DEVELOPMENT OF A HIGH-THROUGHPUT SCREENING METHOD FOR TRANSKETOLASE AND PROTEIN ENGINEERING FOR BIOTECHNOLOGY APPLICATIONS

A thesis submitted to University College London

for the degree of

Doctor of Philosophy

By

Roberto Icken Hernández López

Supervisor Paul Dalby
Department of Biochemical Engineering
University College London
Gower Street
London
WC1E 6BT

2020

I, Roberto Icken Hernández López, confirm that results and work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in this thesis.

Signature

Abstract

Transketolase (TK) is an interesting enzyme for the biotechnology industry because it can catalyse the formation of specific carbon-carbon bonds with high stereospecificity and selectivity. These characteristics make TK interesting for the formation of high-value chemicals and pharmaceutical intermediates such as those used to synthesise antibiotics and others according to the substrates on the bioconversion. However, its application within large-scale processes is currently limited by low activity on new reactions, and poor stability at the high temperatures often used during industrial processes. One route to overcome these limitations is to use site-directed mutagenesis or directed evolution to improve the enzyme function and stability. The success of directed evolution relies upon designing a suitable screening method that can directly identify the best mutants from large numbers of variants, with the desired set of attributes. This thesis aims to develop an improved screening platform by adaptation of a previous screening method based upon colorimetric reactions. To assess and quantify TK activity towards the conversion of lithium hydroxypyruvate (Li-HPA) and propionaldehyde (PA) to (3S)-1,3-dihydroxypentan-2-one (HK), over a wide range of substrate concentrations. Moreover, several experiments were performed to establish the best conditions to grow *E. coli* for TK production, protein extraction methods and quantification of TK. In addition, PCR conditions were established for the development of mutagenic libraries using the MEGAWHOP. Finally, five different truncated TK variants were generated, all of them showed activity using Li-HPA, glycolaldehyde (GA) and PA as substrates. Results obtained in this project set up the basis to generate TK variants with better stability and activity, screen large numbers of variants using the high-throughput platform developed and finally it was shown that truncated versions of TK could keep its activity.

Impact Statement

Biocatalysis has several advantages for the synthesis of chemicals with potential uses in the food, chemical and pharmaceutical industries. The stereo-selectivity of enzymes make them interesting for the biocatalytic synthesis of high-value compounds such as complex carbohydrates and hydroxyketones that are used as valuable building blocks for preparation of pharmaceutical components such as antibiotics. This thesis was focused on the TK, which is a prominent enzyme that can catalyze the formation of asymmetric carbon-carbon bonds with high stereo specificity and selectivity. TK has been engineered to have significantly altered substrate specificity, however, the application of enzymes into large-scale processes is currently limited by their poor thermostability and activity under bio-catalytic process conditions. In this context, it is necessary to design TK variants that can maintain high activity and stability under different process conditions. On the other hand, directed evolution is one of the most useful techniques to generate variants with these characteristics, and in consequence these generated variants could have a potential benefit to the industrial synthesis of enantiopure pharmaceuticals in biocatalysis and pharma industry. However, the bottleneck to get a suitable variant is the screening method. Results from this work showed basic but fundamental results for TK production using different fermentation conditions, moreover this project was also focused in the development of an effective screening method for selection of thermostable mutants, using a colorimetric reaction that can differentiate the bioconversion product of active TK variants between Li-HPA and propionaldehyde qualitatively and quantitatively. Results from the colorimetric method stablished in this project, were used as a method to identify activities of a range of TK variants and in consequence results have been published in the Federation of European Biochemical Societies (FEBS) (Yu, Hernández et al., 2019). In addition, the high-throughput screening methodology here reported is in process to be published as a significant improvement of previous screening methodologies using tetrazolium salts, however, scientific journal is still in process to be defined. Moreover, conditions of directed evolution for TK variants were stablished to generate larger quantity of mutants for further screening with the method previously mentioned. Finally, five different truncated TK versions were generated, minimising the overall structure, giving the basis for future studies of enzymes evolution. In conclusion, this project has potential results that could be applied even in research studies or industrial processes.

Table of contents

Abstract **3**
Impact Statement **4**
Table of contents **5**
List of Figures **8**
Table of Tables **15**
Abbreviations **18**
Acknowledgements **20**
1 Introduction **23**
1.1 History and Overview of Biotechnology and Biocatalysis 24
1.2 Protein and Amino Acids Structure 29
1.3 Protein Stability, Hydrophobic Effect and Amino Acids 35
1.4 Enzymes and Their Different Production Methods 40
1.5 Recombinant Enzymes and their Importance 43
1.6 Extraction and Purification Methods of Enzymes 46
1.7 Fermentation Types and Modes of Bioreactors 49
1.8 Metabolic and Protein Engineering for Biocatalyst Improvement 53
1.9 Strategies of Protein Engineering 57
1.9.1 Rational Design 58
1.9.2 Directed Evolution 59
1.9.2.1 Error Prone PCR 63
1.9.2.2 DNA Shuffling 64
1.9.2.3 Combined Methods 65
1.10 High-Throughput Screening Methods 67
1.11 Transketolase, Structure and Reactions In vivo 73
1.12 Application of TK 78
1.13 Thiamine Diphosphate (ThDP) the main Cofactor of TK 79
1.14 Development of TK Variants 80
1.15 Possible Uses of Hydroxyketones 81
1.16 Colorimetric Assays and Tetrazolium Salts 83
Objectives **87**
2 Materials and Methods **88**
2.1 Laboratory equipment 89
2.2 LB Media 89
2.3 Terrific Broth Media 90

2.4 Microbial Medium Powder 90
2.5 Magic Media E. coli Expression Medium 90
2.6 Buffers 91
2.7 Antibiotic stocks (1000x) 93
2.8 Agar Plates 93
2.9 Microorganisms 94
2.10 Glycerol Stocks 95
2.11 TK Production 96
2.12 Protein Extraction by Sonication 97
2.13 Protein Extraction by Bug Buster Reagent 97
2.14 Protein Extraction by Freezing-Thawing 97
2.15 Protein Purification by Ni-NTA Columns 97
2.16 SDS-PAGE Protein Gel 98
2.17 Synthesis of Li-HPA 99
2.18 Conditions of Bioconversions 99
2.19 Protein Quantification by Bradford Assay 101
2.20 Colorimetric Method 102
2.21 High-pressure Liquid Chromatography (HPLC) Analysis 102
2.22 Generation of Truncated Transketolases 103
2.23 PCR conditions for Truncated TK variants 104
2.24 Agarose Gel Preparation 105
2.25 Mutagenic Libraries 106
3 Results: TK expression, Purification and Analysis 109
3.1 Introduction 110
3.2 Importance of Media, Buffers and Reagents 111
3.3 TK Growth Curves and Protein Extraction 112
3.4 Fermentations of E. coli for TK Production 114
3.5 Growth Curves of E. coli XL10Gold expressing TK variants G425Z, R520Z, D469E and D469T in LB 115
3.6 Calibration Curves of Total Protein Using Bradford Method 117
3.7 WT-TK Quantification Using Different Disruption Methods 119
3.8 D469E and D469T Protein Quantification 122
3.9 Magic Medium as an Alternative for High Biomass Production 124
3.10 TK Production and Quantification in Different Media 126
3.11 TK Fermentation in 96 and 24 Well Plates 128
3.12 EGTA Buffer as an Alternative for TK Purification 130

3.13 Conclusions 133
4 Colorimetric Method 135
4.1 Introduction to High-Throughput Screening Methods 136
4.3 Conditions for the Colorimetric Method Using WST-1 139
4.4 MP Resin and Absorption Capacity 141
4.5 Calibration Curves of Li-HPA Using the Colorimetric Method 143
4.6 Calibration Curve of (3S)-1,3-Dihydroxypentan-2-one (HK) Using the Colorimetric Method 146
4.7 Calibration Curve of Li-HPA and (3S)-1,3Dihydroxypentan-2-one (HK) in HPLC 149
4.8 Application of Colorimetric Method in TK Bioconversion 151
4.9 Is MP Resin Necessary in the Colorimetric Method? 153
4.10 Relationship Between Colorimetric Method and HPLC 155
4.11 Conclusions of Colorimetric Method 159
5 Development of Mutagenic Libraries 161
5.1 Development of Mutagenic Libraries Using MEGAWHOP Method 162
5.2 Development of Mutagenic Libraries 164
5.3 Conclusions of Mutagenic Libraries 173
6 Truncated Transketolase Variants 174
6.1 Introduction of Truncated TK variants 175
6.2 Generation of Truncated TK Variants 177
6.3 Bioconversion of Truncated Variants 181
6.4 Conclusions of Truncated Variants 184
7 Final Conclusions 185
8 Future Work 187
9 References 189

List of Figures

Figure 1.1. The twenty amino acid side chains specified by the genetic code.

Figure 1.2. The four levels of protein structure. National Human Genome Research Institute (Clark et al., 2018).

Figure 1.3. Schematic diagram showing the packing of hydrophobic side chains between the two α helices in a coiled-coil structure. Every seventh residues in both α helices is a leucine, labeled “d”. Residues labeled “a” are also hydrophobic and participate in forming the hydrophobic core along the coiled-coil (Branden & Tooze 2012).

Figure 1.4. Schematic representation of (a) all α helix, (b) all β sheet, (c) α/β, (d) α + β proteins. (Guruprasad, 2019).

Figure 1.5. Schematic representation of triose-phosphate isomerase built from α helices and β strands of adjacent motifs and connected by loop regions in a three-dimensional structure (Branden & Tooze 2012).

Figure 1.6. The Ni-NTA resin (Nickel - Nitrilotriacetic Acid).

Figure 1.7. Typical variation in working volume and cell substrate concentrations under batch, fed batch and continuous fermentation mode (Fengxue, 2019).

Figure 1.8. Enhancing the thermostability of the lipase from Bacillus subtilis (Lip A) by iterative saturation mutagenesis.

Figure 1.9. Structural model of Bacillus subtilis lipase A (Lip A) displaying the sites A-H chosen for saturation mutagenesis based on high B-factors (Reetz et al., 2007).

Figure 1.10. Reaction in vivo of TK. (A and B) Natural reactions catalysed by TK in the reductive pentose phosphate pathway. (C) Biocatalytic reaction in which Li-HPA is used to make the reaction irreversible (Hibbert et al., 2007).

Figure 1.11. Schematic view of the subunit of transketolase.

Figure 1.12. Schematic view of the dimer of transketolase.

Figure 1.13. Cofactor-protein interactions in the thiamine diphosphate binding site of TK (Schneider et al., 1998).

Figure 1.14. Structure of thiamine diphosphate (ThDP).

Figure 1.15. TK reaction within lithium β- hydroxypyruvate (Li-HPA) and an aldehyde acceptor to produce an α, α'-dihydroxy ketone product.

Figure 1.16. Schematic reaction of the TK bioconversion that produce a 2-hydroxyketone and the colorimetric method between tetrazolium red, which is colourless and the 2-hydroxyketone to form the corresponding formazane which has an intense red colour ƛ~ 485 nm, Smith et al., (2006).

Figure 1.17. Formation of the WST-1 formazan from WST-1 tetrazolium (Ishiyama et al., 1993).

Figure 1.18. Mechanism of reaction of the colorimetric assay between WST-1 (1) and 2 hydroxyketone (2) conversion in an alkaline condition using NaOH 3M.

Figure 2.1. Schematic representation of plasmid pQR791.

Figure 3.1. TK Fermentation curve of *E. coli* WT samples in LB media, in baffled flasks at 37 °C-250 rpm – for 28 hours.

Figure 3.2. TK Fermentation curve of *E. coli* variants G425-stop and R520-stop in LB media, baffled flasks at 37 °C-250 rpm – for 23 hours.

Figure 3.3. Fermentation curve of *E. coli* variants D469E and D469T in LB media. Fermentation in baffled flasks at 37 °C-250 rpm for 24 hours.

Figure 3.4. Calibration curve of protein using Bradford method and BSA as a standard protein. The samples were measured at different times after the addition of Coomassie reagent.

Figure 3.5. Calibration curve of protein using Bradford method and BSA as a standard

protein, measured after 5 minutes of Coomassie reagent addition.

Figure 3.6. (A) Quantification with Bradford method of WT TK (TK lysate) and pure TK using Ni-NTA columns for three different methods of extraction: Bug Buster, Sonication and Freezing – Thawing.

Figure 3.7. SDS-PAGE gel of TK lysate extracted with Bug Busted and pure TK samples from Ni-NTA columns. M: Marker, 1: lysate D469E, 2: pure TK D469E, 3: lysate D469T, 4: pure TK D469T.

Figure 3.8. Fermentation curve of *E. coli* variant D469T (50 mL) using three different growth media Luria Bertani, Magic Media and Terrific Broth in baffled flasks at 37 °C-500 rpm for 25 hours.

Figure 3.9. Fermentation curve of *E. coli* variant D469T, D469E and WT (50 mL) in magic media using baffled flasks at 37 °C - 500 rpm for 24 hours.

Figure 3.10. (A) Total protein quantification using Bradford assay after 24 hours of fermentation at standard conditions, using Tk from D469T in three different growth media, LB, Magic Media, and Terrific Broth.

Figure 3.11. 96 well plate fermentation curve of *E. coli* variant D469T in 1 mL of different growth media in 96 well plates at 37 °C - 500 rpm for 24 hours.

Figure 3.12. Fermentation curve of *E. coli* variant D469T in 3 mL of different growth media in 24 well plates at 37 °C-500 rpm for 24 hours.

