz, J.L. Donahue, M. Rosbash, A. Jacobson, Stabilization and ribosome association of unspliced pre-mRNAs in a yeast upf1- mutant, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 7034–7038. https://doi.org/10.1073/pnas.90.15.7034.

[212] H. Tang, P. Zhang, X. Luo, Recent Technologies for Genetic Code Expansion and their Implications on Synthetic Biology Applications, J. Mol. Biol. 434 (2022) 167382. https://doi.org/10.1016/j.jmb.2021.167382.

[213] K.C. Hsu, C.Y. Liu, T.E. Lin, J.H. Hsieh, T.Y. Sung, H.J. Tseng, J.M. Yang, W.J. Huang, Novel Class IIa-Selective Histone Deacetylase Inhibitors Discovered Using an in Silico Virtual Screening Approach, Sci. Rep. 7 (2017). https://doi.org/10.1038/s41598-017-03417-1.

[214] D. Cervettini, S. Tang, S.D. Fried, J.C.W. Willis, L.F.H. Funke, L.J. Colwell, J.W. Chin, Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase–tRNA pairs,

Nat. Biotechnol. (2020). https://doi.org/10.1038/s41587-020-0479-2.

[215] M. Pellis, E. Pardon, K. Zolghadr, U. Rothbauer, C. Vincke, J. Kinne, I. Dierynck, K. Hertogs, H. Leonhardt, J. Messens, S. Muyldermans, K. Conrath, A bacterial-two-hybrid selection system for one-step isolation of intracellularly functional Nanobodies, Arch. Biochem. Biophys. 526 (2012) 114–123. https://doi.org/10.1016/j.abb.2012.04.023.

[216] M. Pellis, S. Muyldermans, C. Vincke, Bacterial two hybrid: A versatile one-step intracellular selection method, Methods Mol. Biol. 911 (2012) 135–150. https://doi.org/10.1007/978-1-61779-968-6_9.

[217] T. Kouzarides, Chromatin Modifications and Their Function, Cell. 128 (2007) 693–705. https://doi.org/10.1016/j.cell.2007.02.005.

[218] S. Sharma, T.K. Kelly, P.A. Jones, Epigenetics in cancer, Carcinogenesis. 31 (2009) 27–36. https://doi.org/10.1093/carcin/bgp220.

[219] T.G. Gillette, J.A. Hill, Readers, writers, and erasers: Chromatin as the whiteboard of heart disease, Circ. Res. 116 (2015) 1245–1253. https://doi.org/10.1161/CIRCRESAHA.116.303630.

[220] A.B. Kunnumakkara, D. Bordoloi, C. Harsha, K. Banik, S.C. Gupta, B.B. Aggarwal, Curcumin mediates anticancer effects by modulating multiple cell signaling pathways, Clin. Sci. 131 (2017) 1781–1799. https://doi.org/10.1042/CS20160935.

[221] S. Reuter, S.C. Gupta, B. Park, A. Goel, B.B. Aggarwal, Epigenetic changes induced by curcumin and other natural compounds, Genes Nutr. 6 (2011) 93–108. https://doi.org/10.1007/s12263-011-0222-1.

[222] F. Chimenti, B. Bizzarri, E. Maccioni, D. Secci, A. Bolasco, P. Chimenti, R. Fioravanti, A. Granese, S. Carradori, F. Tosi, P. Ballario, S. Vernarecci, P. Filetici, A Novel Histone Acetyltransferase Inhibitor Modulating Gcn5 Network: Cyclopentylidene-[4-(4′-chlorophenyl)thiazol-2-yl)hydrazone, J. Med. Chem. 52 (2009) 530–536. https://doi.org/10.1021/jm800885d.

[223] Q. Jin, L.-R. Yu, L. Wang, Z. Zhang, L.H. Kasper, J.-E. Lee, C. Wang, P.K. Brindle, S.Y.R. Dent, K. Ge, Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation, EMBO J. 30 (2011) 249–262. https://doi.org/10.1038/emboj.2010.318.

[224] S.M. Hollenberg, R. Sternglanz, P.F. Cheng, H. Weintraub, Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system., Mol. Cell. Biol. 15 (1995) 3813–3822.

