Metabolic channelling for biofuel production

Co-localization of Pdc and Adh

Isabel Moreno de Palma

Degree project in biology, Master of science (2 years), 2016
Examensarbete i biologi 45 hp till masterexamen, 2016
Biology Education Centre and Department of Chemistry, Ångström laboratory, Uppsala University
Supervisors: Pia Lindberg and Elias Englund
External opponent: Claudia Duall de la Fuente
# Contents

Abstract .................................................................................................................................................. 4

1. Introduction ........................................................................................................................................ 5
   1.1. Goal of the master thesis .............................................................................................................. 5
   1.2. Ethanol as a model for biofuel ..................................................................................................... 5
   1.3. Metabolic channelling effect ....................................................................................................... 6
   1.4. Experiment overview .................................................................................................................... 12

2. Materials and methods .................................................................................................................... 14
   2.1. Modelling of enzymes .................................................................................................................. 14
   2.2. Bacterial strains and culture conditions ........................................................................................ 14
   2.3. Generation of genetic constructs ................................................................................................ 14
   2.4. Deletion of adhE ......................................................................................................................... 16
   2.5. Ethanol detection ....................................................................................................................... 16
   2.6. Protein detection ....................................................................................................................... 17

3. Results .............................................................................................................................................. 17
   3.1. Genetic constructs ....................................................................................................................... 17
   3.2. Protein detection ....................................................................................................................... 22
   3.3. Knockout of adhE ..................................................................................................................... 23
   3.4. Fusion ......................................................................................................................................... 24
   3.5. Peptide ....................................................................................................................................... 28
3.6. DNA and RNA enzymes without scaffolds .................................................. 31
3.7. DNA scaffold ................................................................................................. 32
3.8. RNA scaffold ................................................................................................. 33
3.9. Media and growing conditions optimization .................................................. 34

4. Discussion ............................................................................................................ 36

Abbreviations ........................................................................................................... 38
Acknowledgement ..................................................................................................... 39
References ................................................................................................................ 42
Abstract

Enhancing productivity in bioprocesses, especially for biofuel production, is crucial for achieving an environmentally and economically sustainable biotechnology industry. Metabolic channelling occurs in nature when the intermediate between two consecutive enzymes in a pathway is directed from the first enzyme to the second avoiding diffusion in the cytosol. This would be very advantageous in bioprocesses as it would increase efficiency of a particular pathway, reducing side products and protecting the cells from potential toxic intermediates. In recent years different strategies for emulating channelling effect were proposed and used with very promising results. Clustering of enzymes seems to be the simplest way to create metabolic channelling. In this master thesis, four different strategies to co-localize enzymes in clusters are compared. The metabolic pathway chosen as a model was ethanol production by pyruvate decarboxylase (Pdc) and alcohol dehydrogenase (Adh). Chimeric proteins were genetically engineered and transformed in E. coli creating different strains. Ethanol production by the different strains was measured to compare production efficiency. Cell growth and protein expression were used for further understanding of the results. Strengths and weaknesses of each strategy and proposals for further improvement were discussed.
1. Introduction

1.1. Goal of the master thesis

As a reaction to growing concerns about the highly contaminating and not-renewable usage of fossil fuels, biofuels appeared as a very promising substitution. Initial optimism about biofuels encountered soon economic, environmental and ethical challenges\textsuperscript{1,2}. Several generations of biofuels have been following each other to attend these issues\textsuperscript{3}. Genetically modified algae for direct biofuel production may overcome most of the environmental and ethical challenges\textsuperscript{4,5}. But production of biofuel from algae is still far from being an efficient system. Yields still need to be improved to obtain economically competent biofuel production\textsuperscript{6}.

Organisms, even simple ones as bacteria, have complex metabolic routes interconnected. Usually the same metabolite is substrate downstream for different enzymes in different routes. In nature, flux equilibrium of metabolites through different routes is optimized for the survival of the organism. Production of any metabolite of interest in biotechnology requires changing this equilibrium towards production enhancement instead of survival.

The goal of this thesis is to explore metabolic channelling as a strategy for yield improvement in biofuel production. \textit{E. coli} was used as a model for later usage in cyanobacteria.

1.2. Ethanol as a model for biofuel

Bioethanol is the most common biofuel nowadays. It can be directly blended with gasoline and its combustion is cleaner than the combustion of the latter. Ethanol has traditionally being biologically produced and has been studied for a long time. It is a well-known process which allows ethanol synthesis from pyruvate, present in all living organisms. During fermentation organisms produce highly reduced compounds, such as ethanol, in order to eliminate the reduction equivalents of this process. The simplest pathway for ethanol production involves only two enzymes, pyruvate decarboxylase (\textit{Pdc}) and alcohol dehydrogenase (\textit{Adh}) $\text{CO}_2$ (Figure 1). Therefore any organism could potentially be made to

\begin{equation*}
\begin{align*}
\text{Pyruvate} & \xrightarrow{\text{Pdc}} \text{Acetaldehyde} & \xrightarrow{\text{Adh}} \text{Ethanol}
\end{align*}
\end{equation*}

\textbf{Figure 1. Ethanol production from pyruvate by \textit{Pdc} and \textit{Adh}.}
produce ethanol by expressing these enzymes.

1.3. Metabolic channelling effect

Metabolic channelling is the property when the intermediate metabolite between two consecutives enzymes in a pathway avoids diffusion in the cytosol and is catalysed directly by the next enzyme, thus favouring that pathway from others\(^7\). This way metabolic channelling can enhance production by preferentially enforce a specific metabolic route and protect the cell from potentially toxic intermediates.

Direct channelling can be achieved by physical connection between active sites, where the intermediate is funnelled from the first active site to the next (Figure 2A).

Proximity channelling happens when active sites of consecutive enzymes are close enough, less than 10 Å\(^7\) (Figure 2B). These two strategies of channelling are difficult to emulate for biotechnological purposes, every method would need a thorough analysis of the structure of the protein, linkers and binding domains. But it also has been theorized that channelling can be achieved by clusters of enzymes\(^8\). The cluster-mediated channelling model stands that when two consecutive enzymes are agglomerated together, creating a high concentration of enzymes in one region, channelling effect can be achieved (Figure 2C).

Different metabolites production has been reported to increase by co-localization of enzymes, supposedly due to channelling effect. Methods used for this purpose have been very different: simple fusion of enzymes, use of micro-compartments in the cell, binding

---

**Figure 2. Representation of channeling effect.** Enzyme 1 (E1) transforms the substrate to the intermediate which is used by enzyme 2 (E2) to produce the product. Direct channeling (A): the intermediate is funneled from E1 to E2 through a physical tunnel. Proximity channel (B) occurs when E1 and E2 are 0.1 – 1 nm apart, then intermediate avoids diffusion and is directly used by E2. If enzymes are further apart the intermediate is diffused in the cell. Cluster mediated channeling (C): when E1 and E2 are clustered together there is a local high concentration of both enzymes. When the intermediate is produced inside this cluster, with high concentration of both enzymes, it is very likely that it will bind E2 before leaving the cluster, even if E1 and E2 catalytic domains are further 0.1-1 nm apart.
enzymes to the membrane and using synthetic peptide, DNA or RNA scaffolds.\textsuperscript{9–23}

Fusion is the simplest and it has shown good results with of two enzymes, but for more enzymes it is very likely to have folding problems. Enclosing enzymes in micro-compartments would allow proper folding of the enzymes, protecting the organisms of toxic intermediates, even allow to have specific conditions for the enzymes; nevertheless with micro-compartments it also is more difficult for previous metabolites of the pathway to reach the enzymes and for the product of these enzymes to go to the next step of the pathway, or to be collected if it is the metabolite of interest. Using scaffolds could allow regulation of stoichiometry and binding of the enzymes, but it would also mean adding one more external element to the cluster and more resources for the host organism. Using peptide binding domains and ligands fused to the enzymes would avoid using the scaffold, but it could only be used for two consecutive enzymes.

Nevertheless all of these strategies have been shown to enhance production, but no comparison among them has been done. Generalizing co-localization methods would be very desirable for biotechnology. Is it possible to know which of these methods would be the best for generalized use? In this master thesis four different were compared: (a) fusion of enzymes, (b) SH3 peptide binding domains, (c) DNA scaffolds and (d) RNA scaffolds.