Figure 3.13. SDS-PAGE gel of TK samples eluted at different concentrations of EGTA: 5, 10, 15, 20, 25, 50, 75 and 100 mM in tris buffer HCl 20 mM and NaCl 500 mM, pH 8.5.

Figure 3.14. SDS-PAGE gel of TK samples eluted at different concentrations of EGTA: 200, 300, 400 and 500 mM in tris buffer HCl 20 mM and NaCl 500 mM, pH 8.5.

Figure 3.15. SDS-PAGE gel of TK samples: L: lysate, F: filtrate, W: washing, EGTA: EGTA buffer 500 mM, Imida: elution buffer 250 mM of imidazole.

Figure 4.1. Colorimetric assay between Li-HPA and TZC 0.2%, first picture showed the colouration immediately after addition of NaOH. Second picture showed the same reactions after 1 minute.

Figure 4.2. Reaction between different concentrations of Li-HPA (mM), TZC 0.2% and NaOH 3M measured at 450 nm every 22 seconds for 10 minutes at room temperature.

Figure 4.3. Calibration curve of Li-HPA after reaction with WST-1 (0.2%) and NaOH 3M at 600 nm.

Figure 4.4. Calibration curve of different concentrations of Li-HPA (mM) after reaction with WST-1 (0.2%) and NaOH 3M measured at 450 nm.

Figure 4.5. Calibration curve of Li-HPA after reaction with WST-1 (0.2%) and NaOH 3M incubated at room temperature for 60 minutes at 450 nm.

Figure 4.6. Calibration curve of (3S)-1,3-dihydroxypentan-2-one (HK) at low concentrations (1 – 10 mM) after reaction with WST-1 (0.2%) and NaOH 3M incubated at room temperature for 10 minutes and measured at 450 nm (A).

Figure 4.7. Calibration curve of (3S)-1,3-dihydroxypentan-2-one (HK) at high concentrations (10 – 100 mM) after reaction with WST-1 (0.2%) and NaOH 3M incubated at room temperature for 60 minutes and measured at 450 nm (A).

Figure 4.8. Calibration curve of (3S)-1,3-dihydroxypentan-2-one (HK) at high concentrations (20 – 70 mM) after reaction with WST-1 (0.2%) and NaOH 3M incubated at room temperature for 60 minutes.

Figure 4.9. Calibration curve of (3S)-1,3-dihydroxypentan-2-one (HK), 0 – 10 mM in HPLC using UV detection at 210 nm, retention time of product 15.25 minutes.

Figure 4.10. Calibration curve of (3S)-1,3-dihydroxypentan-2-one (HK), 0 – 50 mM in HPLC using UV detection at 210 nm, retention time of product 15.25 minutes.

Figure 4.11. Calibration curve of Li-HPA, 0 – 50 mM in HPLC using UV detection at 210 nm, retention time of product 8.26 minutes.

Figure 4.12. Colorimetric reactions of 1: Li-HPA 200 mM, 2: Product of bioconversion after MP resin, 3: mix of Li-HPA, cofactors and PA after MP resin, 4: sample 3 with no MP resin treatment.

Figure 4.13. Correlation curve between HPLC (x axis) results and absorbance (y axis) after 60 minutes of reaction using the colorimetric method for WT-TK lysate.

Figure 4.14. Correlation curve between HPLC (x axis) results and absorbance (y axis) after 60 minutes of reaction using the colorimetric for quadruple variant (H192P-A282P-I365L-G506A).

Figure 5.1. 1% Agarose gel of Megaprimers (MP) from the first (mutagenic) PCR step. M (Marker), 1, 2, 3, 4 and 5 (Megaprimers 1, 2, 3, 4 and 5) and C (Control) template H192-282P.

Figure 5.2. 1% Agarose gel of MEGAWHOP PCR products. M (Marker), 1, 2, 3, 4 and 5 (MEGAWHOP 1, 2, 3, 4 and 5) C (Control TK template).

Figure 5.3. Sample of a mutagenic library after transformation. Around 1000 of colonies can be observed on LB agar with ampicillin.

Figure 6.1. Each coloured region, corresponding to the different regions of TK, the pyrophosphate (PP)-binding domain (2-322 aa) green, the pyrimidine (Pyr)-binding domain (323-539 aa) blue and the C-terminal domain (540-663 aa) yellow.

Figure 6.2. 1% Agarose gel of PCR's for construction of TK truncated variants. M (Marker), C (TK - H192P/A282P), 1 (Q453Z), 2 (G425Z), 3 (N403Z), 4 (F375Z), 5 (Q349Z).

Figure 6.3. SDS PAGE gel of transketolase (WT-TK) lysate and pure protein using different methods of TK extraction.

Figure 6.4. Li-HPA remained in bioconversion of truncated TK versions with GA at different times.

Figure 6.5. Production of erythrulose in bioconversion of truncated TK versions with GA at different times.

Figure 6.6. Li-HPA remained in bioconversion of truncated TK versions using PA as a substrate at different times.

Figure 6.7. Production of 1,3-dihydroxypentan-2-one in bioconversion of truncated TK. versions with PA at different times.

Table of Tables

Table 1.1. Important discoveries in biotechnology (Hulse, 2004).

Table 1.2. Classification of enzymes (Johannes et al., 2006).

Table 1.3. Advantages and disadvantages of different modes of bioreactors (Riet et al., 1991; Gary et al., 2003; Fengxue 2019).

Table 1.4 Comparison of approaches for engineering enzyme activity (Chica R, et al., 2005).

Table 1.5. Summary of HTP screening methods (Xiao, H., et al., 2015).

Table 2.1. LB Broth composition (tablet).

Table 2.2. LB-Broth composition.

Table 2.3. Terrific Broth composition (TB).

Table 2.4. Microbial Medium Powder composition.

Table 2.5. Lysis buffer composition with lysozyme.

Tables 2.6. Purification buffers for 6xHis-tagged Proteins using Ni-NTA columns (Superflow BioRobot Handbook). pH was adjusted to 8.0 using NaOH.

Table 2.7. Cofactors buffer, 10x.

Table 2.8. Buffers for protein purification using EGTA for Ni-NTA columns.

Table 2.9. Tris-glycine-SDS (TGS) Buffer for SDS PAGE gel (BIORAD).

Table 2.10. Agar LB-Broth.

Tables 2.11. Bioconversion volumes and concentrations.

Table 2.12. Bioconversion volumes and concentrations.

Table 2.13. Truncated versions and sequence showing the stop codons.

Table 2.14. Mix of components for PCR.

Table 2.15. PCR conditions for generation of truncated TK's.

Table 2.16. Mix for PCR products and marker in agarose gel.

Table 2.17. Mix for standard DNA in agarose gel.

Table 2.18. Position and sequence of Megaprimers.

Table 2.19. PCR mix for Megaprimers generation.

Table 2.20. PCR conditions for Megaprimers generation.

Table 2.21. PCR mix for MEGAWHOP generation.

Table 2.22. PCR conditions for MEGAWHOP generation.

Table 3.1. Samples TK-WT from the figure 3.7 (B).

Table 3.2. Protein content of D469E and D469T using two different protein extraction reagents.

Table 3.3. Results of total protein from TK lysates.

Table 4.1. Solutions and volumes used for colorimetric reaction with tetrazolium WST-1.

Table 4.2. Controls of reaction to determine the colouration after mixing different solutions with WST-1 (0.2%) and NaOH (3M) at room temperature.

Table 4.3. Conditions for determination of maximum binding capacity of MP-Carbonate resin for Li-HPA at room temperature after 1 hour.

Table 4.4. Solutions and volumes used for TK bioconversion.

Table 4.5. Solutions and volumes used for TK bioconversion using pure and lysate solution.

Table 4.6. Results of bioconversion after 72 hours.

Table 4.7. Solutions and volumes used for TK bioconversion.

Table 5.1. Average number of mutations of four different mutagenic libraries compared with template H192P- A282P. SD represent one standard deviation of the mean (n=6).

Table 5.2. Observed amino acid mutations generated in each Mutagenic Libraries. H192P-A282P was used as a template.

Table 6.1. Number of colonies in each Petri dish after 16 hours of incubation at 37 °C.

Table 6.2. Total protein in lysate of truncated TK variants.

Tables 6.3. Volumes and concentrations for bioconversions of truncated TK versions.

Table 6.4. Concentration of TK in variants G425-stop and R520-stop after 8 hours of fermentation at standard conditions. Last column shows the protein loaded in SDS gel.

Abbreviations

B-FIT	B-Factor Iterative Test
BSA	Bovine Serum Albumin
DE	Directed Evolution
DHAP	Dihydroxyacetone Phosphate
DHA	Dihydroxyacetone
DI	Digital Imaging
E. coli	*Escherichia coli*
EcoTK	Transketolase from *E. coli*
FACS	Fluorescence-Ativate Cells Sorting
GA	Glycolaldehyde
GC	Gas Chromatography
h	hours
HPLC	High Performance Liquid Chromatography
HTS	High-throughput Screening
ISM	Iterative saturation mutagenesis
IUB	International Union of Biochemistry
IVTC	*In vitro* Compartmentalization
Li-HPA	Lithium hydroxypyruvate
L-ery	L-erythrulose
ME	Metabolic Engineering
MeOH	Methanol
MFCs	Microbial Chemical Factories
min	Minutes

ML	Mutagenic Library
MS	Mass Spectrometry
NADH	Nicotinamide-Adenine Dinucleotide
NMR	Nuclear Magnetic Resonance
PA	Propionaldehyde
PDC	Pyruvate decarboxylase
Ppm	Parts per million
R5P	Ribose-5-phosphate
RD	Rational Protein Design
RE	Recombinant Enzyme
SB	Synthetic Biology
S. cerevisiae	*Saccharomyces cerevisiae*
Ser	Serine
SDSM	Site Directed Saturation Mutagenesis
Tam	Transaminase
ThDP	Thiamine diphosphate
TK	Transketolase
TLC	Thin Layer Chromatography
WT	Wild-type

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Dr. Paul Dalby for his expert advice, extraordinary support during the entire PhD journey and for giving me the opportunity to be part of his team. His guidance helped me all the time of research and writing of this thesis.

My sincere thanks to my second advisor Dr. John Ward for his support and key advices.

I would like to thank my thesis examiners Dr. Frank Baganz and Ewan Main.

Thanks to CONACYT and Biochemical Engineering Department of UCL for the financial support and assistance in making my project come to life, this study might have never happened without them.

I would also like to thank to my lab colleagues Dr. Pierre Affaticati, Dr. Pierre Yves, Dr. Fabiana Subrizi and Dr. Cheng Zhang and academic staff for their important support on this project.

A very special gratitude goes out to my lab friends, especially to Gerardo Santiago and Haoran Yu who helped me out in completing this work.

I owe a sincere thank you to Patrick Mugisha for all his time, advices, and unconditional support every single day in the Department.

I am grateful to my friends Penta Zero Miedo, Rey Fénix, Drago and Misael Rodríguez who I met during this journey and I really admire. All my respect to them.

I want to thank to my friends who gave me personal support during this entire journey, especially to Armando Ruíz, August Bekkers, Edgar Sosa, Esaú Mondragón, Eugenio Ruíz, Gabriel Rojas, José Jiménez, Oswaldo Pérez, Salvador Rea, Servando García, Tonatihu Valdez and Victor Acuña.

Thanks to all my teachers and professors that helped me and inspired me to achieve this goal with a special mention to Dr. Agustín López Munguía-Canales and Eduardo Bárzana García.

I wish to give special thanks to my brother of life Erick Mota for the unconditional support, inspiration, motivation, patience, understanding, advices, time, music, travels, and unforgettable experiences together.

Finally, nobody has been more important for me in the completion of this project than my family members, I would like to thank my parents, sister and brother for all their support, patience and love throughout all my life to pursue my dreams.

To

My mom Elsa María López Ríos who gave me life, education, and love

My dad Jesús Rafael Hernández Aldama who has supported me every single day of my life

My sister Elsy Hernández López and

My brother Hegel Rafael Hernández López

that I love with all my heart and soul

My grandmother Ofelia Reyna Ríos Rodríguez

My unconditional friend Erick Mota

My real friends

My family

I love you all

In memoriam to

My grandmother Margarita Aldama Gallardo

My Olympic Wrestling trainer and hero Zaprian Petrov Doychev

My best friends

Elian Elizabeth Lozano Soria &

Armando Ruíz Medina

My friend, teacher, uncle, and father Armando López Ríos

My lovely uncle José Antonio Hernández Aldama

My champion Cali

1 Introduction

1.1 History and Overview of Biotechnology and Biocatalysis

There are several definitions of biotechnology, nevertheless for this study we can define biotechnology as “a multidisciplinary activity that uses the knowledge of different disciplines (molecular biology, biochemical engineering, microbiology, genomics and immunology) to study, modify and apply biological systems (microorganisms, plants, animals and virus). Biotechnology seeks to make intelligent, responsible and sustainable use of biodiversity through the development of effective, clean and competitive technology, to facilitate the solution of important problems such as in healthcare, agriculture, industry and the environment” (Bolívar, 2004).

Other definitions that are also valid and are important to mention, is the one used by the Organization for Economic Co-operation and Development (OECD): “Biotechnology is the application of scientific engineering principles to the processing of materials by biological agents to provide goods and services” (Bud, 1993). On the other hand, modern biotechnology started in 1976, with recombinant DNA methodologies, chemical synthesis of oligonucleotides and genetic expression, which enabled the production of somatostatin (growth hormone) from synthetic DNA (Bolívar, 2004). However, Biotechnology has a long history and has been part of the culture of ancient civilizations including Babylon and ancient Egypt, 2300 B.C. Those cultures used to have different fermentation processes, mainly for food production as cheese, wine, yoghurt bread and beer. All these fermentation processes were enabled by microorganisms and enzymes without any deep knowledge of the fundaments in the process. At the end of the 17th century, fermentation techniques were described in the book *Fundamental Zymotechnology* by Georg Ernst Stahl, in 1697. This text could be considered as one of the first documents of biotechnology (Bud, 1993).