[225] D. Gietz, A. St. Jean, R.A. Woods, R.H. Schiestl, Improved method for high efficiency transformation of intact yeast cells, Nucleic Acids Res. 20 (1992) 1425–1425. https://doi.org/10.1093/nar/20.6.1425.

[226] B.D. Strahl, C.D. Allis, The language of covalent histone modifications, Nature. 403 (2000) 41–45. https://doi.org/10.1038/47412.

[227] F.A. Verza, U. Das, A.L. Fachin, J.R. Dimmock, M. Marins, Roles of Histone Deacetylases and Inhibitors in Anticancer Therapy, Cancers (Basel). 12 (2020) 1664. https://doi.org/10.3390/cancers12061664.

[228] M. Hatamipour, T.P. Johnston, A. Sahebkar, One Molecule, Many Targets and Numerous Effects: The Pleiotropy of Curcumin Lies in its Chemical Structure, Curr. Pharm. Des. 24 (2018) 2129–2136. https://doi.org/10.2174/1381612824666180522111036.

[229] H. Huang, A.M. Maertens, E.M. Hyland, J. Dai, A. Norris, J.D. Boeke, J.S. Bader, HistoneHits: A database for histone mutations and their phenotypes, Genome Res. 19 (2009) 674–681. https://doi.org/10.1101/gr.083402.108.

[230] L. Cai, B.M. Sutter, B. Li, B.P. Tu, Acetyl-CoA Induces Cell Growth and Proliferation by Promoting the Acetylation of Histones at Growth Genes, Mol. Cell. 42 (2011) 426–437. https://doi.org/10.1016/j.molcel.2011.05.004.

[231] L. Cai, M.A. McCormick, B.K. Kennedy, B.P. Tu, Integration of Multiple Nutrient Cues and Regulation of Lifespan by Ribosomal Transcription Factor Ifh1, Cell Rep. 4 (2013) 1063–1071. https://doi.org/10.1016/j.celrep.2013.08.016.

[232] G. Bora-Tatar, D. Dayangaç-Erden, A.S. Demir, S. Dalkara, K. Yelekçi, H. Erdem-Yurter, Molecular modifications on carboxylic acid derivatives as potent histone deacetylase inhibitors: Activity and docking studies, Bioorg. Med. Chem. 17 (2009) 5219–5228. https://doi.org/10.1016/j.bmc.2009.05.042.

[233] Z. Liu, Z. Xie, W. Jones, R.E. Pavlovicz, S. Liu, J. Yu, P. Li, J. Lin, J.R. Fuchs, G. Marcucci, C. Li, K.K. Chan, Curcumin is a potent DNA hypomethylation agent, Bioorg. Med. Chem. Lett. 19 (2009) 706–709. https://doi.org/10.1016/j.bmcl.2008.12.041.

[234] M. Sun, Z. Estrov, Y. Ji, K.R. Coombes, D.H. Harris, R. Kurzrock, Curcumin (diferuloylmethane) alters the expression profiles of microRNAs in human pancreatic cancer cells, Mol. Cancer Ther. 7 (2008) 464–473. https://doi.org/10.1158/1535-7163.MCT-07-2272.

[235] J. Zhang, T. Zhang, X. Ti, J. Shi, C. Wu, X. Ren, H. Yin, Curcumin promotes apoptosis in A549/DDP multidrug-resistant human lung adenocarcinoma cells through an miRNA signaling pathway, Biochem. Biophys. Res. Commun. 399 (2010) 1–6. https://doi.org/10.1016/j.bbrc.2010.07.013.

[236] F.J. Dekker, T. van den Bosch, N.I. Martin, Small molecule inhibitors of histone acetyltransferases and deacetylases are potential drugs for inflammatory diseases, Drug Discov. Today. 19 (2014) 654–660. https://doi.org/10.1016/j.drudis.2013.11.012.

[237] B. Mutlu, P. Puigserver, GCN5 acetyltransferase in cellular energetic and metabolic processes, Biochim. Biophys. Acta - Gene Regul. Mech. 1864 (2021) 194626.

https://doi.org/10.1016/j.bbagrm.2020.194626.

[238] L. Chen, W. Fischle, E. Verdin, W.C. Greene, Duration of Nuclear NF-κB Action Regulated by Reversible Acetylation, Science (80-.). 293 (2001) 1653–1657. https://doi.org/10.1126/science.1062374.