Fusion

In nature, many enzymes evolved by fusion of different catalytic domains. Inspired in nature, artificial fusion of enzymes is a simple technique which only requires deletion of the stop codon of one of the genes and addition the ORF of the next, usually connected by a flexible glycine rich linker. The biggest advantage of this method is the simplicity and economy for the cell as it does not need extra domains or scaffolds.

Fusion of enzymes has been used with different purposes and it has been shown to be useful to increase production

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
Fusion & 5.8 ± 1.2 \\
\hline
FPPS – PTS & 9.5 ± 0.6 \\
\hline
PTS – FPPS & 5.7 ± 1.2 \\
\hline
\end{tabular}
\caption{Example of protein fusion strategy from Albertssen et al. 2010. FPPS and PTS were fused to a single big protein either with C-terminus of FPPS fused to N-terminus of PTS C-terminus or C-terminus of PTS to N-terminus of FPPS. One of the fusion produced nearly twice while the fusion with the enzymes in the opposite order did not produce any enhancement.}
\end{table}
both in vitro and in vivo. In fact, $Pdc$ and $AdhB$ from Z. mobilis have already been fused with enhancement in ethanol production. Nevertheless this strategy can disrupt the proper folding of enzymes, leading to decrease of activity or loss of function of the enzymes. For this reason it is very difficult to extend to more than two enzymes, which is the greatest limitation of this method for enhancing pathways.

The order of the fused proteins is a key factor for the efficiency of the process (Figure 3). The length and nature of the linker is also an important factor. Short linkers bring closer together the enzymes facilitating channelling, but it might cause a negative effect in the folding process of the enzymes. Longer linkers may facilitate folding but are exposed areas which can be targets for peptidases which can lead in separation of the fused proteins. Flexible linkers would allow proper folding of the enzymes. On the other hand, rigid linkers can be design to achieve a specific orientation of the enzymes in order to bring closer both catalytic domains. The latter, nevertheless, requires thorough understanding of the structure of both enzymes and good prediction and designing tools for the linker structure.

**Peptide binding domains**

Peptide binding domains are commonly used in nature to control flux of metabolites between enzymes. They can be easily regulated thus in nature they are usually found in signalling pathways.

There are two main approaches for co-localization of enzymes with peptide binding domains, both inspired in nature: direct binding of the enzymes and binding of the enzymes to a peptide scaffold. In the first case a binding domain is fused to one of the enzymes and its ligand is fused to the second enzyme (Figure 4A). In the second case different ligands can be fused to the enzymes and a scaffold peptide containing all different binding domains (Figure 4B). There are many known binding domains representing enough building blocks for the generalization of the method. One type of well-known peptide binding domains are Src homology 3 (SH3) domains, this type of domain is used in the experiment of this master thesis, based on previous work from Dueber et al. where they showed a 77 fold increase in mevalonate production. They are small domains, and can bind short proline rich sequences with specific directionality. They are intensively studied at the moment, and advances in
the knowledge of the binding mechanism could help engineering specific ligands and domains.

The biggest advantage in comparison with fusion is that stoichiometry can be controlled by changing the relative number of binding domains (both when using scaffolds or direct binding). Another advantage is the variety of peptide binding domains available. This supposes more work during the design of the strategy but it allows being more specific when trying to achieve proximity channelling. In the case of peptide scaffolds it is very interesting for the possibility of co-localizing more than two enzymes together.

A disadvantage of this strategy is degradation of the enzymes by host organism, as the intrinsically disorder sequences that characterizes binding domains are recognised by proteases.

The key factor which needs to be optimized when using peptide binding domain is stoichiometry, controlling the relative number of the different enzymes bound together. It has also been shown that the number of scaffolds expressed in the cell is a decisive factor. It is necessary to have enough scaffolds to have all the enzymes bound to a scaffold to achieve

![Figure 4: Example of a peptide scaffolding strategy for mevalonate production from Dueber et al. A. Direct binding domains strategy: different number of ligands was attached to HMGS and one binding domain was attached to HMGR, the next enzyme in the pathway. When no ligand was attached to HMGS it could not bind to HMGR (A1). The ligand attached to HMGS binds to the domain attached to HMGR, co-localizing both enzymes (A2). Several ligands can be attached to HMGS to bind more HMGR (A3). Thus, stoichiometry can be regulated controlling the number of ligands attached to HMGS (A4). B. Peptide scaffold strategy. Several binding domains can be synthesized together to build a synthetic protein scaffold that can be recognized by ligands with high specificity. This type of co-localization allows the arrangement of more than two enzymes of a pathway. In protein scaffolds the stoichiometry is regulated by the number of the specific binding domains for each protein.]
co-localization but, when exceeding the optimal number of scaffolds, it leads to decrease efficiency. Yet in this experiment no peptide scaffold was used so only the first factor needs to be optimized.

**DNA scaffold**

This method of scaffolding requires two different elements: DNA binding domains to be attached to the enzymes and a DNA scaffold which. The DNA binding domains recognise specific regions of the DNA scaffold. The number and order of these recognition regions can be engineered to decide the stoichiometry and sequence in which the enzymes are bound (Figure 5). DNA is an attractive scaffold because of the easy predictable structure of double stranded DNA. This could be very useful to engineer cascades of enzymes and it has already been used in-vitro and in-vivo\(^9,15,18\). Even though rigidity of DNA does not allow so many shapes as peptide or RNA scaffolds to fit the enzymes in an optimal position, it has been used to create in-vitro computer predicted nanostructures of a great variety of different shapes\(^28\).

In nature there are different DNA binding domains; zinc fingers are relatively well known ones. These are small B-DNA binding domains formed only by one β-sheet and one α-helix bound by hydrophobic interactions and stabilized by a Zn ion\(^29\). Zinc finger domains recognize combinations of three consecutive bp in guanine rich regions. When several zinc fingers are together in tandem they form a helix that fits into the major groove of the DNA, binding consecutive triplets of nucleotides. Thus, different combinations of zinc fingers can be engineered to recognise longer DNA regions for a more specific binding\(^30\). This is a building block system very

![Figure 5. Example of a DNA scaffolding strategy from Conrado et al. 2012. A specific combination of zinc finger domains (a, b, c) is attached to every enzyme (E1, E2, E3). Different DNA scaffolds are built with specific sequences recognized by the zinc fingers domains. Stoichiometry can be changed by creating more DNA recognition sequences for the desired enzyme.](image)
attractive for synthetic biology. The modular nature and easy engineering of DNA is the most attractive characteristic of this method. The main disadvantage, especially when clustering enzymes together, is the big size of the DNA scaffold.

Conrado et al. 2012, use these domains for co-localization of 4CL and STS achieving five-fold enhancement of 1, 2 propanediol. The factors that have been shown to affect the most for this method are the order, number and distance between zinc fingers recognition sequences and the number of scaffold plasmids in the cell.

**RNA scaffold**

The extra oxygen in ribonucleotides allows RNA more variety in shapes and reactivity than DNA. RNA scaffolds can be design to present different 3D structures which allows spatial organization of the enzymes. Recognition of RNA can be achieved not only by its specific nucleotide sequence, but also by specific RNA structures. Aptamers are RNA structures that recognise specific protein domains mainly by shape rather than sequence. Using different aptamers, Delebeque et al. successfully produced RNA scaffolds forming ribbons and nanotubes in vivo (Figure 6). The scaffolds were tested for Hydrogen production with a maximum increase of 48 fold for 2D scaffolds. This strategy was later on successfully applied for pentadecane synthesis and for multiple enzymes of the
succinate pathway. As a result they were able to increase production while reducing side products\textsuperscript{12}.

Length and orientation of the aptamers (related to the relative orientation of the enzymes) and level of organization (0, 1 and 2 dimensions) of the RNA structures seem to be the most important factors to optimize production with RNA scaffolds\textsuperscript{12,13}. The higher the level of organization and concentration of enzymes, the greater production obtained, which could be related with cluster channelling.

New and more specific protein motif-aptamer interactions are being designed as RNA nanotechnology evolves\textsuperscript{31}, providing new building blocks for this co-localization method.