In the following years, some crucial discoveries in microbiology were made by Louis Pasteur, Emil Christian Hansen, and others. All these discoveries were the beginning to reinvigorate zymotechnology and consequently allowed application of science and technology for production of wine, vinegar, cheese, yogurt, and even in the leather industry for removing hair. At the end of the 19th century, chemicals such as lactic acid, citric acid, and the enzyme takaminase were first made using biotechnology. This was the beginning of a new era in the industry (Bud, 1993).

Biotechnology was, and still is, a fundamental industry for the pharmaceutical and food sectors. The progress from the first fermentations 6000 years ago until now has been dramatic. In the 19th century, science influenced the food and drug industry

to move from using extracts of local plants, to the isolation and synthesis of chemical compounds. On the other hand, biocatalysis has a close relationship with biotechnology, and both started with fermentation processes thousands of years ago.

Biocatalysis has been defined as the use of enzymes, as part of a cell lysate, or whole cell, to convert a molecular substrate (particularly a substrate not known to be transformed by that enzyme in natural systems) into a product (Hughes, Gregory, Lewis, 2018). Millions of years of evolution have generated an enormous diversity of microorganisms and some of them have been able to survive in unusual environmental conditions (temperature, salt concentration and pressure). Such microorganisms are denominated as extremophiles. These organisms are very interesting for the industry, because they produce enzymes that can be active at high temperatures (more than 100 °C), or under extreme pH conditions (alkalophiles or acidophiles), which can be isolated for applications in chemical synthesis and different biotechnological processes (Grunwald, 2018). Application of enzymes includes their use in the food manufacturing, animal nutrition, cosmetics, pharma industries, and as tools for research and development. Table 1.1 shows some interesting examples of the history of biotechnology regarding food and drugs (Hulse, 2004).

Table 1.1. Important discoveries in biotechnology **(Hulse, 2004).**

Years	Historical Fact
4000 to 3000 BC	Egyptians developed grain milling, baking brewing.
2000 BC	Chinese distilled ethanol.
5th/4th Century BC	Hippocrates School of Medicine Hippocratic Oath; 70 medical treatises, 300 remedies. Aristotle classified known plants and animals; Greek work "Pharmacon" means both medicine and poison.
10th/11th Century	Arabians *Materia Medica:* medicinal plants, extracts from wood and tree bark.
17th Century	Anton von Leeuwenhoek (Netherlands) invented a microscope (magnification 30x) described blood corpuscles, cells in fish tissues and "animalcules" (protozoa) present in organic matter in ponds.
18th Century	French pharmacist Pelletier isolated morphine from opium poppy seed and caffeine from coffee berries.
19th Century	Alkaloids first therapeutics isolated: morphine (1806), strychnine (1818), quinine (1820). 1838 Berzelius (Swedish) coined the name "Protein" (Greek πρωτεΐνη – Proteos: "that which comes first" for nitrogenous organic compounds. 1859 Charles Darwin published the book "On the Origin of Species by means of Natural Selection" 1860s – 80s Louis Pasteur (French) proved microorganism are the cause not the result of fermentation in decaying matter. He also revealed the link between bacteria and disease by studying infected silkworms. He also showed bacteria as the cause of acidity in beer. 1880 First pharmaceutical industries in UK. Vaccines for cowpox, cholera, typhus and typhoid developed.
20th Century	More scientific progress in pharmacology, biosciences than in past 5000 years 1921 In Toronto> Insulin isolated from Langherens islets in pig pancreas. 1929 Alexander Fleming accidentally discovered penicillin 1953 Watson and Crick discovered double helix of DNA 1973 First gene cloned 1980 Name "Genomics" first used Increased understanding of genomics, bioinformatics, genetic transformations, molecular breeding, diagnostics, and improved immunology in humans, animals, plants and fish by recombinant DNA vaccines.
21st Century	Human genome sequenced. Nutraceutical: foods purported to contain substances pharmacologically beneficial.

As shown in Table 1.1., biotechnology and biocatalysis has had an important influence on areas as diverse as pharmacy, medicine, food, nutritional products, environmental technology, and others. Enzymes are catalysts, usually proteins that speed up the reactions in living organisms without being a reactant, which means that enzymes decrease the activation energy for the reactions, and the amount of energy that is required for the reaction to occur. It is important to clarify that enzymes do not change a reaction's ΔG (Gibbs energy change) value. In consequence, they do not change the direction of the reaction, or whether a reaction is producing or requiring energy, since the enzymes do not affect the free energy of the reactants or products. In summary, enzymes decrease the free energy of the transition state of a reaction (Robinson, 2015).

Natural enzymes can catalyse reactions with up to 10^{17}-fold rate accelerations, and with specific control of regio-and stereo-chemistry (Hibbert et al., 2005). According to the International Union of Biochemistry and Molecular Biology (IUBMB) nomenclature system, all the enzymes are classified into six classes based in the type of reactions that they catalyse (Table 1.2), and the range of subclasses within each class represents an enormous range of potential chemical reactions that can be catalysed by enzymes.

Hundreds of enzymes are used industrially, over half are from fungi, over one third are from bacteria, from animal around 8% and plant only 4% sources. Over 500 commercial products are made using enzymes. The industrial enzyme market reached U$ 1.6 billion in 1998 and in 2009, the market was U$ 5.1 billion. In the 1980s and 1990s, microbial enzymes replaced many plant and animal enzymes and have been used in many industries including food, detergents, textiles, leather, pulp and paper, diagnostics, and therapy (Sanchez & Demain, 2011). Nevertheless, even when the number of enzymes has been increasing for industrial application, only about 20 enzymes are produced on a truly industrial scale, with Novozymes, DuPont and Roche as the three top companies that cover nearly 75% of the total worldwide enzyme production (Yang et al., 2012).

Table 1.2. Classification of enzymes **(Johannes et al., 2006).**

Enzymes	Type of reactions	Representative subclasses
Oxidoreductases	Catalyze the transfer of hydrogen or oxygen atoms or electrons from one substrate to another	Oxidases, oxygenases, peroxidases, dehydrogenases
Transferases	Catalyze the group transfer reactions	Glycosyltransferases, transketolases, methyltransferases, transaminases, transaldolases, acyltransferases
Hydrolases	Catalyze hydrolytic reactions	Estereases, lipases, proteases, glycosidases, phosphatases
Lyases	Catalyze the nonhydrolytic removal of groups	Decarboxylases, aldolases, ketolases, hydratases, dehydratases
Isomerases	Catalyze isomerization reactions	Racemases, epimerases, isomerases
Ligases	Catalyze the synthesis of various types of bonds with the aid of energy-containing molecules	Synthetases, carboxylases

Enzyme activity under mild conditions is a valuable attribute for producing labile compounds, having profound technological implications as significantly reducing the costs of equipment, energy, and downstream operations. However, the use of enzyme catalysts in organic synthesis has been difficult to adopt by an industry not sufficiently acquainted to deal with biological materials. The bottlenecks of enzyme technology are their high cost, instability and poor performance under reactor conditions, low substrate specificity and requirements of complex cofactors (Dalby, 2011; Illanes et al., 2012).

Biocatalysis is becoming one of the most powerful tools in biotechnology, having a profound social impact on health, food supply, environmental protection, and sustainable fuel production. Moreover, biocatalysis is gaining a prominent place in the present and new scenarios of Biotechnology being nurtured by the advances in several fields as genetics, molecular biology, fermentation technology, bioinformatics, nanotechnology, material sciences, advanced spectroscopy, and others. Major challenges are referred to the production of robust enzyme catalysts at reduced price, being properly addressed by a wide spectrum of technological developments that have flourished in recent decades (Illanes et al., 2012).

1.2 Protein and Amino Acids Structure

Proteins are linear polymers of amino acids (aa) linked by peptide bonds, and the specific biological function of each protein depends on its sequence of amino acid residues within the polymer. The amino acid sequence of a protein contains all the necessary information to dictate its final three-dimensional structure or fold. Most of the proteins have a well-defined three-dimensional structure that is sufficiently stable to fulfil their biological role. Indeed, unfolding or denaturation of a protein typically leads to a loss of biological activity (Guruprasad, 2019). The structures of proteins are generally classified by their topology, which includes the arrangement of secondary structures, and the overall secondary structure content, and is dependent on the sequence of amino acids. The number of possible protein structures is astronomical given the possible combinations of amino-acid sequence, although protein structure classification indicates a relatively few number of topologies that are commonly observed even with very different protein sequences (Sillitoe et al., 2015). Moreover, complete DNA sequences for organisms ranging from bacteria such as *E. coli* to humans suggest that the total number of proteins necessary for life lies in the range of 4200 to 50000, however the number of genes coding for proteins in higher organisms is still not clear (Rayment, 2003).

Proteins consist of sequences of only twenty of the available natural amino acids. Each amino acid has a central carbon atom (C_α) that is attached to a hydrogen atom, an amino group (NH_2), a carboxyl group (COOH), and then a fourth group, the side chain R. It is important to mention that what distinguishes one amino acid from another is the structure of R linked to the C_α through its fourth valence. Amino acids are also classified into different categories depending on the chemical nature of their side chains (see Figure 1.1). Glycine (G) is the simplest amino acid which has only a hydrogen atom as a side chain (Branden et al., 1999). Although there is still much to be learned about protein structure, some fundamental features, folding rules, and structural motifs have been observed and studied in many of the three-dimensional structures determined to date (Guruprasad, 2019).

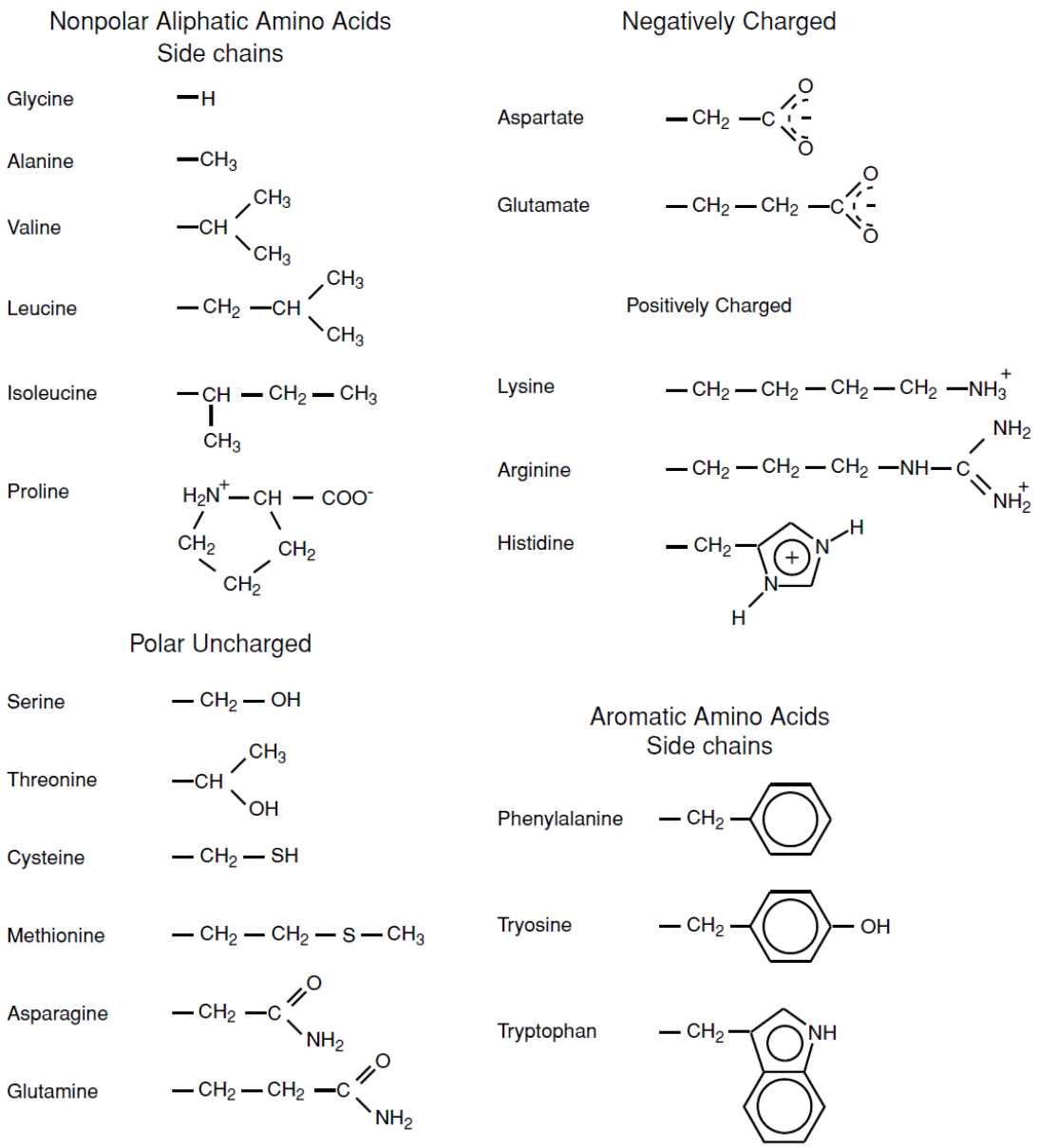


Figure 1.1. The twenty amino acid side chains specified by the genetic code. All except glycine have β-carbon. Proline is technically an imino acid since it is a secondary amine **(Guruprasad, 2019).**

Protein structures are defined using a hierarchical system that links primary structure, to secondary, tertiary and quaternary structure, originally envisaged as a possible pathway through which proteins might fold up into their final structure. The primary structure of proteins is defined as its amino-acid sequence, which represents the chain of amino acids linked by peptide bonds between the carboxyl group of one amino acid and the amino group of the next amino acid, forming a polypeptide. The primary structure or peptide, can fold up to form secondary structures.