[239] W.-C. Hsieh, B.M. Sutter, H. Ruess, S.D. Barnes, V.S. Malladi, B.P. Tu, Glucose starvation induces a switch in the histone acetylome for activation of gluconeogenic and fat metabolism genes, Mol. Cell. 82 (2022) 60-74.e5. https://doi.org/10.1016/j.molcel.2021.12.015.

[240] G. Andrieu, A.C. Belkina, G. V. Denis, Clinical trials for BET inhibitors run ahead of the science, Drug Discov. Today Technol. 19 (2016) 45–50. https://doi.org/10.1016/j.ddtec.2016.06.004.

[241] E.F. Joest, C. Winter, J.S. Wesalo, A. Deiters, R. Tampé, Light-guided intrabodies for on-demand in situ target recognition in human cells, Chem. Sci. 12 (2021) 5787–5795. https://doi.org/10.1039/D1SC01331A.

[242] S.B. Rothbart, B.D. Strahl, Interpreting the language of histone and DNA modifications, Biochim. Biophys. Acta - Gene Regul. Mech. 1839 (2014) 627–643. https://doi.org/10.1016/j.bbagrm.2014.03.001.

[243] A. Gavrilov, S. V. Razin, G. Cavalli, In vivo formaldehyde cross-linking: it is time for black box analysis, Brief. Funct. Genomics. 14 (2015) 163–165. https://doi.org/10.1093/bfgp/elu037.

[244] M.E. Sardiu, M.P. Washburn, Building Protein-Protein Interaction Networks with Proteomics and Informatics Tools, J. Biol. Chem. 286 (2011) 23645–23651. https://doi.org/10.1074/jbc.R110.174052.

[245] Y. Sato, M. Mukai, J. Ueda, M. Muraki, T.J. Stasevich, N. Horikoshi, T. Kujirai, H. Kita, T. Kimura, S. Hira, Y. Okada, Y. Hayashi-Takanaka, C. Obuse, H. Kurumizaka, A. Kawahara, K. Yamagata, N. Nozaki, H. Kimura, Genetically encoded system to track histone modification in vivo, Sci. Rep. 3 (2013) 2436. https://doi.org/10.1038/srep02436.

[246] R. Villaseñor, R. Pfaendler, C. Ambrosi, S. Butz, S. Giuliani, E. Bryan, T.W. Sheahan, A.L. Gable, N. Schmolka, M. Manzo, J. Wirz, C. Feller, C. von Mering, R. Aebersold, P. Voigt, T. Baubec, ChromID identifies the protein interactome at chromatin marks, Nat. Biotechnol. 38 (2020) 728–736. https://doi.org/10.1038/s41587-020-0434-2.

[247] M. Vermeulen, H.C. Eberl, F. Matarese, H. Marks, S. Denissov, F. Butter, K.K. Lee, J. V. Olsen, A.A. Hyman, H.G. Stunnenberg, M. Mann, Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers, Cell. 142 (2010) 967–980. https://doi.org/10.1016/j.cell.2010.08.020.

[248] S.A. Myers, J. Wright, R. Peckner, B.T. Kalish, F. Zhang, S.A. Carr, Discovery of proteins associated with a predefined genomic locus via dCas9–APEX-mediated proximity

labeling, Nat. Methods. 15 (2018) 437–439. https://doi.org/10.1038/s41592-018-0007-1.

[249] X. Liu, Y. Zhang, Y. Chen, M. Li, F. Zhou, K. Li, H. Cao, M. Ni, Y. Liu, Z. Gu, K.E. Dickerson, S. Xie, G.C. Hon, Z. Xuan, M.Q. Zhang, Z. Shao, J. Xu, In Situ Capture of Chromatin Interactions by Biotinylated dCas9, Cell. 170 (2017) 1028-1043.e19. https://doi.org/10.1016/j.cell.2017.08.003.

[250] D.G. Gibson, L. Young, R.-Y. Chuang, J.C. Venter, C.A. Hutchison, H.O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases, Nat. Methods. 6 (2009) 343–345. https://doi.org/10.1038/nmeth.1318.