1.4. Experiment overview

Four methods of co-localization were used to achieve cluster-mediated channelling (Figure 7): fusion of enzymes, peptide binding domains, DNA scaffolding and RNA scaffolding. The two enzymes \textit{Pdc} and \textit{Adh} used in this study are homotetramers formed as dimer of dimers, therefore cluster-mediated channelling can be achieved by co-localizing every monomer of \textit{Pdc} with another monomer of \textit{Adh}\textsuperscript{32,33}. The channelling effect was measured by ethanol production improvement by \textit{Pdc} and \textit{Adh}. Relevant factors affecting each method were modified in order to achieve optimal production for each method.

\textit{Pdc} and \textit{Adh} were fused separated by a short GSG linker in both orders \textit{Pdc-Adh} and \textit{Adh-Pdc}. Controls of the method were overexpression of both \textit{Pdc} and \textit{Adh} separately and a negative control with the empty plasmid (Figure 7, Fusion).

Peptide binding domain co-localization method was design following previous work from Dueber et al.\textsuperscript{16}. Four different combinations were tested using SH3 domains and one or two SH3 linkers attached to the N-terminus of \textit{Pdc} and \textit{Adh}. For this method control strains were the same as for fusion method (Figure 7, Peptide).

For DNA scaffolding Zif268 land PSB2 zinc fingers domain were attached to N-terminus of \textit{Pdc} and \textit{Adh} respectively. Zinc finger domains and flexible linkers were chosen following Conrado et al previous work\textsuperscript{9}. Four different plasmids were engineered as scaffolds with different stoichiometry, separation between recognition sequences and repetitions per plasmid. The controls for this method was an strain overexpressing \textit{Pdc} and \textit{Adh} without the zinc finger domains and with the empty scaffold plasmid and a negative control with both plasmids empty of enzymes and scaffold in order to be grown in the same
antibiotics. Another strain overexpressing Zif268-Pdc and PSB2-Adh with the empty scaffold plasmid was created to analyse the effect of the zinc finger domains to the activity of the enzymes (Figure 7, DNA scaffold).

Aptamer domains were chosen according to Delebecque et al. previous work. BIVtat and PP7 aptamer recognition sequences were attached to Pdc and Adh respectively. Three different RNA scaffolds were used: an RNA molecule with both aptamers to bind both enzymes with no polymerization capacity (dimension 0, d₀); an RNA structure that can bind both enzymes and with the capacity to form polymers (d₁); and two RNA molecules that interact with each other forming two dimension structures (d₂) (Figure 7, RNA scaffold).

Figure 7. Overview of the different strategies used in this experiment for co-localize enzymes. Different strategies are represented from left to right: Fusion: expression of pdc and adh fused genes; Peptide: Fusion of SH3 domains and ligands to Pdc and Adh; DNA scaffold: use of DNA scaffolds recognized by zinc fingers domains attached to Pdc and Adh; RNA scaffolds: use of RNA scaffolds with aptamers recognized by specific domains attached to Pdc and Adh. The schematic process is represented from top to bottom: a: Chimeric Pdc (green) and Adh (orange) are transcribed; b: interaction domains bind their target domain (peptide) or scaffold (DNA and RNA scaffolds); c: Clusters are formed as a result of the interaction of the Adh and Pdc monomers to form the active tetramer forms (not represented for RNA). In the case of RNA scaffolds the RNA molecules are also synthesized (a’) and interact with each other to form different structures (b’).
Control strains were two expressing or not expressing Pdc and Adh with the empty plasmid for the RNA scaffolds, plus a strain expressing BIVtat-Pdc and PP7-Adh with the empty vector for the RNA scaffolds, to see the effect of this aptamer recognition modules in the enzymes activity.

Some promising results were obtained for the fusion constructs but they could not be confirmed. Many factors seemed to affect and obscure results from these experiment. We finally discuss which these factors were and how to overcome them best for future experiments.

2. Materials and methods

2.1. Modelling of enzymes

The modelling of the chimeric proteins (monomers of the tetrameric enzyme) was carried out with the the protein fold recognition server Phyre2\textsuperscript{34}, using the intensive option. The tetramer was built fitting the chimeric proteins with the original pdb structures 2WVA and 4W6Z of Pdc and Adh1 respectively, and energy was minimized using SPDBV\textsuperscript{35}. Sterical clashes were used as an indicator of how the fusion of domains would affect the structure of the tetramer to form the active enzymes.

2.2. Bacterial strains and culture conditions

DH5α ZI E. coli were used for the cloning of plasmids. They were grown in LB at 37°C with the correspondent antibiotic.

Expression of chimeric proteins and ethanol production was carried in T7 Express lysY/Iq E. coli (NEB), here on IqT7, and a ΔadhE strain derived from IqT7. Cells were grown at 37°C in LB 2.5% glucose and 20 mM HEPES buffer with Ampicillin (for fusion and peptide) or Ampicillin and Kanamycin (for DNA and RNA scaffold). Protein expression was induced adding 1 mM or 0.01 mM of IPTG to the media.

2.3. Generation of genetic constructs

Plasmid construction.

pdc from Z. mobilis and adh2 from S. cerevisiae were taken from previous work. New adh1 from S. cerevisiae and slr1192 (adh) from Synechocystis sp. PPC 6803 were amplified by PCR using isolated DNA from correspondent organisms as template.
All interaction domains were synthesized de novo using the sequences from previous work by Dueber et al.\textsuperscript{16} Conrado et al.\textsuperscript{9} and Delebecque et al.\textsuperscript{12,13}. For fusion enzymes short GSG were added as an overhang in the primers used for amplification.

All enzymes genes, \textit{pdc} and \textit{adh2} with the corresponding adaptor domains were inserted in pUCBB in the XbaI and PstI restriction sites, under the influences of pLac promoter preceding each of the genes a Ribosomal Binding Sequence (RBS); this obviously was not the case of the fusion proteins where only one RBS was placed at the beginning of both fusioned proteins. RNA encoding scaffolds were inserted in the low copy number pSB3K3 using EcoRI and PstI under control of a T7 promoter. DNA scaffold constructs were inserted in pSB3K3 either by EcoRI and SpeI or with XbaI and PstI. Four plasmids with the \textit{de novo} synthesized sequences (binding domains, recognition domains and RNA scaffolds sequence) inserted in pUC19 with BamH1 and SacI in order to store and be able to clone the pieces.

\textbf{Overlapping Extension PCR.} De novo synthetized pieces, with overlapping sequences of around 25 nucleotides were amplified by PCR and purified with the DNA Clean & Concentrator\textsuperscript{TM} kit (Zymo Research). The purified DNA was used in a second short PCR reaction without primers. The product was amplified with flanking primers. In every step, proper size of the PCR product were checked in a 1\% agarose gel with Thiazole Orange. When agarose gels showed unspecific bands, inserts were gel purified, running 200 \(\mu\)l of PCR product in a 1\% agarose gel, the band of the right size was carefully cut off and DNA was purified with the Zymoclean\textsuperscript{TM} Gel DNA Recovery kit (Zymo Research).

\textbf{Transformation.} Competent cells stored at -80 °C were thawed on ice, purified plasmids or ligation product were added and kept on ice for 30 min. Competent cells underwent a 42 °C heat shock for 45 s, except in the case of commercially competent cells which, following the makers protocol this heat shock lasted only for 10 s. LB was added and incubated for an hour at 37 °C. Cells were centrifuged supernatant was discarded, leaving around 100 \(\mu\)l. Pellets were resuspended in the remaining LB and plaqued on LB-agar plates with the correspondent antibiotic. Plates were incubated overnight at 37 °C.

\textbf{Colony PCR.} Single colonies were picked and diluted in 10 \(\mu\)l sterile ddH\textsubscript{2}O and used as template for the PCR. The PCR was performed adapting DreamTaq\textsuperscript{TM} DNA
Polymerase (Fermentas) protocol. Positive colonies were grown overnight in LB with the appropriate antibiotic.