There are two dominant forms of secondary structure, namely alpha helices and beta sheets, whose structures are stabilised primarily through maximising the number of internal hydrogen-bonds between the backbone amino and carboxyl groups of the amino acids (Figure 1.2).

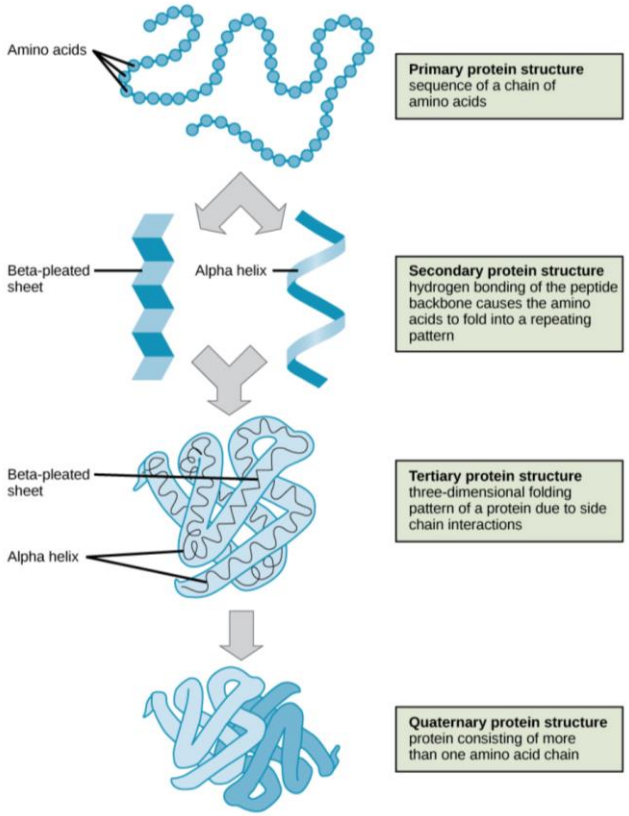


Figure 1.2. The four levels of protein structure. National Human Genome Research Institute **(Clark et al., 2018).**

These structures have relatively low flexibility compared to an unstructured peptide. Alpha helices can be found either completely buried, or at the surface of proteins. Those at the surface are often found to be polar on one face along the helical cylinder, and hydrophobic on the other, with the polar face at the surface of protein structure exposed to solvent, and the hydrophobic one packed against other protein structures (Figure 1.3). They are characterised by intramolecular hydrogen-bonding between every fourth amino acid, and which run in parallel to the helical axis. The side chains of the amino acids project radially outwards from the helix, and are typically found staggered in position relative to the side chains at subsequent turns of the helix. The tightness of the helix is wound such that the length of the hydrogen bonds are optimal, and results in 3.6 residues per turn of the helix.

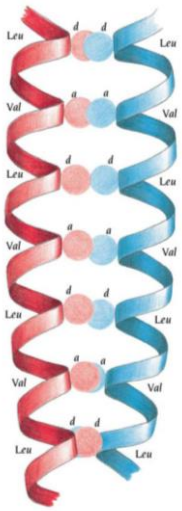


Figure 1.3. Schematic diagram showing the packing of hydrophobic side chains between the two α helices in a coiled-coil structure. Every seventh residues in both α helices is a leucine, labeled “d”. Residues labeled “a” are also hydrophobic and participate in forming the hydrophobic core along the coiled-coil **(Branden & Tooze 2012).**

Beta sheets are comprised of one or more beta strands arranged side by side into sheets, to maximise intermolecular hydrogen bonds, ie. those between strands. Contrary to the alpha helix that is formed by continuous regions of polypeptides, beta sheets can be formed through a combination of several regions of the polypeptide chain (Branden et al., 1999). The beta strands can be arranged either in parallel, or anti-parallel to the neighbouring strand, with respect to the orientation of their N- and C-termini. Anti-parallel beta sheets are far more common, which reflects their higher stability compared to parallel beta sheets, resulting from more optimal hydrogen bonding in the anit-parallel form.

A small population of sequences also form a left-handed helix form. The remaining regions of structure form so-called random coil or loop regions, which are less structured than helices, and so can adopt a many conformations and with a wide range of flexibility, that depends on the degree of interactions and packing against other regions of the structure (Figure 1.4). Loops regions form the connections between alpha helices and beta strands, and most often appear at the surface or solvent exposed regions of the molecule. Unlike helices or beta sheets, loop regions do not optimally maximise the number of internal hydrogen bonds, and so tend to form a greater number of hydrogen bonds to water molecules.

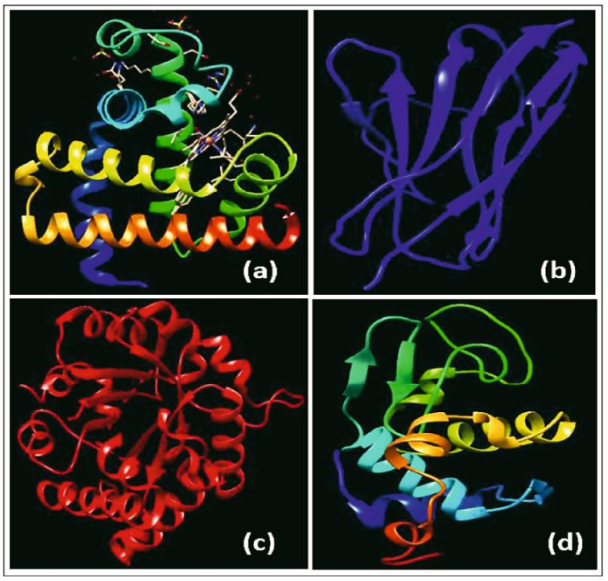


Figure 1.4. Schematic representation of (**a**) all α helix, (**b**) all β sheet, (**c**) α/β, (**d**) α + β proteins. **(Guruprasad, 2019).**

Tertiary structure is the arrangement of multiple secondary structures into a larger three-dimensional structure. This is driven by interactions and forces described in the next section. Finally, the quaternary structure is the union of several distinct polypeptide chains (or subunits), each themselves folded into tertiary structures, that associate into a defined multimeric molecule in a specific way. These subunits can work independently of each other, with a low degree of interactions between them, or otherwise can be highly dependent on the functional state of other subunits, leading to cooperative change sin structure and function between the subunits (Branden et al., 1999).

Besides the hierarchical arrangement of primary through to quaternary structure, structural biologists have classified a number of other structural elements, including domains, motifs and super secondary structures. Domains are units of tertiary or quaternary structure that can fold and function independently, often as discrete compact globular structures. Mulitple domains can be joined together through linking loop regions.

Motifs and super secondary structures are combinations of secondary structures with specific geometric arrangements that are commonly seen in protein structures (Figure 1.5). Super-secondary structures are small elements, such as the combination of two alpha helices joined by a loop region. By contrast, motifs are larger elements of structure that have commonly observed topologies, and include alpha-

helix bundles, the Rossmann fold, Greek key, beta-barrels, and EF-hand structures.

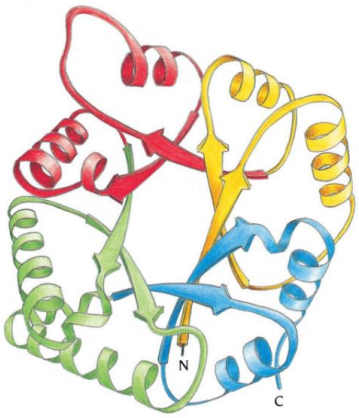


Figure 1.5. Schematic representation of triose-phosphate isomerase built from α helices and β strands of adjacent motifs and connected by loop regions in a three-dimensional structure **(Branden & Tooze 2012).**

1.3 Protein Stability, Hydrophobic Effect and Amino Acids

Protein stability is a term that refers to the free energy difference between the folded and unfolded state of the protein in solution. The free energy difference between these states is typically between 20 and 80 kJ/mol, which corresponds to the magnitude of five to twenty hydrogen bonds in secondary protein structures (α-helix and β-sheet) (Sheh-Yi et al., 2003.). These few net interactions are enough to prevent spontaneous unfolding at room temperature. In practice, this small net free energy difference is a precarious balance between the enthalpy change from a large number of specific interactions, and a large change in entropy due to changes in conformational freedom of the polypeptide, and due to desolvation effects.

One of the main challenges in protein engineering is the design of proteins with enhanced thermostability and activity, that are directly derived from a detailed understanding and calculation of underlying thermodynamic principles. Protein stability is determined by an enormous number of interactions (peptide bonds, ionic bonds, disulfide bonds, hydrogen bonds) observed in the folded state. Protein stability is a complex problem since every amino acid residue has several possibilities of interaction such as polar interaction via the peptide bond, and a variety of ionic, polar, and nonpolar interactions through its twenty different types of side chains.

Hydrophobic effects have been accepted to be one of the main driving forces (poor interaction of hydrocarbons with water) for protein folding. The hydrophobic effect can be defined as “the process by which non-polar groups are removed from contact with water” (Doig & Williams, 1990; Kronberg, 1991) and it is considered as the major stabilizing force in proteins, membranes, and nucleic acids. In other words, hydrophobic molecules isolate themselves in the interior of the protein to stay away from contact with water (Guruprasad, 2019). The hydrophobic effect thus describes the energetic preference of non polar molecular surfaces to interact with other nonpolar molecular surfaces and thereby to displace water molecules from the interacting surfaces. As a consequence, protein folding tends to lead to most (but not all) hydrophobic side chains becoming buried in the interior of the protein (De Jesus, 2017; Rayment, 2003). This is partly driven by van der Waals interactions formed between hydrophpbic atoms. However, another major driver of the free-energy change in the burial of hydrophobic residues is the entropy associated with desolvation. Water molecules are highly ordered around exposed hydrophobic surfaces, and these water moleules become disordered upon desolvation and dispersion into the bulk solvent.

Overall, these changes in enthalpy and entropy must offset the lowering of entropy due to a structuring process of the polypeptide. Generally, the enthalpy change ΔH is of the same order of magnitude as the change of TΔS, the product of temperature (in Kelvin) and entropy for the process of folding in water. For enzymes used in biocatalysis, the stability in organic solvents is also of interest. Experiments have shown that the entropy change upon transferring a hydrocarbon from a non-polar environment into water is large and negative (Kronberg, B. 2016) such that

$$\Delta G \text{ (large and positive)} = \Delta H \text{ (small or zero)} - T\Delta S \text{ (large and negative)}$$

All proteins are formed from 20 amino acids as it was specified in Figure 1.1. These amino acids exist as L-isomers, that are the only ones that are utilised in protein biosynthesis, wheres the D-isomers have only occasionally been found in nature as residues in proteins. The use of only one stereoisomer of the amino acids establishes a structural uniqueness that is essential for biochemical specificity. The genetic code classifies the amino acids into four classes: (1) aliphatic amino acids, (2) aromatic amino acids, (3) polar amino acids, (4) charged amino acids. These groups of amino acids provide the range of properties necessary to create a stable, functional, and folded protein (Guruprasad, 2019).

The aliphatic amino acids that include glycine, alanine, valine, leucine, isoleucine, and proline provide the range of a small hydrophobic amino acids necessary to fill the gaps in the interior of the protein. Due to the fact that glycine is the smallest amino acid and it is unique because it lacks a side chain, it serves a special role in the protein structure, and provides more conformational freedom than any other amino acid. Glycine is often found in turns and loops due to its chemical structure, whereas other amino acids would be sterically unacceptable. Phenylalanine, tyrosine, and tryptophan are large aromatic residues that are normally found buried in the interior of a protein and are important for protein stability.

In addition, the polar amino acids that include serine, threonine, cysteine, methionine, asparagine, and glutamine provide many of the functional groups found in proteins. These could serve as a nucleophile in many enzyme sites, metal ligand (particularly for iron an zinc), sites of phosphorylation and glycosylation, which are important for enzyme regulation and cell signaling and its ability to form disulfide bonds, which often make an important contribution to the stability of the extracellular proteins. On the other hand, charged amino acids that include aspartate, glutamate,

lysine, arginine and histidine are relatively abundant and are important for making proteins soluble, these residues are frequently located on the surface of the protein unless they play a specific biological role. These amino acids could also work as general acids or bases in enzyme catalysed reactions and metal ions ligands (Guruprasad, 2019).

Some studies have shown that the stability of a protein can be increased by rational modification of protein structure in one amino acid substitution that could decrease the configurational entropy of unfolding. In internal helical positions, alanine has been reported as the most stabilizing residue, whereas glycine and proline are the most destabilizing amino acids. This difference in stabilization can be explained by various physical reasons such as the hydrophobic effect and the entropy folding. Alanine consistently stabilizes the helical conformation relative to glycine because it buries more apolar area upon folding and because its backbone entropy is lower. In alanine-based peptides, alanine has been reported to be 2 kcal/mol more stabilizing than glycine, whereas in other systems the difference is of just 0.7 kcal/mol. (López-Llano et al., 2006).