[251] A.L. Goldstein, J.H. McCusker, Three new dominant drug resistance cassettes for gene disruption inSaccharomyces cerevisiae, Yeast. 15 (1999) 1541–1553. https://doi.org/10.1002/(SICI)1097-0061(199910)15:14<1541::AID-YEA476>3.0.CO;2-K.

[252] A. Wach, A. Brachat, R. Pöhlmann, P. Philippsen, New heterologous modules for classical or PCR-based gene disruptions inSaccharomyces cerevisiae, Yeast. 10 (1994) 1793–1808. https://doi.org/10.1002/yea.320101310.

[253] M.K. Chee, S.B. Haase, New and Redesigned pRS Plasmid Shuttle Vectors for Genetic Manipulation of Saccharomyces cerevisiae, G3: Genes|Genomes|Genetics. 2 (2012) 515–526. https://doi.org/10.1534/g3.111.001917.

[254] S. Mehta, X.M. Yang, C.S. Chan, M.J. Dobson, M. Jayaram, S. Velmurugan, The 2 micron plasmid purloins the yeast cohesin complex, J. Cell Biol. 158 (2002) 625–637. https://doi.org/10.1083/jcb.200204136.

[255] G.R.F. Christine Guthrie, Guide to Yeast Genetics and Molecular Biology, 1991. https://www.sciencedirect.com/bookseries/methods-in-enzymology/vol/194/suppl/C.

[256] J.M. Gardner, S.L. Jaspersen, Manipulating the Yeast Genome: Deletion, Mutation, and Tagging by PCR, in: 2014: pp. 45–78. https://doi.org/10.1007/978-1-4939-1363-3_5.

[257] D.S. Wilson, A.D. Keefe, Random Mutagenesis by PCR, in: Curr. Protoc. Mol. Biol., 2001. https://doi.org/10.1002/0471142727.mb0803s51.

[258] K.U. Sprague, Transcription of Eukaryotic tRNA Genes, in: TRNA, ASM Press, Washington, DC, USA, 2014: pp. 31–50. https://doi.org/10.1128/9781555818333.ch4.

[259] C.A. Otter, K.B. Straby, Transcription of eukaryotic genes with impaired internal promoters: the use of a yeast tRNA gene as promoter, J. Biotechnol. 21 (1991) 289–293. https://doi.org/10.1016/0168-1656(91)90049-2.

[260] C.A. Otter, J. Edqvist, K.B. Stråby, Characterization of transcription and processing from plasmids that use polIII and a yeast tRNA gene as promoter to transcribe promoterdeficient downstream DNA, Biochim. Biophys. Acta - Gene Struct. Expr. 1131 (1992) 62–68. https://doi.org/10.1016/0167-4781(92)90099-L.

[261] A. Crnković, O. Vargas-Rodriguez, D. Söll, Plasticity and Constraints of tRNA

Aminoacylation Define Directed Evolution of Aminoacyl-tRNA Synthetases, Int. J. Mol. Sci. 20 (2019) 2294. https://doi.org/10.3390/ijms20092294.

[262] F. Beranger, Getting more from the two-hybrid system: N-terminal fusions to LexA are efficient and sensitive baits for two-hybrid studies, Nucleic Acids Res. 25 (1997) 2035–2036. https://doi.org/10.1093/nar/25.10.2035.

[263] D. Cao, R. Parker, Computational Modeling and Experimental Analysis of Nonsense-Mediated Decay in Yeast, Cell. 113 (2003) 533–545. https://doi.org/10.1016/S0092-8674(03)00353-2.

[264] M. Peccarelli, B.W. Kebaara, Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway, Eukaryot. Cell. 13 (2014) 1126–1135. https://doi.org/10.1128/EC.00090-14.

[265] M. Fantini, S. Lisi, P. De Los Rios, A. Cattaneo, A. Pastore, Protein Structural Information and Evolutionary Landscape by In Vitro Evolution, Mol. Biol. Evol. 37 (2020) 1179–1192. https://doi.org/10.1093/molbev/msz256.

[266] D.F. Steele, S. Jinks-Robertson, An examination of adaptive reversion in Saccharomyces cerevisiae., Genetics. 132 (1992) 9–21. https://doi.org/10.1093/genetics/132.1.9.