2.4. **Deletion of adhE**

Transition of the cytosine to thymine in position 509 found in *Z. mobilis pdc* gene was restored by mutagenesis PCR. *adhE* sequence for designing primers for knocking out the gene was selected from the genome of BL21(DE3). Primers were design with flanking nucleotides of the Sp resistance cassette in the 3’ end homologous recombination sequences for regions at the beginning (nucleotides from 81-120) and the end (nucleotides from 2572 - 2610) of the *adhE* sequence. These primers were used to amplify the Sp resistance cassette.

Knock out of *adhE* gene, replacing most of its sequence with a Sp resistance cassette was carried out by Red/ET Recombinaion (Gene Bridges) according to manufactures.

A colony PCR was carried out using flanking primers of the *adhE* gene, outside the deleted sequence. Positive colonies were grown overnight in LB with Sp and 1ml kept as stock in 80°C with 15% glycerol. A confirmation PCR was carried out in these overnight cultures one more time.

To confirm *adhE* deletion 10 μl of overnight cultures of IqT7 and IqT7 ΔadhE in LB Sp were grown overnight in chloramphenicol and samples were taken for ethanol measurement.

2.5. **Ethanol detection**

Samples from cultures were centrifuge at 13000 rpm in a microcentrifuge, supernatant was used for measuring ethanol. If the concentration of ethanol was too high samples were diluted in ddH₂O.

Ethanol was measured by gas chromatography. Samples were run in a custom packed column with affinity for organic carbons, separating organic molecules mainly by size (Cat #N9305013-ZW5531, Perkin Elmer).

Program was carried out with a flow of 20 ml/m N2 as a carrier gas. 1 ul of sample was inyected with a PKD inyector at 220 °C. Oven was kept 5.5 min with a constant temperature of 130 °C, temperature was raised to 220 °C and this temperature was kept constant for 5 min for eliminating residues from the sample. The standard curve was constructed diluting ethanol in LB to 8 different concentrations (25, 50, 100, 250, 1000, 2500
and 5000 ppm) in three replicates. The standard curve \( y = 5637.8 \times \) with a \( R^2 \) of 0.9968 was created, where \( y \) is peak area and \( x \) ethanol concentration in ppm.

### 2.6. Protein detection

Expression of both enzymes was confirmed by WB analysis. Samples form overnight induced cultures were diluted in dH\(_2\)O to obtain equal OD600. Dilutions were mixed with the same volume of loading buffer with \( \beta \)-mercaptoethanol and heated to 95 °C for 5 min. Samples were run in polyacrylamide gels (200 mV) and stained with Coomassie blue or transferred to a membrane.

Membranes were blocked with TBS-tween with 5 % milk powder during one hour at room temperature or overnight at 4 °C. After washing several times with TBS-tween they were incubated for one or two hours with goat anti-\( Pdc \) IgG \(^{36} \) or rabbit anti-\( Adh1 \) IgG (Rockland). After removing the antibody by washing several times with TBS-tween they were incubated with secondary antibodies HRP conjugated anti-goat from donkey or anti-rabbit from llama. Antibody was removed and washed several times with TBS-tween and reveal with luminol (Biorad).

### 3. Results

#### 3.1. Genetic constructs

To investigate the effect of co-localizing the ethanol producing enzymes \( Pdc \) and \( Adh \) (from \( Z. \) mobilis and \( S. \) cerevisiae respectively), four different co-localization strategies were tested and compared; enzyme fusion, peptide scaffold, DNA scaffold and RNA scaffold (Figure 8). For the fusion strategy, Adh and Pdc were fused together by removing the stop codon of one and inserting a GSG linker between the genes, thereby getting the genes to translate together as one protein. Two plasmid were made where \( pdc \) was fused to \( adh \), one where \( pdc \) was translated first (P-A) and the other were \( adh \) was first (A-P)(Figure 8 top). As a control, a plasmid where the two genes were translated separately was made (P+A).

For peptide binding domains four constructs were created (A-D). Construct A encoded \( Pdc \) with a SH3 linker and for \( Adh \) with a SH3 domain attached; construct B encoded for \( Pdc \) with two SH3 linkers and for \( Adh \) with a SH3 domain attached; construc C encoded \( Pdc \) with an SH3 domain and \( Adh \) with a SH3 linker and construct D encoded for \( Pdc \) with a SH3 domain and \( Adh \) with two SH3 linkers. Therefore we have A and C with a
1:1 stoichiometry, B with two Adh per Pdc and D with two Pdc per Adh (Figure 8 third group from the top). The control strain used was the same used for the fusion strategy, overexpression of both enzymes with no peptide binding domains.

For the DNA scaffold, one plasmid was made where the sequence of the DNA binding motif Zif 268 was attached to the gene sequence of pdc and PSB2 to adh. Four different DNA scaffolds in pSB3K3 had already been synthesized by previous work from other master students, these were named α, β, γ and δ. These scaffolds differ in the proportion of binding sites for Zif268 and PBSII and in the length of the space between them: (α) One binding site for each separated by 4 bp, (β) one binding site for each separated by 12 bp, (γ) one Zif268 binding domain flanked by two PBSII binding domains separated by 12 bp, (δ) one Zif268 binding domain flanked by two PBSII binding domains separated by 12 bp and repeated twice in the plasmid with a 50 bp separation (Figure 8 second group from the top). All strains tested for the DNA scaffolding strategy contained two plasmids, one with the DNA-binding domains fused adh and pdc, and one with a DNA scaffold plasmid. As a control, an Ecoli strain was transformed with the ethanol plasmid and an empty plasmid without any scaffold.

For RNA scaffold strains a construct was created encoding Pdc with the recognition domain for the aptamer BIVtat attached and Adh with the recognition domain for PP7 (Figure 8). Three pSB3K3 based plasmids encoding different dimensions RNA scaffolds were built. Dimension 0 (d0) transcribed a single RNA with both recognition hairpin structures for BIV-Tat and PP7, d1 an RNA with both recognition structures and a cohesive site that produced polymerization forming a chain, d2 encoding two RNA interacting between them forming an RNA tetramer that interacts with other tetramers creating a plane scaffold (Figure 8 bottom). Each RNA strains had the plasmid with the aptamers recognition domains fused to the enzymes and one of the RNA scaffold encoding plasmid

All constructs encoding native or chimeric Pdc and Adh were cloned in pUCBB under the control of an IPTG inducible lac promoter. DNA and RNA scaffolds were inserted in the pSB3K3 plasmid and the RNA scaffolds were expressed under the control of a T7 promoter. For a complete list of constructs made during this study, see Error! Reference source not found. for pUCBB based plasmids and Table 1. Table summarizing all pUCBB derived plasmids encoding for enzymes created during this experiment with its characteristic
features. For each plasmid it is indicated the order in which *adh* and *pdc* are arranged in the construct (1st enzyme/2nd enzyme), the sequence added at the N-terminus of each enzyme (Nt), the amino acid sequence of the linker (linker) and the size expected in KDa in a WB with anti-*Pdc* and with anti-*Adh* (anti-*Pdc*/anti-*Adh*)

<table>
<thead>
<tr>
<th>pUCBB based plasmids containing enzymes</th>
<th>1st enzyme</th>
<th>2nd enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gene</td>
<td>gene</td>
</tr>
<tr>
<td>Control plasmids</td>
<td>Nt</td>
<td>Linker</td>
</tr>
<tr>
<td>P*+A2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P+A2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P+A1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fusion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P*-A2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A2-P*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P-A2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A2-P</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P-A1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A1-P</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Peptide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>SH3 L</td>
<td>GGSGGGSGG</td>
</tr>
<tr>
<td>B2</td>
<td>SH3 L - SH3 L</td>
<td>GGSGGGSGG</td>
</tr>
<tr>
<td>C2</td>
<td>SH3 D</td>
<td>GGSGGGSGG</td>
</tr>
<tr>
<td>D2</td>
<td>SH3 D</td>
<td>GGSGGGSGG</td>
</tr>
<tr>
<td>A1</td>
<td>SH3 L</td>
<td>GGSGGGSGG</td>
</tr>
<tr>
<td>B1</td>
<td>SH3 L - SH3 L</td>
<td>GGSGGGSGG</td>
</tr>
<tr>
<td>C1</td>
<td>SH3 D</td>
<td>GGSGGGSGG</td>
</tr>
<tr>
<td>D1</td>
<td>SH3 D</td>
<td>GGSGGGSGG</td>
</tr>
<tr>
<td>DNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA2</td>
<td>Zif 268</td>
<td>GGSGGGGGSSGGS</td>
</tr>
<tr>
<td>DNA1</td>
<td>Zif 269</td>
<td>GGSGGGGGSSGGS</td>
</tr>
<tr>
<td>RNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA2</td>
<td>BIV-tat</td>
<td>GGSGGGGGGS</td>
</tr>
<tr>
<td>RNA1</td>
<td>BIV-tat</td>
<td>GGSGGGGGGS</td>
</tr>
</tbody>
</table>