As it was mentioned in Figure 1.1, glycine lacks a β-carbon and has more backbone conformational flexibility than alanine, this means that the backbone of a glycine residue in solution has greater configurational entropy than alanine. In consequence, more free energy is required during the folding process to restrict the conformation of glycine than alanine. The stabilization energy introduced by mutation of Gly to Ala and the change in solvent-accessible surface of the protein, strongly suggests that interactions of alanine in the folded protein plays the dominant role in the stabilization of protein helices. In this context, the stability of some proteins should increase through the replacement of some glycines with alanines, because Ala increases the hydrophobic effect (or with other residues containing a β-carbon). However, it is necessary to consider that not all the glycines are potential sites of substitution and an analysis must be done to avoid the introduction of unfavorable steric interactions in the mutated protein (Matthews et al., 1987; Scott et al., 2006). One successful example was the substitution of glycine to alanine from *Bacillus stearothermophilus* that demonstrated to provide protein stabilization by rational amino acid replacements based on the concept of entropic stabilization (Hencht et al., 1986).

Some other factors also influence the stability and conformation of the protein structure, such as hydrogen bonds, which involve an interaction between a hydrogen

attached to an electronegative atom and another electronegative acceptor atom with a free pair of electrons. These important interactions are dipolar in nature, and primarily formed between nitrogen and oxygen atoms as the most common electronegative atoms in biological systems. Most of the hydrogen bonds in proteins occur in networks where each donor participates in multiple interactions with acceptors and each acceptor interacts with multiple donors. In principle, every hydrogen donor and acceptor in a protein can form an interaction with the folded structure or the external solvent. Proteins maximise the internal number of hydrogen bonds, and minimize the number to solvent where possible, because internal hydrogen bonds are very slightly more stable thermodynamically, than those with water (Rayment, 2003). The total number of hydrogen bonds do not contribute to a significant stability of the protein, as discussed above, due to counteracting entropy changes. However, they do provide the main driver of protein conformation through the formation of α-helixes and β-strands that are in protein structures (Guruprasad, 2019).

Several authors have reported that disulfide bonds are able to enhance protein stability, however the addition of a new disulfide bond is not always found to increase the stability of the protein compared with the wild type (Zhou, Kay, & Hodges,1993). The natural presence of disulfide bonds adds considerable stability to the folded state, and in many cases reduction of the cystine linkages is sufficient to induce unfolding. The introduction of disulfide bonds reduces the entropy of the unfolded state by reducing the degrees of freedom available to the disordered polypeptide chain. This helps to stabilize the folded state by decreasing the entropy difference between the folded and unfolded state. Although this strategy looks a simple task, the geometry of the disulfide bond is rather restricted, and some other considerations should be made before the introduction of disulfide bonds aimed at improving protein stability (Guruprasad, 2019). Disulphie bonds are ideally introduced in the most closely packed part of the protein interior. In addition to the position of the disulfide bridge, the distance between the two sulfur atoms, and the angle between the sulfide and β-carbon of each Cys residue, are determinant factors to improve the protein stability. On the other hand, disruptions of pre-existing interactions in the WT structure due to the introduction of new cysteine residues must be avoided. Finally, it must be considered that the replacement of larger hydrophobic residues could significantly change the protein structure and in consequence decrease the protein stability.

Salt bridges or ion pairs are the association of two oppositely charged ionic groups in a protein. These interactions do not have a high impact on the overall protein

stability due to the fact that the isolated ionic groups are so effectively solvated by water. As a consequence only a few unsolvated salt bridges are found in the interior of proteins. Finally, the dipole-dipole interactions known as Van der Waals interactions, are weak interactions that arise from the close association of permanent or induced dipoles. Although these forces are very weak, the high number of interactions in a folded protein allows them to contribute significantly to the overall stabilization of proteins, and to the interactions between proteins and their complementary ligands such as proteins or small molecules (Guruprasad, 2019).

1.4 Enzymes and Their Different Production Methods

The term *enzyme "in yeast"* was described for the first time in 1876 by W. Kuhne (Kuhne, 1876). Enzymes are naturally tailored protein catalysts synthesised to perform under physiological conditions. Enzymes are natural catalysts that can be produced from any living organism to increase the rate of a massive and multiple set of chemical reactions required for life. The importance of enzymes lies in the fact that they are involved in all the processes needed for life, for example protein synthesis, DNA replication and transcription and metabolism. Enzymes are commonly used in food processing, production of food ingredients, as tools for production of medical products that can be used for diagnosis, prevention, and treatment of diseases, being the production of antibiotics one of the most common examples. Moreover, one important characteristic of enzymes, is their capacity to execute very specific chemical and biochemical reactions, making them increasingly convenient in industrial processes (Steele et al., 2008; Yang, et al., 2012).

Up until 1970s, most of the commercial application of enzymes required plant and animal sources. At that moment, the use of the enzymes was predominantly for food processing industry. Importantly, as the enzymes were extracted from animals and plants, they were considered free from problems of toxicity and contamination. This is not the case when enzymes are of microbial origin. Nevertheless, as demand increased and as fermentation technology developed, economical, technical, and ethical advantages of the microbial enzymes were recognized (Robinson, 2015). Biotransformations imply the use of enzymes under conditions that are not the natural or ideal conditions and that could affect the structure and functionality of the protein. Thus, the challenge is to develop enzymes that can preserve their functions and properties, even under harsh process conditions. Their specificity and stereoselectivity make enzymes very interesting and well appreciated for the synthesis of pharmaceuticals and fine chemicals (Woodley, 2008).

The use of microorganisms provides higher quantities of enzyme, in a shorter time and in a small production facility. A historical example is the production of chymosin also called rennin (a milk coagulating natural enzyme used for cheese production) that was traditionally obtained from the stomach of a calf (a young cow still feeding on its mother's milk). The obtention of 10 kg of rennet implies several months of intensive farming to produce a calf, without mention the ethical issues that surround the use of animals. On the other hand, a 1000-litre fermenter of recombinant *Bacillus*

subtilis can produce 20 kg of enzyme in 12 hours. Indeed, most of the cheese available nowadays are made from milk coagulated with microbial enzymes. Moreover, another advantage of microbial enzymes are their extraction methods. Many of the enzymes that are used for biotechnological application are produced extracellularly, making them much easier to obtain than the equivalent in animals or plants that usually require several steps of extraction and purification (Robinson, 2015).

A further advantage is that the microorganisms can be produced in the same fermentation facility for production and extraction, whereas animals and plants need to be transported to the extraction facility. Additionally, commercial enzymes derived from plants and animals are usually located within only one organ or tissue, making the rest of the material a waste product. Lastly, enzymes from plant and animal sources generally have yield differences and are not available at any time of the year, a problem that is completely discarded using microorganisms. In terms of technical advantages, microbial enzymes are more stable, for example the stability of enzymes from thermophilic microorganisms. Microorganisms are very amenable to genetic modification with the use of molecular biology techniques, making them suitable for production of novel or altered enzymes. Contrary to the genetic manipulation of plants and animals that are more expensive, difficult and is still a controversial topic in different countries including the U.K. (Robinson, 2015).

Another important characteristic of enzymes is that they can be promiscuous catalysts, which means they can catalyse several reactions and/or transforming many substrates, in addition to the ones for which they are physiologically specialized or evolved (Khersonsky & Tawfik, 2010). One of the main problems in the use of enzymes for biocatalysis is their poor operational stability at high temperatures (around 60°C) or in processes with organic solvents. Johannes et al., (2006) claimed that the main disadvantages of biocatalysis are their susceptibility to substrate, their operating range being limited to neutral pH, their denaturation at high temperatures, as well as their preferred activity and solubility in water. For biotransformation in industrial processes with poorly soluble substrates, it is often necessary to include organic solvents, which does not provide the best environment for most enzymes. Water is also often not a desired solvent in chemical synthesis as it is difficult to remove due to its high boiling point and high heat of vaporization. However, protein engineering and directed evolution have shown that it is possible to improve the activity of enzymes in organic solvents (Johannes et al., 2006).

Despite the enzyme constraints, many complementary strategies were developed to improve their performance; protein engineering could be the answer for the next generation of enzymes with desirable characteristics, based on physical modifications of the protein. Moreover, it can be used to achieve the production of cheaper and more robust enzymes for industrial use in several commercial sectors as food, cosmetics, agrochemicals, and pharmaceutical (Ran et al., 2008). Biocatalysis can bring important advantages such as high stereo-selectivity, regio-selectivity, and chemo-selectivity. Along with their high catalytic efficiency, such features are considered advantageous compared with traditional chemical synthesis, due to their reduced need for protecting groups. Finally, the reactions are environmentally friendly due to their operation in water, and at ambient temperature and pressure (Johannes et al., 2006).

1.5 Recombinant Enzymes and their Importance

Native enzymes could be improved using protein engineering to get recombinant enzymes (RE) that have even more advantages compared to the native enzymes, although both are manufactured to meet the same rigorous quality control standards. Recombinant enzymes are formed by expressing foreign genes in a host cell. RE have shown several advantages like production of a purer product with less processing time compared with the native enzymes. This is translated to significant higher yields of recombinant enzymes than native enzymes, greater consistency, and less lot-to-lot variation. Enzyme producers like New England BioLabs, have reported that introduction of recombinant enzymes resulted in lower prices per unit, which allows for the consumers substantial savings. For example, the thermal stable enzyme Vent® DNA Polymerase shows 5x higher fidelity than native *Taq* and an improvement of thermostability with a half-life of 6.7 hours at 95° C for maximum activity during PCR compared to the wild type.

Moreover, medical biotechnology has provided multiple technologies and products for health care. These technologies have been providing significant changes in the treatments of diseases such as the recombinant protein of human insulin in 1982. More than 130 recombinant proteins have been approved by the US/FDA for clinical use, currently produced and used in medicine worldwide. Some of these recombinant proteins are enzymes, or contain enzyme domains, including blood clotting factors (e.g Factor Xa), clot-breaking agents (eg. tissue plasminogen activator), asparaginase (Erwinase), and even botulinum neurotoxin (BoNT), which each improve human health and life quality (Pham, 2018).

As it was mentioned before, chymosin was the milestone in the history of food-processing enzymes for the Food Industry, due that it was the first recombinant enzyme regulated by the U.S. government for the use in cheese and other dairy products. In 1990, FDA affirmed as a Generally Recognized As Safe (GRAS) the chymosin enzyme preparation derived from *E. coli* K-12. The chymosin is a recombinant enzyme that contains the bovine prochymosin gene inserted in *E. coli* K-12 and it is accumulated in the form of inclusion bodies that are subsequently extracted, isolated solubilized, purified and finally converted into chymosin by acid treatment (Zofia et al., 2006).

Another area of major importance for enzymes is in industrial biocatalysis. Traditionally, chemical processes transform a low-value starting material into a high-value product through a series of unit operations. This kind of processes require several steps to remove contaminants, performing reactions, and an additional step to separate the product of interest from unreacted substrates and byproducts (Fisher, 2014). With the recent advances in biomolecular engineering and synthetic biology, several biologically and active chemical compounds have been successfully produced using enzyme catalysed approaches, and also whole cell-based approaches in the last few decades. Many of these high-valued chemicals are very attractive in the industry as many of them have been demonstrated to have important pharmacological activities or biotechnological significance (Guo et al., 2017).

Production of industrial or research enzymes must be controlled and produced under sterile conditions and Good Manufacturing Practices to avoid any contamination during the fermentation, and the selected microorganisms for enzyme production must be well defined. In the laboratory, a master culture must be labeled and stored on strict aseptic conditions with a specific low temperature to avoid degeneration and securing genetic stability. All the aseptic conditions must be kept during all operations to avoid contamination, formation of undesirable products, and deterioration. One of the main advantages of enzyme production via fermentation is the constant quality of the product and a high production yield under specific conditions for each strain, with the objective of increasing as much as possible the amount of enzyme. Enzyme activity and operational parameters such as temperature, pH and oxygen content must be monitored during the process and kept in a predetermined range based on experiences.

It is important to clarify that even when the enzyme was produced by a Genetically Modified Micro-Organisms (GMMs), the enzyme itself is not and cannot be a GMM because enzymes are substances and not organisms. Moreover, the commercial enzymes are listed by the Association of Manufacturers and Formulators of Enzymes Products (AMFEP), this list shows the enzyme, organism for production, original microorganism or donor and if is authorized for application in the food industry, being *Aspergillus sp., Bacillum sp.,* and *Saccharomyces sp.,* the organisms that produce the highest number of enzymes (Association of Manufacturers and Formulators of Enzymes Products, 2020). The AMFEP in collaboration with the European Commission, decided to classify the food process in which enzymes are used, this classification helps in the practical implementation of enzymes. The food

processes that involves enzymes are baking, brewing, dairy, egg, fats and oil, flavour, fruit and vegetable, protein, potable alcohol, grain and starch, wine, and yeast processing. Application of enzymes in the food industry is enormous, and each process area have many specific examples. In the case of fruits and vegetables, the use of cellulases, beta-glucanases, xylanases and pectinases are necessary for improving the process of solid-liquid separation, filtration, depectinisation and concentration of the raw material to produce juices, jams, marmalades etc. In the case of dairy processing, enzymes can be used to hydrolyse lactose to produce lactose reduced dairy products.

An industrial example of enzymes developed by the combination of microbial screening and rational protein engineering is the second-generation detergent enzymes that includes the development of novel amylases. This enzyme has been upgraded by enhancing its activity at lower temperatures and alkaline pH, while maintaining the necessary stability under detergent conditions. In addition, even when the enzymes like proteases have been isolated from nature, they have been modified in the laboratory by directed evolution to keep their activity at low temperatures and alkaline pH. Another example of enzymes that have been modified via protein engineering is the new α-amylases that are used in the industrial bioprocess for the conversion of starch to high fructose corn syrup. This modified α-amylase has been optimized to improve its thermal stability, acid tolerance and ability to function without the addition of calcium (Ole et al., 2002).