[267] V. Beránek, J.C.W. Willis, J.W. Chin, An Evolved Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase/tRNA Pair Is Highly Active and Orthogonal in Mammalian Cells, Biochemistry. 58 (2019) 387–390. https://doi.org/10.1021/acs.biochem.8b00808.

[268] M. Kaishima, J. Ishii, T. Matsuno, N. Fukuda, A. Kondo, Expression of varied GFPs in Saccharomyces cerevisiae: Codon optimization yields stronger than expected expression and fluorescence intensity, Sci. Rep. 6 (2016). https://doi.org/10.1038/srep35932.

[269] V.E. DeLey Cox, M.F. Cole, E.A. Gaucher, Incorporation of Modified Amino Acids by Engineered Elongation Factors with Expanded Substrate Capabilities, ACS Synth. Biol. 8 (2019) 287–296. https://doi.org/10.1021/acssynbio.8b00305.

[270] E. De Genst, C.M. Dobson, Nanobodies as Structural Probes of Protein Misfolding and Fibril Formation, in: 2012: pp. 533–558. https://doi.org/10.1007/978-1-61779-968-6_34.

[271] P. Arosio, T. Müller, L. Rajah, E. V. Yates, F.A. Aprile, Y. Zhang, S.I.A. Cohen, D.A. White, T.W. Herling, E.J. De Genst, S. Linse, M. Vendruscolo, C.M. Dobson, T.P.J. Knowles, Microfluidic Diffusion Analysis of the Sizes and Interactions of Proteins under Native Solution Conditions, ACS Nano. 10 (2016) 333–341. https://doi.org/10.1021/acsnano.5b04713.

[272] T. Guilliams, F. El-Turk, A.K. Buell, E.M. O'Day, F.A. Aprile, E.K. Esbjörner, M. Vendruscolo, N. Cremades, E. Pardon, L. Wyns, M.E. Welland, J. Steyaert, J. Christodoulou, C.M. Dobson, E. De Genst, Nanobodies Raised against Monomeric α-Synuclein Distinguish between Fibrils at Different Maturation Stages, J. Mol. Biol. 425 (2013) 2397–2411. https://doi.org/10.1016/j.jmb.2013.01.040.

[273] A.K. Hopper, Transfer RNA Post-Transcriptional Processing, Turnover, and Subcellular Dynamics in the Yeast Saccharomyces cerevisiae, Genetics. 194 (2013) 43–67. https://doi.org/10.1534/genetics.112.147470.

[274] D. Belin, P. Puigbò, Why Is the UAG (Amber) Stop Codon Almost Absent in Highly Expressed Bacterial Genes?, Life. 12 (2022) 431. https://doi.org/10.3390/life12030431.

[275] Y. Kato, Translational Control using an Expanded Genetic Code, Int. J. Mol. Sci. 20 (2019) 887. https://doi.org/10.3390/ijms20040887.

[276] D.B.F. Johnson, J. Xu, Z. Shen, J.K. Takimoto, M.D. Schultz, R.J. Schmitz, Z. Xiang, J.R. Ecker, S.P. Briggs, L. Wang, RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites, Nat. Chem. Biol. 7 (2011) 779–786. https://doi.org/10.1038/nchembio.657.

[277] D. Botstein, S.A. Chervitz, M. Cherry, Yeast as a Model Organism, Science (80-.). 277 (1997) 1259–1260. https://doi.org/10.1126/science.277.5330.1259.

[278] C.L. Quinn, N. Tao, P. Schimmel, Species-specific microhelix aminoacylation by a eukaryotic pathogen tRNA synthetase dependent on a single base pair, Biochemistry. 34 (1995) 12489–12495. https://doi.org/10.1021/bi00039a001.

[279] T.H. Yoo, D.A. Tirrell, High-Throughput Screening for Methionyl-tRNA Synthetases That Enable Residue-Specific Incorporation of Noncanonical Amino Acids into Recombinant Proteins in Bacterial Cells, Angew. Chemie Int. Ed. 46 (2007) 5340–5343. https://doi.org/10.1002/anie.200700779.

[280] J. Mehla, J.H. Caufield, N. Sakhawalkar, P. Uetz, A Comparison of Two-Hybrid Approaches for Detecting Protein–Protein Interactions, Methods Enzymol. 586 (2017) 333–358. https://doi.org/10.1016/bs.mie.2016.10.020.