Table 2 for pSBK3 based plasmids.
Table 1. Table summarizing all pUCBB derived plasmids encoding for enzymes created during this experiment with its characteristic features. For each plasmid it is indicated the order in which *adh* and *pdc* are arrange in the construct (1st enzyme/2nd enzyme), the sequence added at the N-terminus of each enzyme (Nt), the amino acid sequence of the linker (linker) and the size expected in KDa in a WB with anti-*Pdc* and with anti-*Adh* (anti-*Pdc/anti-*Adh*)

<table>
<thead>
<tr>
<th>pUCBB based plasmids containing enzymes</th>
<th>1st enzyme</th>
<th>2nd enzyme</th>
<th>Anti <em>Pdc</em></th>
<th>Anti <em>Adh</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nt</td>
<td>Linker</td>
<td>gene</td>
<td>Nt</td>
</tr>
<tr>
<td>Control plasmids</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P*+A2</td>
<td>-</td>
<td>-</td>
<td><em>pdc</em> C509T</td>
<td>-</td>
</tr>
<tr>
<td>P+A2</td>
<td>-</td>
<td>-</td>
<td><em>pdc</em></td>
<td>-</td>
</tr>
<tr>
<td>P+A1</td>
<td>-</td>
<td>-</td>
<td><em>pdc</em></td>
<td>-</td>
</tr>
<tr>
<td>P+*A2</td>
<td>-</td>
<td>-</td>
<td><em>pdc</em> C509T</td>
<td>-</td>
</tr>
<tr>
<td>A2+P*</td>
<td>-</td>
<td>-</td>
<td><em>adh2</em></td>
<td>-</td>
</tr>
<tr>
<td>P-A2</td>
<td>-</td>
<td>-</td>
<td><em>pdc</em></td>
<td>-</td>
</tr>
<tr>
<td>A2-P</td>
<td>-</td>
<td>-</td>
<td><em>adh2</em></td>
<td>-</td>
</tr>
<tr>
<td>P-A1</td>
<td>-</td>
<td>-</td>
<td><em>pdc</em></td>
<td>-</td>
</tr>
<tr>
<td>A1-P</td>
<td>-</td>
<td>-</td>
<td><em>adh1</em></td>
<td>-</td>
</tr>
<tr>
<td>Peptide</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>SH3 L</td>
<td>GGSGGGGSG</td>
<td><em>pdc</em></td>
<td>SH3 D</td>
</tr>
<tr>
<td>B2</td>
<td>SH3 L - SH3 L</td>
<td>GGSGGGGSG</td>
<td><em>pdc</em></td>
<td>SH3 D</td>
</tr>
<tr>
<td>C2</td>
<td>SH3 D</td>
<td>GGSGGGGSG</td>
<td><em>pdc</em></td>
<td>SH3 L</td>
</tr>
<tr>
<td>D2</td>
<td>SH3 D</td>
<td>GGSGGGGSG</td>
<td><em>pdc</em></td>
<td>SH3 L - SH3 L</td>
</tr>
<tr>
<td>A1</td>
<td>SH3 L</td>
<td>GGSGGGGSG</td>
<td><em>pdc</em></td>
<td>SH3 D</td>
</tr>
<tr>
<td>B1</td>
<td>SH3 L - SH3 L</td>
<td>GGSGGGGSG</td>
<td><em>pdc</em></td>
<td>SH3 D</td>
</tr>
<tr>
<td>C1</td>
<td>SH3 D</td>
<td>GGSGGGGSG</td>
<td><em>pdc</em></td>
<td>SH3 L</td>
</tr>
<tr>
<td>D1</td>
<td>SH3 D</td>
<td>GGSGGGGSG</td>
<td><em>pdc</em></td>
<td>SH3 L - SH3 L</td>
</tr>
<tr>
<td>DNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA2</td>
<td>Zif 268</td>
<td>GGSGGGGSGGS</td>
<td><em>pdc</em></td>
<td>PSB2</td>
</tr>
<tr>
<td>DNA1</td>
<td>Zif 269</td>
<td>GGSGGGGSGGS</td>
<td><em>pdc</em></td>
<td>PSB2</td>
</tr>
<tr>
<td>RNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA2</td>
<td>BIV-tat</td>
<td>GGGSGGGGS</td>
<td><em>pdc</em></td>
<td>PP7</td>
</tr>
<tr>
<td>RNA1</td>
<td>BIV-tat</td>
<td>GGGSGGGGS</td>
<td><em>pdc</em></td>
<td>PP7</td>
</tr>
</tbody>
</table>
Table 2. Table summarizing all different pSB3K3 derived plasmids: DNA scaffold plasmids (DNA) or RNA scaffold encoding plasmids (RNA). For DNA scaffold plasmids it is indicated the sequence inserted, indicating the order and number of recognition sequences for the Zinc finger domains (Zif268 and PBSII), the number of bp separating the binding domains in the sequence (separation of the binding domains), number of repetitions of the sequence (number of repetitions per plasmid) and the separation between this repetitions. For RNA scaffold plasmids it is indicated the number of RNA sequences encoded by the plasmid, the level of organization of the RNA scaffold and the predicted structure that they will acquire according to previous work by Delebeque et al.

<table>
<thead>
<tr>
<th>pSB3K3 based plasmids containing scaffolds</th>
<th>DNA</th>
<th>RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequence</strong></td>
<td><strong>Separation of binding domains</strong></td>
<td><strong>Number of repetitions per plasmid</strong></td>
</tr>
<tr>
<td>α  Zif268- PBSII</td>
<td>4 bp</td>
<td>1</td>
</tr>
<tr>
<td>β  Zif268- PBSII</td>
<td>12 bp</td>
<td>1</td>
</tr>
<tr>
<td>γ  PBSII -Zif268- PBSII</td>
<td>12 bp</td>
<td>1</td>
</tr>
<tr>
<td>δ  PBSII -Zif268- PBSII</td>
<td>12 bp</td>
<td>2</td>
</tr>
<tr>
<td><strong>RNA</strong></td>
<td><strong>Number of RNA molecules</strong></td>
<td><strong>Organization</strong></td>
</tr>
<tr>
<td>d’</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>d¹</td>
<td>1</td>
<td>Polymerization in a line (1D)</td>
</tr>
<tr>
<td>d²</td>
<td>2</td>
<td>Polymerization in a surface (2D)</td>
</tr>
</tbody>
</table>
Figure 8. Representation of the constructs used for this experiment. Genetic structures are represented on the left. Schematic figures of the synthesized peptides and scaffolds are represented on the right. Yellow arrows-PLac; red arrows-T7 promoters; grey dots-RBS; Red lines-terminators; Red crosses-stop codons; adh-adh1 or adh2 (see table 2); ORF are represented with thicker lines while non-translated sequences of the genes are represented with thin lines.

Fusion constructs: both ORF of pdc and adh were fused together eliminating the stop codons and RBS in between and adding a flexible linker. The resulting peptide is a chimeric big protein with both monomers for pdc and adh joined by a flexible linker. DNA scaffold: a plasmid with the enzymes attached to the zinc finger domains (Zif268 and PBS2) were co-transcribed with four different DNA scaffolds (α-δ) creating four different strains. Peptide constructs: An SH3 domain or one or two SH3 ligands were added to the N terminus of the pdc or adh sequences, resulting in four different strains (A-D). RNA scaffold constructs: A pUCBB plasmid containing Pdc and adh attached to the adaptor domains (BIVtat PP7 respectively) were created. Four constructs encoding three different dimensions RNA were created (d₀, d₁, d²). Three strains were created by co-transcribing these plasmids with the enzymes containing plasmid.
3.2. Protein detection

Cells transformed with the final constructs for enzymes and scaffolds were tested for both Adh and Pdc detection. For Pdc it was possible to detect a band of the expected size in all constructs, nevertheless, every time there were many other bands (Figure 9 top). Most of these bands could be due to degradation, but some of them are bigger than the expected size. Unspecific binding of the antibody might be argued, but consistently, in all the western blot columns, there were no bands in the negative control (Figure 9). Strains with the same Pdc but different scaffold plasmids or associated Adh present the same pattern for these bands, while it was different in others.