1.6 Extraction and Purification Methods of Enzymes

Enzymes can be used as whole cell lysates or in purified form, and their use will have different advantages depending on their application process. Lysate is typically used when it is difficult/costly to purify the enzyme, when the biotransformation requires multiple enzymes, or where it is necessary to recycle the cofactors, such as in redox reactions (Johannes et al., 2006). Nevertheless, due to other metabolic processes, undesirable reactions and products may be formed during biotransformation or cell growth, and sometimes these products can be toxic to the cell or difficult to separate from the rest of the cell culture. Moreover, the cell membrane can act as a mass transport barrier between the substrates and the protein, making whole-cell biocatalysis inefficient. On the other hand, the main advantage of purified enzymes is that the reaction has higher productivity with fewer side-reactions due to the absence of other enzymes. Nevertheless, enzyme purification must be analysed to know if it is suitable by considering several factors that are explained below (Sakkos et al., 2019).

Firstly, the cost of purification of the enzyme must be considered. Diagnostic reagents and clinical therapeutics require a extremely pure enzyme and thus have high purification costs. In contrast, some bulk industrial enzymes do not require a high level of purification. This is the case of proteases used in washing powders. Their low enzyme purification costs account for only 5 to 10% of the final product value. (Robinson, 2015). Once the level of purity required is known, the purification procedure can be modified to suit. As one might expect, different laboratory techniques have been optimized to remove contaminating material in order to obtain pure enzyme and in consequence increasing its specific activity, these are discussed below:

Once the fermentation process has achieved the optimal harvest point (i.e. produced the most enzyme), the microorganisms must be inactivated (generally decreasing the temperature to 4°C to prevent further microbial growth and stabilize the enzyme product) and isolated from the fermentation broth, usually by centrifugation at different speeds depending on the microorganism used. The pH must be measured and adjusted if it is necessary to optimize enzyme stability. The process to extract the enzyme may vary if the enzyme is intracellular for example transketolase, glucose oxidase, glucose-6-phosphate dehydrogenase or cholesterol oxidase that requires more extensive purification treatment or extracellular like several industrial enzymes for example cellulase, dextranase, protease, α-amylase and amyglucosidase (Association of Manufacturers and Formulators of Enzymes Products, 2020).

For intracellular enzymes, cells must be ruptured using sonication, osmotic shock, freeze-thaw cycles, enzymatic removal of the cell wall, ball milling or high pressure using French press to release the cell content. Some of these processes are explained as follows. Separation of enzyme normally involves separation of the protein by different processes as salting out or salt induced precipitation, that requires the addition of ammonium sulphate salts to increase the ionic strength of the solvent. This method produces an interaction of water molecules with salt ions and in consequence decreases the number of water molecules available to interact with the protein. If the protein is not able to interact with the solvent, it interact with each other and finally form a precipitate that comes out of the solution. Additional steps like centrifugation and chromatography must be performed to get the pure enzyme (Robinson, 2015).

Ion-exchange chromatography is an effective technique in the early stages of the purification process. This technique separates charged biological molecules such as proteins based on their affinity to the ion exchanger. It depends on the pH of the solution, proteins may carry a net positive, negative or no charge. Protein solution must be added into a column containing an insoluble polymer with a specific net charge that will determine the type of mobile ion it will attract. All proteins that are oppositely charged to the column will bind to it and all the other proteins with the same charge of the column will pass through it. Finally, change of pH or introduction of a salt solution must be made to alter the electrostatic forces, allowing the retained protein to unbind and elute from the column (Robinson, 2015).

A particularly useful technique is Immobilized-Metal Affinity Chromatography (IMAC), developed by Porath in 1975. IMAC is based on the interactions between a transition metal ion (Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}) immobilized on a matrix and specific amino acid side chains. Histidine is often used as it exhibits the strongest interaction with immobilized metal ion matrices. Proteins with a tail of at least 6 histidine residues are efficiently retained on IMAC matrices or columns containing a nickel-nitrilotriacetic acid (Ni-NTA) agarose resin allowing a specific separation (Figure 1.6). Elution of the protein can be easily made adding an imidazole solution to the column. The main advantage of this technique is that it can be utilised to rapidly purify a His-tag protein, giving 100-fold enrichments in a single purification step with a high yield up to 95%. One disadvantage of this method is that it is suitable only for proteins expressed at high levels (Falke, 2010).

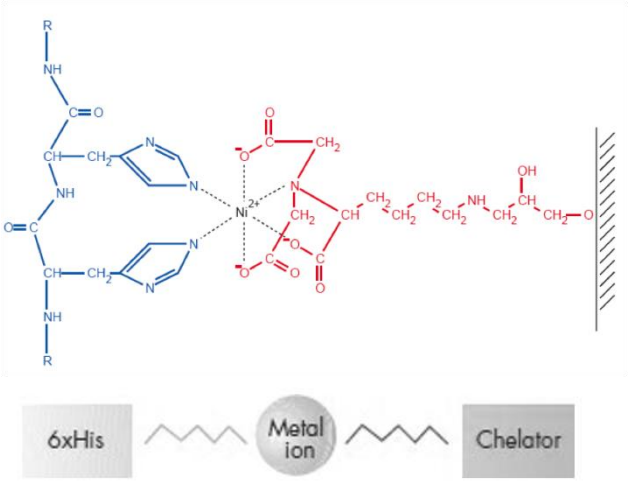


Figure 1.6. The Ni-NTA resin (nickel - nitrilotriacetic acid) is specifically designed for the purification of recombinant proteins with 6x histidine (6xHis) tag expressed in bacteria or mammalian cells. The high affinity and selectivity of the resin (red) bind the proteins that are tagged with six tandem histidine residues (blue). Proteins bound to the resin, can be eluted modifying the pH or with imidazole solution. Ni-NTA resin works to purify 6xHis tagged proteins under native and denaturing conditions. ***(QIAGEN, Ni-NTA Superflow BioRobot Handbook 09/2002, BIOSCIENCES Ni-NTA Resin Spin Columns Catalogue 2015).***

1.7 Fermentation Types and Modes of Operation

Historically, fermentation technology has been one of the most popular techniques used in the biotechnology field. It has been used for production of traditional foods and beverages. Fermentation is defined as a metabolic process caused by microorganism that produces chemical changes in organic substrates through the action of enzymes. The term bioreactor also called fermenter, could be described as a self-contained vessel, capable of growing living microorganisms. It must provide the optimal environment and specific conditions for growth of a biological system (cells, viruses, bacteria, or live tissue). Other factors like mechanical conditions and fluid flow to induce shear, also play an important role for the correct growth of the cells for producing different metabolites. The cost of each fermentation is related with the overall volumetric productivity (defined as the total moles of product which are produced by 1 mol of biocatalyst during its operational life) where the maximization of it, means less expensive fermentations or lower operating costs (Riet et al., 1991; Valdevit et al., 2018; Pino et al., 2018; Gauri et al., 2018).

Every single bioprocess requires the initial study of the biological system involved, aspects like pH, temperature, agitation, substrate characteristics, other conditions such as mass and heat transfer must be specifically analyzed in order define the optimal conditions to improve the biological activity in the bioreactor and finally get the highest possible concentration of the desired products (Valdevit et al., 2018). In the last years, different types of bioreactors have been developed in order to improve the yield in growth, regardless the bioreactor design and geometry. Operating modes can be classified, according to the way the substrate is fed, as discontinuous, semi-continuous and continuous mode.

The batch mode also known as discontinuous mode was the most used method in the early days of biotechnology, it was also used for vaccine production. Nowadays is normally used for routine maintenance culture, research and development purposes for seed-train cultures in the scale-up processes and in the industry for bioethanol production. The main advantage of this system is the simplicity for the set-up while the main disadvantage is the low efficiency for maximizing volumetric productivity (Singh et al., 2014).

To start a batch fermentation, all necessary media components and the inoculum must be added at the beginning of the culture, the mass flows and volumes

are kept constant until the end of the process. However, even as this mode can be considered as a closed process, gases and in some cases, solutions for pH control, antifoam and additions of inducing agents can be performed during the process. The time of the harvest will vary depending on the product and the cell kinetics. However, the entire culture volume is usually harvested at the same time normally at a lowest concentration of substrate and the highest concentration of product (Pino et al., 2018; Gauri et al., 2018).

Fed batch or semi-continuous mode is similar to the batch mode with the difference that in this process fresh substrate or nutrients are continuously (continuous feed) or intermittently (bolus feed) added during the fermentation to extend the growth phase and improve productivity. In this case, acids or bases must be added to the fermentation in order to keep the pH at optimum levels. For aerobic fermentations oxygen must be added continuously, being this a key factor for microbial growth. One of the most critical parameters is the regulation of nutrient transport across the cell membrane, the osmotic pressure, the change of the osmolality could also affect different cellular functions. In some cases, a high osmolality can have a positive impact on the specific productivity of certain cell lines, however, hyperosmotic conditions are harmful for both productivity and growth. Fed-batch has been shown better mixing, yields, higher stability, better efficiency, and higher volumetric productivity compared with batch mode, in combination with a relatively straightforward set-up (Singh et al., 2014). The harvest time (that usually is at the highest peak of cell concentration) is specific for each fermentation and it varies depending upon the host cell, culture medium, feeding schedule and bioreactor control regime. For example, microbial fed-batch using *E. coli* might take from 12 to 48 hours depending on the product, whereas a *P. pastoris* yeast fermentation can take somewhere between 4 and 7 days depending on the protein and the process (Lindskog, 2018). During the last decades, fed-batch mode has become an attractive choice in the biomanufacturing processes for microbial and animal cell fermentations thanks to its robustness and operational simplicity. On the other hand, the main disadvantage of this mode is the non-productive time during the cleaning, verification and preparation for the next process.

In continuous mode process (also called continuous flow reactors) substrates are continuously added to the fermentation while products along with the culture are removed at the same rate, usually keeping a constant concentration of nutrients and cells during the process. In some cases, like wastewater treatment, concentration of substrates and cells can fluctuate. Addition of fresh growth medium could start when

the culture reaches the exponential growth phase, or when the concentration of the substrate decrease dramatically, at the same time equal volume of culture broth must be removed. This mode allows proper mixing and uniform temperature in the system. One advantage of this mode is that the control of pH, optical density and substrate concentration could be performed automatically using a pH-auxostat, a turbidostat and a nutristat correspondingly, making biomass control in this mode much easier compared with batch or fed-batch processes (Lindskog, 2018). Continuous reaction mode, is a superior tool in research and it is used for high volume production; industrial applications as production of vinegar, baker's yeast and wastewater treatment, however the risk of contamination is higher compared with the batch and fed-catch modes (Singh et al., 2014). Figure 1.7 shows graphically the typical behavior and variations of substrate concentration, cell density and volume of fermentations under the three fermentation modes previously described and, Table 1.3 summarizes the advantages, disadvantages and applications of each fermentation mode.

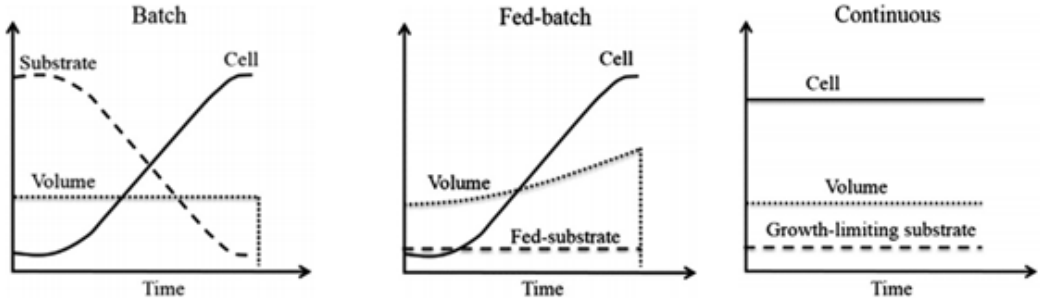


Figure 1.7. Typical variation in working volume and cell substrate concentrations under batch, fed batch and continuous fermentation mode **(Fengxue, 2019).**

Table 1.3. Advantages and disadvantages of different modes of bioreactor operation **(Riet et al., 1991; Gary et al., 2003; Fengxue, 2019).**

Mode	Advantages	Disadvantages	Application
Batch	Simple process. High conversion. Flexibility of operation. Low risk of contamination.	Expensive cost per batch. Large consumption of power. Variability of products from batch to batch. Difficult to use at large scale production. Long growth period. Relatively low productivity.	Microbial fermentation. Cell metabolism. Production of products. To build kinetics models of microorganism. Helpful to study optimal fermentative conditions.
Fed Batch	Good temperature control. High cell densities due to the extension of working time. Control of substrates and products. Reduces the inhibition of substrate. Improve the conversion rate of the products. Control and automation.	Expensive cost per batch. Difficult to use at large scale. Because no medium is removed, the accumulation of the product leads to a decrease of the specific production rate. Higher risk of contamination compared to batch.	Entire fermentation industry. Butyric acid production by *Clostridium butyricum.* L-lactate production from corn cob hydrolysate by *Rhizopus oryzae.* In research for production of cell protein, amino acids, growth hormone, antibiotics, vitamins, enzymes, etc.
Continuous Mode Process	Improve the utilization of equipment. Low operation cost. Easy to clean Parameters in the fermentation tend to be at fixed value and it is easy to control automatically. Good temperature control and continuous production. Easy to clean and simplicity of construction.	High demand for the rationality of the equipment and the accuracy of the feeding. Really high risk of contamination. Susceptibility of strains to degenerate because of mutation. Low conversion per unit volume. High consumption of power due the presence of mechanical pumps.	Industrial production of single cell protein. Production of metabolic products like β-galactosidase by *Monascus spp.* Production of α2 interferon by recombinant *E. coli.*

1.8 Metabolic and Protein Engineering for Biocatalyst Improvement

During the last decades, biocatalysts have been extensively studied and applied to the industry of bulk chemicals and pharmaceuticals (de Carvalho, 2011; Du et al., 2011; Patel, 2011; Rubin-Pitel & Zhao, 2006; Schoemaker et al., 2003; Wen et al., 2009; Zhao, 2011). Biocatalysis has shown several advantages compared with chemical catalysis, some of them are the high efficiency, high degree of selectivity (regio-, chemo-, and enantio-) and friendly environmental reaction conditions (Hudlicky & Reed, 2009; Reetz, 2009). Application of biocatalysts includes the nitrile hydratase catalysed hydrolysis of acrylonitrile to acrylamide for use in plastics, with a production of around 10,000 tons per year, and the production of D-amino acid oxidase catalyzed transformation of cephalosporin C to keto-adipyl-7-aminocephalosporinic acid for production of antibiotics (Wang et al., 2012), among others.