[281] D. Auerbach, I. Stagljar, Yeast Two-Hybrid Protein-Protein Interaction Networks, in: Proteomics Protein-Protein Interact., Springer US, Boston, MA, n.d.: pp. 19–31. https://doi.org/10.1007/0-387-24532-4_2.

[282] M. Yamagata, J.R. Sanes, Reporter–nanobody fusions (RANbodies) as versatile, small, sensitive immunohistochemical reagents, Proc. Natl. Acad. Sci. 115 (2018) 2126–2131. https://doi.org/10.1073/pnas.1722491115.

[283] C. Fan, H. Xiong, N.M. Reynolds, D. Söll, Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids, Nucleic Acids Res. 43 (2015). https://doi.org/10.1093/nar/gkv800.

[284] J.-D. Pédelacq, S. Cabantous, T. Tran, T.C. Terwilliger, G.S. Waldo, Engineering and characterization of a superfolder green fluorescent protein, Nat. Biotechnol. 24 (2006) 79–88. https://doi.org/10.1038/nbt1172.

[285] A. Pörtner-Taliana, M. Russell, K.J. Froning, P.R. Budworth, J.D. Comiskey, J.P. Hoeffler,

In vivo selection of single-chain antibodies using a yeast two-hybrid system, J. Immunol. Methods. 238 (2000) 161–172. https://doi.org/10.1016/S0022-1759(00)00145-9.

[286] H. Koch, N. Gräfe, R. Schiess, A. Plückthun, Direct Selection of Antibodies from Complex Libraries with the Protein Fragment Complementation Assay, J. Mol. Biol. 357 (2006) 427–441. https://doi.org/10.1016/j.jmb.2005.12.043.

[287] L.S. Mitchell, L.J. Colwell, Comparative analysis of nanobody sequence and structure data, Proteins Struct. Funct. Bioinforma. 86 (2018) 697–706. https://doi.org/10.1002/prot.25497.

[288] A.R. Poteete, What makes the bacteriophage Î» Red system useful for genetic engineering: molecular mechanism and biological function, FEMS Microbiol. Lett. 201 (2001) 9–14. https://doi.org/10.1111/j.1574-6968.2001.tb10725.x.

[289] R. Swanson, P. Hoben, M. Sumner-Smith, H. Uemura, L. Watson, D. Söll, Accuracy of in Vivo Aminoacylation Requires Proper Balance of tRNA and Aminoacyl-tRNA Synthetase, Science (80-.). 242 (1988) 1548–1551. https://doi.org/10.1126/science.3144042.

[290] B.J. Bachmann, K.B. Low, A.L. Taylor, Recalibrated linkage map of Escherichia coli K-12, Bacteriol. Rev. 40 (1976) 116–167. https://doi.org/10.1128/br.40.1.116-167.1976.

[291] D.J. Selkoe, J. Hardy, The amyloid hypothesis of Alzheimer's disease at 25 years, EMBO Mol. Med. 8 (2016) 595–608. https://doi.org/10.15252/emmm.201606210.

[292] R. La Joie, A. V. Visani, S.L. Baker, J.A. Brown, V. Bourakova, J. Cha, K. Chaudhary, L. Edwards, L. Iaccarino, M. Janabi, O.H. Lesman-Segev, Z.A. Miller, D.C. Perry, J.P. O'Neil, J. Pham, J.C. Rojas, H.J. Rosen, W.W. Seeley, R.M. Tsai, B.L. Miller, W.J. Jagust, G.D. Rabinovici, Prospective longitudinal atrophy in Alzheimer's disease correlates with the intensity and topography of baseline tau-PET, Sci. Transl. Med. 12 (2020). https://doi.org/10.1126/scitranslmed.aau5732.

[293] C.-W. Chang, E. Shao, L. Mucke, Tau: Enabler of diverse brain disorders and target of rapidly evolving therapeutic strategies, Science (80-.). 371 (2021). https://doi.org/10.1126/science.abb8255.

[294] C. Despres, C. Byrne, H. Qi, F.-X. Cantrelle, I. Huvent, B. Chambraud, E.-E. Baulieu, Y. Jacquot, I. Landrieu, G. Lippens, C. Smet-Nocca, Identification of the Tau phosphorylation pattern that drives its aggregation, Proc. Natl. Acad. Sci. 114 (2017) 9080–9085. https://doi.org/10.1073/pnas.1708448114.