Adh antibody was very weak, nevertheless at least a blur band is noticed in all

<table>
<thead>
<tr>
<th>pSB3K3</th>
<th>pUCBB</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
<th>A1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
</tr>
</tbody>
</table>

Figure 9. Western blot with anti-Pdc (top) and anti-Adh (bottom). In the table on top of the gels it is indicated which insert was ligated in pSB3K3 or pUCBB. Abreviations: P represents pdc, A represent adh, A-C represent the enzymes with the SH3 domains and ligands, DNA represents the Pdc and Adh with the zinc finger domains, RNA represents Pdc and Adh with the RNA aptamer binding domains, numbers 1 and 2 indicate if the Adh in the construct is Adh1 or 2 respectively. Size of Pdc and Adh are indicated with an arrow. The expected size for each chimera is marked with a rectangle.
constructs but the negative control (Figure 9 bottom). Degradation is also noticed this time, but either because of the weakness of the antibody or because Adh is more stable, they were less and weaker. All enzymes are at the expected size but for constructs which appear in a band corresponding to a smaller protein than expected. These bands may represent degradation products.

One conclusion that was possible to achieve is that bands for enzymes with SH3 ligands are weaker: Pdc from A1 and B1; and Adh1 from C1 and D1. It was also noticed that fusion with Adh in the amino terminal side presented a weaker band, but also less degradation.

3.3. **Knockout of adhE**

![Figure 10](image)

**Figure 10.** Summary of the strategy used for the knockout of adhE in IqT7. A Sp resistance cassette was amplified with the AdhE flanking regions. High concentrated DNA fragments of the cassette flanked by recombination sequences were introduced by electroporation in a IqT7 strains that had been expressing λ recombination enzymes. As the Sp resistance DNA fragments containing the Sp resistance do not have capacity of duplication only if recombination occurs cells would grow in Sp containing media. To make sure that the original AdhE was deleted we used short primers flanking recombination area (small arrows) for colony PCR. Strains with the original AdhE gene would give a band of 2.7 Kbp while strains with successful recombination would be 1.1 Kbp. False positives by amplification of possible free DNA fragments remaining in the cells was avoid by using primers from outside the recombination area, therefore not matching the DNA fragments.

*E. coli* produces ethanol during fermentation, the final step of this pathway is also the reduction of acetaldehyde to ethanol. In the negative control some ethanol production was observed even during aerobic conditions. Endogenous Adh activity was eliminated in order to measure exclusively the ethanol produced by heterologous enzymes. Knock out was carried by insertion of a spectinomycin resistance cassette in the ORF of the adhE gene (Figure 10).
Colony PCR from colonies grown with spectinomycin showed a band at the expected size for the knock out gene (around 1 kb), clearly differentiated from the wild type adhE (around 2.7 kb). Ethanol measurements were taken from overnight cultures of wild type colonies and ΔadhE colonies grown in the same conditions. Ethanol production from the ΔadhE cultures was nearly impossible to see and much smaller than the wild type culture (Figure 11).

3.4. Fusion

Structure model for fused enzymes

To see if any of the adh and pdc chimeric proteins made in this study could cause sterical clashes, computational models of the fused monomers of pdc and adh1 were built with more than 90 % confidence using the protein fold recognition server Phyre2.

Using native Pdc and Adh as a model, fusion monomers were fitted to form both enzymes (Figure 12). Analysis of steric clashes showed that A1-P monomers (Figure 12 middle right) were clearly better than P-A1 (Figure 12 middle left) for building the Pdc tetramer. For Adh1 tetramer, nevertheless it was not, more than one third of the residues produced steric clashes (Figure 12 bottom). P-A1 appeared to have many steric classes for both enzymes (Figure 12 bottom): there were more than hundred residues clashing and about hundred and fifty clashes predicted for Adh and Pdc respectively. According to this it is very likely that Pdc and Adh activity in P-A strains will be very affected and that activity of Adh will be lower for A-P strains.

The active site of Pdc is closer to Adh in the P-A constructs, but the active site of Adh is closer to Pdc in the A-P constructs. The ideal conformation in order to bring closer both active sites would be binding the N-terminus of Pdc to the N-terminus of Adh, unfortunately, this is not possible using fusion technique.
Figure 12. Structural model for fusion enzymes. Left represents structures from P-A1, right belong to A1-P. On top are the models for the monomers, underneath structures of Pdc and at the bottom structures for Adh1. The only enzyme that is able to be built without presenting sterical clashes is Pdc from A1-P monomers (middle structure from the right), for the rest of the structures many are found. Pdc amino acids are colored in turquoise; Adh aminoacids are colored in red. Structural metal ions are shown in grey. The catalitic region can be spotted for the presence of the analog of pyruvate coloured in CPK.
**Ethanol production by fusion strains**

Production of ethanol was induced overnight with 1mM IPTG and measured from the supernatant of the pelleted cells. Ethanol was measured in fusion strains with a different fusion order (P-A and A-P), using as controls a strain with the empty vector pUCBB as negative control (“-“ in Figure 13) and a strain over expressing both separated enzymes as a positive control(P+A2 or P*+A2 in Figure 13).

IQT7 strains containing the mutated Pdc gene and Adh2 showed that the fusion strain with Adh2-Pdc* (A2-P*) produced around twice the amount of ethanol than the positive control. Pdc-Adh2 produced around the same amount or slightly more than the positive control.

A transition from a cytosine to a thymine was found in pdc to study the effect of this mutation, ethanol production was measured in both the mutated (P*) and the restored Pdc (P) control strains with Adh2. Restored Pdc constructs presented nearly twice the amount of ethanol per OD600. Therefore all constructs were rebuilt with the restored pdc gene.

Production of ethanol was nearly twice times greater than the control for the adh2-pdc* construct (Figure 13 top left). Nevertheless after the mutation was restored it was not

![Figure 13. Ethanol production and growth from fusion constructs in IQT7. Ethanol concentration per OD600 from negative control (light grey), positive control and fusion constructs with mutated pdc (blue) and native pdc (red). OD600 is represented in grey underneath.](image-url)
possible to see any clear effect of the \textit{adh2-pdc} fusion and the \textit{pdc-adh2} decrease ethanol production down to basal levels (Figure 13 top right).

Interestingly when \textit{pdc} was restored final OD600 was smaller for all constructs, especially for the \textit{pdc-adh2} (Figure 13 bottom). This hinted that accumulation of toxic acetaldehyde produced by restored \textit{Pdc} could be greater than the capacity of \textit{Adh2} was able to reduce to ethanol. \textit{Adh1} in \textit{S. cerevisiae} is supposed to be more specific for ethanol production; therefore it was decided to change the \textit{Adh2} to \textit{Adh1}.

Also to detect better the specific effect of the overexpressed enzymes native \textit{adhE}, encoding alcohol dehydrogenase gene from \textit{E. coli}, was deleted in \textit{IqT7} strain. All constructs (fusion, peptide, DNA and RNA scaffolds) with the \textit{pdc} and \textit{adh1} were then transformed in \textit{ΔadhE} strain. Nonetheless when induced cells were grown overnight strains with DNA and RNA scaffolds cells would not grow properly, not reaching OD600 denser than 0.4, which lead in very little production and too much variation. Therefore after analysing which factors in the media could affect the grow induction for IPTG was decreased in all constructs to 0.01 mM (see paragraph 3.9. \textit{Media and growing conditions optimization} and Figure 22). These were the conditions used for growth in all final experiments for the other methods.

When samples were measured after the deletion of \textit{adhE} and induction with 0.01 mM IPTG, ethanol production in the positive control was less than half than reported before and fused enzymes produced in both cases less ethanol than the positive control (bars in red in Figure 14).