Majority of enzymes that are studied in the laboratory come from microorganisms, due to the advantage of easy cultivation for research, natural abundance, and rich diversity. Application of natural enzymes in the industry is limited because enzymes should have high activity, specificity, and enantio-selectivity towards range of different challenging substrates. Additionally, they must be stable and able to work under several harsh reaction conditions including high temperature, strong acid or alkaline conditions, extreme pH, high substrate/product concentrations and need to tolerate different unnatural solvents. Generation of enzymes with these characteristics is nowadays tailor made by different enzyme engineering techniques (Yang et al., 2016). Advances in modern bioengineering, molecular biology as well as advances in computational tools have increased the number of improved enzymes suitable for different applications in industrial processes. The relatively fast advances and development of biocatalysis make it very interesting, and it is possible that in a near future most of the fine chemical production and pharmaceutical production will be performed with the use of engineered and recombinant enzymes (Madhavan et al., 2017).

In this context, Metabolic Engineering (ME) is a powerful tool that involves modification of metabolic pathways of a host organism for production of useful metabolites, chemicals, fuels, pharmaceutical or medicines. One of the big challenges of ME is the identification of optimal microorganisms and to determine targets to manipulate individual genes or whole pathways. This approach requires the complete

understanding of host cells for genetic modifications, moreover prior knowledge of enzyme network pathway and its kinetics is also needed. The general process of ME could be standardised and summarised in the following steps: bioprospecting and recombinant pathway design; selection and cloning or synthesis of heterologous genes; production host choice, vector choice, and transformation of heterologous genes into a host; troubleshooting expression, folding, and activity of plant proteins in microbial hosts (normally via protein engineering); strain improvement via carbon flux redistribution, toxicity reduction, transporter engineering, removal of regulatory restrictions, enzyme compartmentalization and fermentation optimization. Although the procedure steps are well characterized, several regulatory control mechanisms in nature are not fully understood and the use of other tools like genomic, proteomic and metabolic analyses, enzyme/protein engineering are useful for the success in the pathway modification (Yadav et al., 2016; Joo et al., 2014). A recent example of metabolic engineering was reported by Valle et al (2019), that used *E. coli* for hydrogen biosynthesis instead of its chemical synthesis, being this a promising possibility since the process requires less energy and is more sustainable and ecofriendly without mention that hydrogen is a potential sustainable energy source that is considered as an alternative to fossil fuel combustion, thus helping to reduce greenhouse gas emissions.

Another strategy with exceptional progress over the few decades in biocatalysis is Protein Engineering (PE) that refers to the development of novel biocatalysts with desired features for research or industrial application. The success of protein engineering has been increased due to the fact that several enzymes from living organisms have been modified for better performance like improvement of catalytic power, stability, tolerance under mild reaction conditions (pH, temperature and pressure), specific selectivity for the substrate and promiscuity. These improvements make enzymes very attractive and suitable to be widely used in the chemical and pharmaceutical industry (Sakkos et al., 2019; Yadav et al., 2016). PE can be also seen as a powerful tool that allows artificial evolution (Valetti & Gilardi, 2013).

The main role of protein engineering is to overcome the limitations of natural enzymes as biocatalysts and to engineer process-specific biocatalysts using recombinant DNA technology, genomic and proteomics to modify the sequence of a protein, and hence its structure, to create enzymes with the desired catalytic characteristics (Yadav et al., 2016; Johannes et al. 2006). On the other hand, enzyme mutations of active sites could deliver promiscuous activities or ‘catalytic promiscuity’

that refers to the ability of one enzyme's active site to catalyze or perform secondary chemical reactions. Promiscuity is usually characterized by the involvement of a functional group and reaction mechanism not seen in the native enzyme reaction and has been observed in many proteins (Kazlauskas, 2005). Reactions have different transition states and different mechanism steps; there is not the same reaction for two substrates (Kazlauskas & Bornscheuer, 2012).

During the last decades, the number of examples of catalytic promiscuity is continually growing and the importance of this characteristics lies in the ability to provide opportunities for the evolution of new functions in nature and in the laboratory, that could be applied for therapeutic drugs. Enzyme promiscuity was defined by in 2007 by Hult and Berglund. They classified the unexpected performance of the enzymes in three different types. Firstly, enzyme condition promiscuity that refers to the enzyme that can work under unexpected conditions, such as organic solvents. Secondly, enzyme substrate promiscuity, that refers to enzymes that show the ability to work with unexpected substrates, such as pharmaceutical intermediates. And finally, the enzyme catalytic promiscuity that is an enzyme with the ability to catalyze an unexpected reaction. Even if these characteristics have been studied during the last decades, the understanding of the origins of the phenomenon is still limited, however, it is a relevant area that can be exploited to expand the range of useful applications of enzymes (Copley, 2015; Mohamed & Hollfelder, 2013). In most of the cases, promiscuity reactions are much slower than the natural reactions, however protein engineering can increase the rate of promiscuous reactions or add mechanistic steps to create new catalytic promiscuity.

The use of biocatalysis and synthetic biology is essential to design new enzymes with catalytic promiscuous characteristics that eventually could solve chemical problems (Mohamed & Hollfelder, 2013). Copley et al., (2015) reported that a single mutation in the active site of *Persicaria minor* β-sesquiphellandrene synthase, leads to a more promiscuous enzyme that was able to change the product specificity, to produce additional hydroxylated sesquiterpenes such as sesquicineole, sesquisabinene hydrate and α-bisabolol that are not produced by the wildtype enzyme. These compounds are classified as a terpenoids and have demonstrated to be a potential candidate for drug discoveries in the pharmaceutical industry, one example is the sesquiterpene compound derived from three isoprene units that it is used for malaria treatment.

In some cases, a single point mutation, could turn the enzyme into a much

more stable, efficient/proficient catalyst with completely new enzyme activity. There are several examples of enzyme promiscuity, as the case of Lipase B from *Candida antartica*. Through a single point mutation, Magnusson et al., (2005) were able to change the enantioselectivity 8×10^6-fold and reverse it from R- to S- selectivity, mainly by increasing the activity towards the S-enantiomer. Another example was reported by Seebeck and Hilvert (2003), they induced catalytic promiscuity in pyridoxal-phosphate dependent enzymes (PLP) and converted a racemase into an aldolase by a single point mutation. The mutation of Tyr into Ala in the position 265, showed a 3×10^3-fold reduction in a racemase activity and a 2×10^5-fold improvement in retro-aldol activity using D-phenylserine as a substrate (Hult & Berglund, 2007). Finally, even with good knowledge of an enzyme, the synthesis to produce a specific alternative product using direct evolution and rational design as the main strategies remains a challenge with an unpredictable result (O'Brien et al., 2018; Wang et al., 2017). In the following section the importance, main characteristics, and examples of strategies for protein engineering will be described.

1.9 Strategies of Protein Engineering

Enzymes used in biotransformation processes are frequently prepared as isolated enzymes in solution or immobilized onto resins, and in some cases are used in the presence of organic solvents, harsh chemicals, or under conditions of temperature and pH that are not optimal for enzyme activity. These unnatural conditions result in poor enzyme activity, or complete deactivation due to denaturation or chemical modifications. Development in protein engineering over the last two decades have enabled enzymes to be evolved *in vitro* for enhancing their properties that favour the required industrial process conditions, and to get enzymes variants with altered substrate specificity or enantioselectivity. Although significant advances have been achieved to date in different industries, there still remains a need to improve direct evolution strategies and develop screening or selection tools which make the process of identification and selection of novel enzymes more efficient, and also to achieve better changes to enzyme function. Moreover, some experiments have demonstrated that indeed the majority of mutations that improve the enantioselectivity of enzymes are around 10 Å of the active site, other experiments have shown that not all enzyme properties can be expected to be improved through active site mutations alone, and indeed thermostability was shown to be improved equally by mutations close to and distant from the active site (Hibbert et al., 2005).

Several successful advances have been made to bioengineer enzymes through various technologies like rational design of proteins and directed mutagenesis. But still there is a need for developing most robust and powerful technologies to enhance the biocatalysts to cater for the needs of various industries. The following section will describe in detail and with examples the use of rational design and directed evolution, two strategies that have been demonstrated to be successful approaches to optimize industrial processes through the introduction of directed genetic changes using recombinant DNA technology.

1.9.1 Rational Design

Rational design or site directed mutagenesis is a method that allows to generate novel proteins that serve as efficient biocatalysts to meet industrial applications. Having deep knowledge of the enzyme structure and active site, it is possible to mutate an amino acid at a specific site of a protein, that will modify the structure and in some cases the mechanisms of action of the enzyme. However, this change must be evaluated in order to verify if the modified protein performs better, equal or worse than the original protein. Bioinformatics prediction of structure is fundamental for rational design mutagenesis to enhance specificity, stability, activity, solubility and expression of the biocatalyst using single and combinational mutation (Chica et al., 2005; Winkler & Kao, 2014; A Madhavan et al., 2017). A useful strategy is the selection of amino acids present in positions of interest where the population of proteins have relation with the enzyme activity. Residues that are present in these positions are determined as “allowed” with higher probability in an active protein. On the other hand, residues that are not found in this position are determined as “not allowed” and suggest inactive proteins (Siloto et al., 2012). Several studies have been reported improvement of enzyme properties using site-directed mutagenesis. For example, the Cys22 of Phi-class glutathionine S-transferase from Oryzasativa was replaced with Ala, this modification improved 2.2-fold the k_m value of the new variant compared with the wild type (Jo et al., 2012). Whereas the mutation of the basic histidine residues His275, His293, and His310 of amylase from *B. subtilis* by replacement with aspartic acid via site directed mutagenesis, the k_{cat}/K_m value of the mutant increased by 16.7-fold compared with that of the WT (Yang et al., 2013).

1.9.2 Directed Evolution

Directed evolution (DE) mimics biological evolution in the laboratory. It is considered as an effective strategy to generate enzyme variants with enhanced functional properties such as catalytic turnover. This occurs in a very quickly (the course of a PhD) compared with the natural evolution (millions of years) (Tracewell & Arnold, 2009; Dalby, 2011; Wang et al., 2012). In directed evolution, random genome mutations are generated and mutants with the desired properties selected based on a specific and desired function. Directed evolution has three main steps 1) Construction of a mutant library 2) Screening/selection of mutants based on the improved function, and 3) Isolation of the improved genes. This strategy introduces genetic diversity and it could be repeated for several rounds until the desired changes are observed. Once the gene for the enzyme to be evolved has been isolated and usually placed in a plasmid of choice, a library of mutants can be produced via a number of differing strategies: chemical mutagenesis, random mutagenesis, saturation mutagenesis, or DNA shuffling. One of the most common approaches is to create random mutations using error prone PCR, as the generation of genetic diversity by this technique is relatively easy to perform than other techniques, while also being free of any intellectual property (Madhavan et al., 2017). Once the library is constructed, the most common and best strategy is to transform the library into bacterial cells, from which many thousands or millions of individual colonies can be isolated and grown, where each expresses a different enzyme variant. Once this library is produced, it is necessary to carry out a high throughput screening method to identify those variants with interesting characteristics, such as higher activity or higher thermostability.

One of the most important limitations of this strategy is the development of a high-throughput screening method (HTSM) to identify proteins with the desired properties from a large number or variants (Barahona et al., 2016). In a few decades, researchers have developed several methods for screening mutagenic libraries that includes agar plate and micro titre plates-based screening strategies. The use of agar plates can involve the direct correlation between growth of the host microorganism on selective agar plates for screening of enzyme function that improves the fitness of the cell. This can apply to critical metabolic enzymes that enable growth on a new food source, faster growth, or otherwise secondary metabolism enzymes that confer antimicrobial resistance. Alternatively, the extracellular secretion of the improved variant into the solid media produces a larger halo formation through eg. the breakdown of an opaque or coloured substrate, compared to the wild-type enzyme.

Such screens can be linked to the improvement of the expression level or activity of enzymes such as polyesterases, esterases, lipases, phospholipases, peptidases, acyl transferases, among others (Shim et al., 2004; Molitor et al., 2019). On the other hand, micro titre plates are also widely used as a screening format because they can be used to screen a large number of variants, particularly in automated liquid handling robotics systems, coupled to plate readers. Its principal advantage is that they have a broader dynamic range, which aids the identification of comparatively low improvements in the desired enzyme function compared to the previous method with agar plates (Kelly et al., 2008).