[295] X. Wang, Q. Liu, X.-G. Li, Q.-Z. Zhou, D.-Q. Wu, S.-H. Li, Y.-C. Liu, J.-Z. Wang, T217-Phosphorylation Exacerbates Tau Pathologies and Tau-Induced Cognitive Impairment, J. Alzheimer's Dis. 81 (2021) 1403–1418. https://doi.org/10.3233/JAD-210297.

[296] M. Morris, G.M. Knudsen, S. Maeda, J.C. Trinidad, A. Ioanoviciu, A.L. Burlingame, L. Mucke, Tau post-translational modifications in wild-type and human amyloid precursor protein transgenic mice, Nat Neurosci. 18 (2015) 1183–1189.

https://doi.org/10.1038/nn.4067.

[297] A. Ittner, S.W. Chua, J. Bertz, A. Volkerling, J. van der Hoven, A. Gladbach, M. Przybyla, M. Bi, A. van Hummel, C.H. Stevens, S. Ippati, L.S. Suh, A. Macmillan, G. Sutherland, J.J. Kril, A.P.G. Silva, J.P. Mackay, A. Poljak, F. Delerue, Y.D. Ke, L.M. Ittner, Site-specific phosphorylation of tau inhibits amyloid-β toxicity in Alzheimer's mice, Science (80-.). 354 (2016) 904–908. https://doi.org/10.1126/science.aah6205.

[298] K. Stefanoska, M. Gajwani, A.R.P. Tan, H.I. Ahel, P.R. Asih, A. Volkerling, A. Poljak, A. Ittner, Alzheimer's disease: Ablating single master site abolishes tau hyperphosphorylation, Sci. Adv. 8 (2022). https://doi.org/10.1126/sciadv.abl8809.

[299] H.H. Shih, C. Tu, W. Cao, A. Klein, R. Ramsey, B.J. Fennell, M. Lambert, D. Ní Shúilleabháin, B. Autin, E. Kouranova, S. Laxmanan, S. Braithwaite, L. Wu, M. Ait-Zahra, A.J. Milici, J.A. Dumin, E.R. LaVallie, M. Arai, C. Corcoran, J.E. Paulsen, D. Gill, O. Cunningham, J. Bard, L. Mosyak, W.J.J. Finlay, An Ultra-specific Avian Antibody to Phosphorylated Tau Protein Reveals a Unique Mechanism for Phosphoepitope Recognition, J. Biol. Chem. 287 (2012) 44425–44434. https://doi.org/10.1074/jbc.M112.415935.

[300] C. Di Primio, V. Quercioli, G. Siano, B. Kovacech, M. Novak, A. Cattaneo, Conformational dynamics of Tau in the cell quantified by an intramolecular FRET biosensor in physiological and pathological context, BioRxiv. (2016). https://doi.org/10.1101/041756.

[301] C. Di Primio, V. Quercioli, G. Siano, M. Rovere, B. Kovacech, M. Novak, A. Cattaneo, The distance between N and C termini of tau and of FTDP-17 mutants is modulated by microtubule interactions in living cells, Front. Mol. Neurosci. 10 (2017). https://doi.org/10.3389/fnmol.2017.00210.

[302] T. Melchionna, A. Cattaneo, A Protein Silencing Switch by Ligand-induced Proteasome-targeting Intrabodies, J. Mol. Biol. 374 (2007) 641–654. https://doi.org/10.1016/j.jmb.2007.09.053.

[303] S. Biocca, F. Ruberti, M. Tafani, P. Pierandrel-Amaldi, A. Cattaneo, Redox State of Single Chain Fv Fragments Targeted to the Endoplasmic Reticulum, Cytosol and Mitochondria, Bio/Technology. 13 (1995) 1110–1115. https://doi.org/10.1038/nbt1095-1110.

[304] D.P. Nguyen, M.M. Garcia Alai, P.B. Kapadnis, H. Neumann, J.W. Chin, Genetically Encoding N ϵ -Methyl- l -lysine in Recombinant Histones, J. Am. Chem. Soc. 131 (2009) 14194–14195. https://doi.org/10.1021/ja906603s.