**Figure 14.** Ethanol production and growth from fusion constructs in \textit{ΔAdhE}. Ethanol production is expressed in mM per OD600. The bar representing the production of the negative control strain (containing the empty pUCBB plasmid) is in grey. Production of ethanol by the strains containing \textit{Pdc} and \textit{Adh1} is in yellow and the strains containing \textit{Pdc} and \textit{Adh2} is in red. OD600 is represented in grey underneath.
Unexpectedly there was slightly more ethanol produced in constructs with \textit{Adh2} than with \textit{Adh1}. Moreover growth was still very affected, with OD600 smaller than 1 for \textit{pdc-adh1/2} constructs and barely more than 0.5 for \textit{adh1/2-pdc} ones (Figure 14 bottom). These results seemed to point that the endogenous \textit{Adh} activity seemed to be more important than the fusion itself. Very likely heterologous \textit{Adh1} and \textit{Adh2} activity is not optimal and should be optimized.

### 3.5. Peptide

#### Structure model for enzymes with SH3 domains

When the Pdc and Adh enzymes fused to the SH3 domains or ligands were modelled the same way as the fusion enzymes described above, the SH3 domains did not seem to affect the structure of both enzymes (Figure 15). In all cases enzymes fit nearly perfectly the models and no steric clashes were predicted by SPDV, which is very promising. Nevertheless we did not analyse how the structure would look once both enzymes bind together through the SH3 binding domains. Nevertheless the flexible linkers are longer than for the fusion which would allow more flexibility to fit the enzymes.

#### Ethanol production by peptide binding strains

First studies carried out with \textit{Adh2} enzymes in IqT7. Production of ethanol was around the same as the negative control for the strains with A2, C2 and D2. Strains with B showed the best result but still smaller than the positive control.

Analysis of the ethanol content showed that none of the co-localization strains produced more ethanol than the positive control. Before deletion of \textit{adhE} the best producing strain was strain B, all the other strains produced nearly as little as the negative control (Figure 16 up left). Strain B has a stoichiometry of two \textit{Adh} monomers every \textit{Pdc} monomer and it is also the strain where both \textit{Adh} and \textit{Pdc} were detected in the WB, however this WB belongs to later experiments, after \textit{adhE} is deleted. After deletion of \textit{adhE} growth was very affected, especially for C and D, but less than OD600 of 1 for all of them (Figure 16 bottom). Deletion of \textit{AdhE} also affected ethanol production: strain A produced no ethanol at all, strains B and C produced very little (even less than the negative control before deletion of \textit{adhE}) and now it was strain D the one that produced the most ethanol, yet producing only
Figure 15. Structure models for peptide enzymes. *Pdc* and *Adh1* are represented, left and right respectively, with one SH3 ligand (top), two (middle) and the SH3 domain (bottom) bound to its N-terminus. Neither binding domains or ligands attached to the enzymes seemed to interfere with the structure of the enzymes.
half than the positive control (Figure 16 up right). This strain had opposite stoichiometry to strain B, it has two $Pdc$ monomers for every $Adh$ monomer.

Figure 16. Ethanol production and growth from peptide constructs with Adh1, schematic representation of the peptide binding domain interaction is represented on top of each production bar. Ethanol concentration per OD600 is represented as red bars for IqT7 strains and in mustard yellow for IqT7 ΔadhE. OD600 is represented in grey underneath.
3.6. DNA and RNA enzymes without scaffolds

Structure models for DNA enzymes

Structure model for DNA enzymes fused to the Zinc finger structure was modelled as described before for fusion and peptide strains. The zinc finger domains affected in a very different way both enzymes. Zif268-Pdc fitted perfectly with the model producing no clashes (Figure 17 left). Meanwhile PBS2-Adh1 presented around half of its residues with clashes (Figure 17 right).

Structure model for RNA enzymes

The structure of Pdc and Adh fused to the aptamer recognition domains predicated by Phyre2 fitted perfectly with the enzymes. For PP7 the expected beta structure of this domain was predicted (Figure 18 right). For BIV-tat, much smaller than PP7 no defined secondary structure was predicted (Figure 18 left). RNA aptamer binding domains seems to have little effect in the enzymes structure, no
residues were found to produce sterical clashes with SPDV.

**Ethanol production by DNA and RNA strains without scaffolds**

To have an idea of how fusion of DNA and RNA domains themselves could affect the activity of the enzymes, the pUCBB plasmid with these two genes was transformed in IqT7 strains without the pSB3K3 scaffold plasmids. Ethanol was measured then for DNA2 (Zif268-Pdc and PBS2-Adh2) and RNA2 (BIV-tat-PDC and PP7-Adh2) alone. Enzymes with RNA aptamers produced as expected less ethanol than the positive control (RNA in Figure 19 left); however the enzymes with the zinc fingers structures fused to the amino terminus produced more than the positive control, even without DNA scaffold plasmids (DNA in Figure 19 left). It is very unlikely that this is due to unspecific binds to other DNA sequences present in the cell, because these zinc finger domains were chosen because they would not interact with DNA sequences from *E. coli* genome\(^9\). The explanation might be at the protein expression level; the new N-terminus might enhance expression of the enzymes. It could be at the post-translational level, the DNA domains might stabilize the enzymes.

### 3.7. DNA scaffold

When testing the DNA scaffold strains both growth and ethanol production was very small for the strains containing DNA and RNA scaffolds, so it was decided to optimize the media again, reducing IPTG concentration, to try and reach a normal growth (see 3.9. Media and growing conditions optimization)
Enzymes with scaffold plasmids were only co-transformed once adh2 had been switch to adh1 and adhE was deleted. Previous measures were made from strains with only the plasmid with the enzymes without the pSB3K3 derived plasmid so no co-localization was expected in those experiments. Therefore unlike the case of the peptide and fusion, production of ethanol, cannot be compared before and after the deletion. Even with lower IPTG concentrations growth was very small compared to the control strain (Figure 20 right). Absolute ethanol production was around four times less for DNA and the relative production per OD600 was smaller (Figure 20 left). However these results were very inconsistent between repetitions in different days.

3.8. RNA scaffold

Even though the RNA strain produced the least ethanol when tested without the scaffolds (Figure 19) it was the best ethanol producing strain per OD600 after deletion of adhE (Figure 21 top left). Nevertheless, with an OD600 under 0.5, growth was very low for all the RNA strains with scaffolds (Figure 21 bottom left), thus the total amount of ethanol was much smaller than the positive strain.

With such small growth and ethanol production is difficult to obtain conclusions. Nevertheless the results show d\textsuperscript{1} scaffold was the best producing strain with 50% enhancement. The more organized d\textsuperscript{2} scaffold seemed to produce just around the same as the
positive control. $d^0$ was the worst strain, it was very variable and, on average, produced less than the positive control.

---

**Figure 21.** To the left: Ethanol production and growth from RNA1. Ethanol concentration per OD600 from negative control (-), positive control (P+A1) and the RNA constructs, this is Pdc and Adh1 with the aptamer recognition domains (RNA), all with the empty pSB3K3 vector; this three strains were used as controls. RNA co-localization strains which contains pUCBB with RNA1 constructs and pSB3K3 containing the three different scaffolds ($d^0$, $d^1$ and $d^2$) all in IqT7 ΔadhE. OD600 is represented in grey (right) underneath. To the left there is a schematic of the three different scaffolds used for RNA strains. $d^0$ is an RNA molecule with one aptamer for each enzyme; $d^1$ is similar to $d^0$, but with an extra end that can polymerize; $d^2$ is formed by two RNA molecules each recognizing each enzyme that can with each other and between them forming two dimension structures. $d^1$ and $d^2$ had shown to form more organized structures: microtubes and sphere like shapes respectively.

### 3.9. Media and growing conditions optimization

To test how to best grow the E. coli strains for ethanol production and to get a good reproducibility, different concentrations of sugar (0 %, 2.5 %, 5 %, 10 %) and the presence or not of buffer were tested (Figure 23). The results showed clearly that adding sugar increased production of ethanol by a couple orders of magnitude, but different concentrations yield the
same amount of ethanol, showing that at 2.5 % it was already saturated. Addition of HEPES buffer showed around 20 % more ethanol production.

When the final experiment was carried out with cells with deletion of *adhE* and scaffolds growth was so low that it was difficult to measure ethanol. To better understand what was affecting growth the most and try and optimized ethanol production it was analysed the induction level and the concentration of the second antibiotic for the scaffold plasmid (Kanamycin). To study the induction level three different concentrations of IPTG were used (0.01 mM, 0.1 mM, 1mM IPTG). Two kanamycin (km) concentrations were tested (25 ug/ml, 50 ug/ml). Conditions of anaerobiosis were also tested, trying to create a healthier

![Figure 23](image)

**Figure 23.** Media comparison for glucose content (0 %, 2.5 %, 5 %, 10 %) in presence (red) or absence (blue) of HEPES.

![Figure 22](image)

**Figure 22.** Ethanol production by P+A1 with empty pSB3K3 grown in different degrees of anaerobic conditions, two concentrations of Km (25 and 50 mg/ml) and three different concentrations of IPTG.
environment for the cells. Therefore all different media conditions were tested in the semianaerobic conditions used until then (hermetic tubes completely filled with culture, 7.5 ml), even a more gradual anaerobiosis acquiring conditions (hermetic tubes with only 5 ml of culture) and aerobic conditions (culture tubes).

The lowest IPTG concentration yielded the most ethanol (last two groups of columns in Figure 22). Different concentrations of antibiotic did not affect to ethanol production. The semianaerobic conditions used before yielded the most ethanol (red columns in Figure 22), while the aerobic conditions yielded the worst (red columns in Figure 22).

Therefore the only condition changed was the level of induction, from 1 mM to 0.01 mM of IPTG.

4. Discussion

The aims of this master thesis was (i) creation of strains with modified pdc and adh enzymes for co-localization using different methods, (ii) improve ethanol production by achieving channelling effect and (iii) compare ethanol production by these different methods. The first goal was successfully achieved. The second and the third were not, mainly due to differences in expression of the enzymes and to a poor growth of the strains especially after AdhE deletion.

Differences in the level of expression might be caused at the transcription, transduction or post-transduction level.

In all these levels the differences between the N-terminus might be affecting. To homogenise transcription and transduction the use of bicistronic promoters might be helpful in future experiments. These promoters have a small peptide at the beginning of the gene which is translated before the desired gene itself and they have shown to equalize translation of different genes.\textsuperscript{37,38}

Posttranslational modifications could have an effect in stability and activity of the enzymes. Though little is yet known about posttranslational modifications in bacteria, already most of the modifications known in eukaryotes have been found in \textit{E. coli}.\textsuperscript{39}

The poor growth after the deletion of AdhE shows that the activity of Adh from yeast was not sufficient to compensate the activity of the deleted native gene leading in toxic accumulation of acetaldehyde. In fact very low ethanol production was achieved comparing
with previous literature with similar enzymes\textsuperscript{10,36,40}. Overexpression of \textit{Pdc} by itself in \textit{E. coli} has shown to increase ethanol production to similar levels as this experiment overexpressing both \textit{Pdc} and \textit{Adh}\textsuperscript{10}. For this experiment growth conditions were selected to ease reproducibility for comparison of strains rather than to achieve optimal ethanol production. This might be the cause for the low activity of AdhE. Another explanation of low AdhE activity could be the effect of the fused Pdc or domains as fusion of enzymes has shown to decrease activity of AdhB from \textit{Z. mobilis} even if this was compensated by the effect of co-localization\textsuperscript{10}.

As it was a novel approach it is difficult to answer the question if we actually achieved any co-localization during these experiments it is difficult to say. Previous experiments of co-localization of enzymes have shown much better results\textsuperscript{9–12,14,15,17,18,21–23}. The methods for co-localization had been used before, but with different enzymes and different conditions. The closest experiment was fusion of pdc and adhB from \textit{Z. mobilis} but as it was a different and not related alcohol dehydrogenase\textsuperscript{10} it cannot be directly compared especially as acetaldehyde reduction to ethanol seems to be the most problematic reaction in our experiments.

Fused proteins had very promising preliminary results; the strain that was predicted to bring active sites closer was the one with the best results in early experiments. Nevertheless these results were inconsistent in different strains and completely disappeared after deletion of the native \textit{adhE}. Enzymes with SH3 recognition showed in some cases positive results, but these were very small, inconsistent and disappeared completely once \textit{adhE} was deleted.

The use of scaffolds may be interesting as it is possible to arrange the proteins in a specific manner and it allows co-localization of more than one enzyme. Also the expression of scaffolds could be under different promoters than enzymes, making it possible to regulate when the enzymes are co-localized together or not. But use of DNA and RNA scaffolds complicated much more the experiment. It required the use of two antibiotics and seemed to affect the growth of the bacteria. Scaffolds could be expressed in the same plasmid as the enzymes which would simplify the process. But overall seems less efficient than direct binding proteins together.

The answer to which method would be the best for co-localization of enzymes remains still unclear. The co-localization of enzymes itself seems as a difficult method to
generalized as it needs optimization for every particular enzymes and host organism. Better prediction tools for expression, stability and structure of the chimeras are needed for a more predictive design.

**Abbreviations**

**A/B/C/D strains:** Strains containing the pUCBB plasmid with the peptide binding ligands and domains attached to the Adh and Pdc. For more information of the enzymes with peptide binding domains see table1.

- **Adh:** Alcohol dehydrogenase from *S. cerevisiae*.
- **AdhE:** Alcohol dehydrogenase from *E. coli*.
- **bp:** nucleotides bais pairs.
- **d⁰/¹/²:** Strain with a pUCBB plasmid encoding the RNA enzymes and a pSB3K3 plasmid encoding the RNA sequences for the 0, 1 or 2 dimension scaffold respectively.
- **DNA strain:** strain containing only pUCBB plasmid encoding for Zif 268-Pdc and PSB2-Adh, with no scaffold plasmid.
- **IqT7:** *E. coli* strain with overexpression of the lac repressor and T7 polymerase expression under the lac promoter.
- **kb:** kilobase.
- **MW:** Molecular weight.
- **PCR:** Polymerase chain reaction
- **Pdc:** Pyruvate decarboxylase.
- **RNA strain:** strain containing the pUCBB plasmid encoding for BIV-tat-Pdc and PP7-Adh with no scaffold plasmid.
- **SH3:** Src homology 3 domain.
**T7**: RNA polymerase from the T7 bacteriophage.

**WB**: Western Blot.

\((X)1/2\): Construct \((X)\) containing adh1 or adh2 sequence respectively.

\((X)^*\): Construct \((X)\) containing \(pdc\) C509T.

\(\alpha/\beta/\gamma/\delta\) strain: Strain containing both the RNA plasmid and the \(\alpha, \beta, \gamma\) or \(\delta\) scaffold plasmids respectively. See table 2 for information of the scaffolds.

\(\Delta\)adhE strain: IqT7 strain with knock out of the \(adhE\) gene.

**Acknowledgement**

Undoubtedly I am grateful to my supervisor Pia and Elias and my coach Claudia for the guidance they give me and for listening to me as if I had something interesting to say, even if most of the times it was not the case. Thanks also to the rest of Cyanos; I think there is none I did not ask for help at least once. But most importantly for the great environment they created inside and outside the lab, in this matter I have to open the circle to all the people I met at the Chemistry department I will always remember lunch and coffee brakes accompanied with a smile and nice chats.

But I also had people waiting for me outside the lab. Sometimes they had to put up with my horrible stress mood, sometimes I forgot a birthday or a date, or arrived late (even for weddings), sometimes I was so tired that I fell asleep in a coach. The ones that were close had to cope with my company and the ones that I left far away had to cope with my absence worsened with my terrible media communication capacity (this year again using more WhatsApp and Skype is in my New Year’s resolutions). CAFÉS, breakfast, Mc Breakfast and brunches, dinners, nights, skypes, wine nots and whys, kitchen chats and corridor crawls, flights coming and going, holas y holos que son más gordos, cañitas.

Neighbors, lab mates, friends and family, many of you can be in more than one category, I prefer not to say names, you know. To all of you thank you for make me focus and thank you for distract me.
References