In this context we can say that the screening and selection of the mutants with the desired characteristics is the main challenge. For example, complete randomization of a mere decapeptide would yield 10^{13} unique combinations of amino acids that must be subjected to screening and selection (Packer & Liu, 2015). Moreover, it is worth to mention that not all enzyme activities are suitable to develop a High Throughput Screening Method (HTSM) and not all screening methodologies are easy or suitable for implementation (Chica et al., 2005; Lutz, 2010).

Directed evolution has become the most powerful and commonly used tool in protein engineering to generate novel biocatalysts, reagents and therapeutics with the advantage that it is not necessary to have an in-depth understanding of the structure/function relationship (Chica et al., 2005; Packer & Liu, 2015). Several examples of successful improvements have been reported since more than two decades. That is the case for improvement in thermo stability of α- amylase from *Bacillus licheniformis* using random mutagenesis. Using random mutagenesis and screening it was possible to increase the half-time of the enzyme at 90°C 10 times higher than the wild type (Arnold, 1993). More recent examples include mutagenesis of alcohol dehydrogenase in *Pyrococcus furiosus* that improved the catalytic activity 10-fold higher than the wild type enzyme at low temperatures for production of enantiopure hexanediol (Machielsen et al., 2008). Improvement of activity of glucose dehydrogenase by a factor of 13 was obtained using DNA shuffling method for production of a key intermediate in the production of atorvastatin that helps to decrease the low-density lipoprotein (bad cholesterol) and raise the high-density lipoproteins (good cholesterol) in the blood (Ma et al., 2010).

Another strategy that combines both approaches is called semi-rational, smart,

or knowledge-based library design. This strategy uses information of protein sequence, structure, function, and computational predictive algorithms to preselect and identify hot spot residues. This previous knowledge limited amino acid diversity for protein engineering (Korendovych, 2018). One advantage of this approach is that is focuses on specific amino acid positions, translates into smaller library sizes and at the same time it can increase the efficiency of biocatalyst tailoring. The smaller number of mutants gives the opportunity to evaluate the library when no high-throughput assay system is available (Lutz, 2010). Xylose reductase from *Neurospora crassa* is a good example of a promiscuous enzyme, which prefers D-xylose over L-arabinose. This enzyme was engineered via semi-rational design using site-saturation mutagenesis, combinatorial active site-saturation testing and finally a specific mutation (L109Q). This variant showed a 9-fold higher preference for D-xylose than L-arabinose (Nair & Zhao, 2008). The combination of error prone-PCR (ep PCR) and site-directed mutagenesis was able to enhance hydrolytic activity and lipase substrate preferences of the bromoperoxidase A2 from *Streptomyces aureofaciens* (Chen et al., 2009). Modification of substrate specificity of *Cellulomonas uda* cellobiose phosphorylase from cellobiose to lactose was achieved using directed evolution after a single round of random mutagenesis. Selection of variants with improved lactose phosphorylase activity was achieved *in vivo* using a minimal medium containing lactose as a sole carbon source. Afterwards site-directed mutagenesis of T508I/N667A improved 7.5 times the specific activity on lactose compared with the wild-type enzyme (de Groeve et al., 2009). We can say that semi-rational design is an effective strategy to move towards generalist enzymes, however some cases need optimization through several rounds of mutagenesis and screening (Chen, 2012). Table 1.4 summarise the different approaches described above.

Table 1.4 Comparison of approaches for engineering enzyme activity **(Chica et al., 2005).**

Method	Rational Design	Random Mutagenesis	Semi-rational Design
HTP Screening	Not essential	Essential	Advantageous but not essential
Structural and/or Functional Information	Both essential	Neither essential	Either is sufficient
Sequence Exploration	Low	Moderate, random	Experimental: Moderate, targeted Computational: vast targeted
Probability of Obtaining Synergistic Mutations	Moderate	Low	High

The advanced knowledge and continuous development of new mutagenic libraries and high throughput screening methods will make directed evolution even more powerful, and in consequence one of the most interesting strategies for novel enzyme production and application in pharmaceutical, agriculture, food, chemical and energy industries (Sakkos et al., 2019). However, although DE techniques have been extensively used for many years, it is required to be adapted to singular enzymes because each enzyme has different and specific properties that should be improved. Some of the most common directed evolution techniques are described in the following sections.

1.9.2.1 Error Prone PCR

This technique is one of the most common strategies (Chen & Arnold, 1993), which introduces random point mutations into a population of DNA, producing a high number of simultaneous mutations per gene. This is based on the low fidelity amplification of genes due that reaction conditions have some modifications, such as addition of Mn^{2+}, higher concentration of $MgCl_2$, in some cases unbalanced of dNTP's concentrations or low fidelity *taq* polymerase. Another interesting aspect is that mutations during PCR are accumulative with each cycle of amplification, making the number of cycles another parameter to increase the average number of mutations (Packer & Liu, 2015). Under the conditions mentioned above the most common mutations in the libraries are changes from A → G and T → C resulting in sequences with higher content of CG. One of the main advantages of this technique is that is relatively easy to use with commercial kits, allowing high mutation rates and providing a relatively even mutational spectrum, while some disadvantages are that it can generate mutations at nucleotide level but that does not translate to a different amino acid or protein. Even when the use of this technique is fast and efficient, some authors (Stemmer, 1993) have reported that it is useful for single cycles but has some limitations for its application on multiple cycles.

1.9.2.2 DNA Shuffling

This is an in vitro recombination method for homologous sequences that allows directed protein evolution. The first step involves the fragmentation of the gene by digestion with DNase, giving random fragments. The second part consist in the reassembly of the fragments into a full-length gene using PCR. Fragmentation of the gene is based on sequence homology, and recombination happens when different fragments anneal to each other causing a templates switch or crossover event (Stemmer, 1994). This method generates point mutations with a similar rate than error-prone PCR (Sheryl et al., 2007). Moreover, this technique can be applied to sequences >1 kb, giving a similar mutagenesis rate than error-prone PCR. One more advantage of DNA shuffling is that it is the only technique that allows removing neutral mutations by backcrossing with excess parental or wild-type DNA. Even when the DNA shuffling is a powerful technique, the main limitation of the method is that it requires the presence of zones with relatively high sequence homology (Sheryl et al., 2007; Stemmer, 1994).

1.9.2.3 Combined Methods

Combination of directed evolution and rational design can potentially guide to the generation of proteins with enhanced activity, thermostability, substrate acceptance and enantioselectivity; properties that could be able to cover some needs in the pharmaceutical and chemical industry (Illanes et al., 2012). Iterative saturation mutagenesis (ISM) is a relatively new method that combines directed evolution and rational design for the generation of enzymes with enhanced catalytic properties like the previously mentioned. The first step of this approach is the analysis of the protein and identification of regions that may be crucial for enhancing a given catalytic property. Subsequent saturation mutagenesis is only considered to these 'hot sites'. This analysis has been demonstrated to increase the probability of success while reducing cost, time and human effort (Reetz et al., 2007). The choice of these regions varies according to the nature of the catalytic property that is to be improved. In the case of substrate acceptance and/or enantioselectivity the focus is on sites around the complete binding pocket in a systematic manner (Reetz et al., 2006)

Once the sites of the enzyme were identified (for example four sites), each of these sites (composed for 1-3 amino acids or more), must be randomized by saturation mutagenesis to generate four different libraries of mutant enzymes that are subsequently screened, and the best variants of each library is identified and sequenced. Until this point, it seems very similar to the methodology for mutagenic libraries using other methods like error-prone PCR except for the systematic view that was taken. The most crucial parts for ISM are the following steps. The selected variant with the best properties from a given library is used as the template for another round of saturation mutagenesis experiments at the respective other sites. This process can continue until the desired degree of improvement has been achieved (Reetz et al., 2007).

A successful example of this approach was the generation of two variants of lipase (Lip A) from *Bacillus subtilis*, with dramatically improvement of thermostability displaying T50 values of 89 and 93 °C compared to the T50 of the WT of only 48 °C, (see Figure 1.8). Because hyperthermophilic enzymes are more rigid than mesophilic analogs (Jaenicke et al., 1998; Buchner et al., 2005), mutations of regions with high degrees of flexibility were generated. The use of atomic displacement parameters obtained from X-ray data, were key to identify such sites for thermostability improvement.

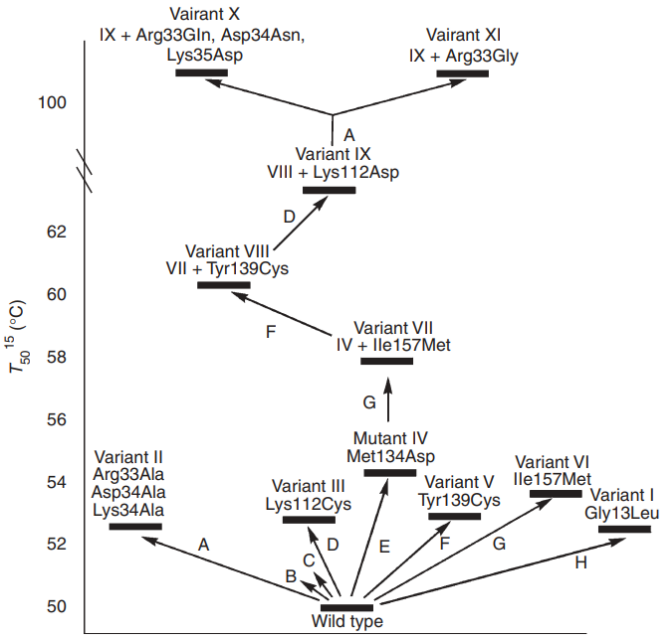


Figure 1.8. Enhancing the thermostability of the lipase from *Bacillus subtilis* (Lip A) by iterative saturation mutagenesis. The sites A, B, C, D, E, F, G and H chosen for saturation mutagenesis are shown graphically (Reetz et al., 2007).

This method also known as B-Factor Iterative Test (B-FIT) can identify the amino acids with the highest flexibility and thereby creates targets for mutagenesis, (see Figure 1.9). Using this approach, each new cycle maximizes the probability of obtaining additive and/or cooperative effects of newly introduced mutations in a defined region of the protein (Reetz et al., 2007).

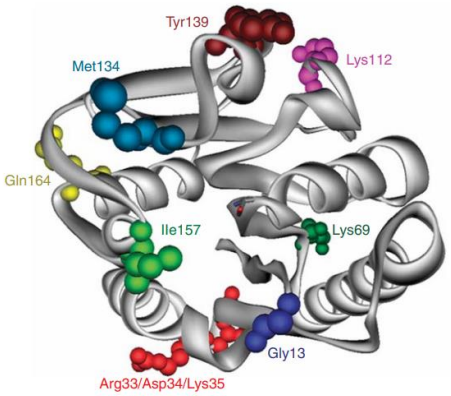


Figure 1.9. Structural model of *Bacillus subtilis* lipase A (Lip A) displaying the sites A-H chosen for saturation mutagenesis based on high B-factors (Reetz et al., 2007).

1.10 High-Throughput Screening Methods

New screening methods are required to enable the identification of improved enzymes from larger libraries, and to obtain the desired properties (Hibbert et al., 2005). Analytical methods including high pressure liquid chromatography (HPLC), gas chromatography (GC), mass spectrometry (MS) and well plate-based spectrometry assays have been traditionally used for analysis and quantification of metabolites. However, these approaches generally are not efficient to screen very large mutagenic libraries (Dietrich et al., 2010). In this sense, massive parallelization, automation and miniaturization of well-plate assays allow to maximally harness the power of large DNA and evolution libraries (often 10^7-10^9 variants/week) to screen pharmaceutical and industrial active compounds (Dragosits & Mattanovich, 2013; Lin et al., 2017).

High-throughput screening (HTS) methods for catalysis have a key role in the identification of new enzymes and biocatalysts. It is a process that allows automated testing of many chemical or biological compounds for a specific biological target. These methods are necessary for screening libraries generated by methods such as error-prone PCR or gene shuffling of an existing enzyme gene or gene family (Wahler & Reymond, 2001). The main goal of the HTS methods is the identification from mutant libraries of a candidate that affects the target in the desired way using robots (for better reproducibility and low data scattering), plate readers as detectors and dedicated software for instrumentation control and data processing. The selected candidates also called ‘hits’ or ‘leads’, give a good starting point for further steps, such as more rounds of random mutagenesis, specific modification using site-directed mutagenesis for better optimization of the molecule. In consequence, HTS methods could be seen as a fast scan of biological processes for compounds or enzymes that show desirable characteristics (Wiggelsworth, 2020).

The success of directed evolution depends mainly on two aspects: genetic diversity and high-throughput screening or selection methods. HTS methods are key to identify desired mutants in directed evolution (Xiao et al., 2015). Screening and selection are common words that must be differentiated, screening refers to evaluation of every protein to identify the desired property, whereas selection automatically eliminates non-functional variants (Leemhuis et al., 2009). One of the most challenging parts for enzyme engineering is to couple phenotype analysis with a compatible HTS method.

Screening methods give the possibility to evaluate individual protein variants, and in consequence it reduces the possibility of missing a desired mutant. Automation of the HTS methods can streamline traditional screening processes. More recently, researchers have developed methods to detect the desired mutant more easily, such as fluorescence-activated cell sorting (FACS). Microtiter plates are another way to miniaturize test tubes to multiple wells. The most common format is the 96-well plate, however there are other commercial formats like the 1536-well and 9600-well for high density of samples. Generally, the enzyme activity assay is performed with the addition of reaction components either to the crude cell extracts or the pure enzymes, being the colorimetric reactions or fluorometric assays some of the most convenient ones among a several number of microtiter plate-based enzyme activity assays (Behrendorff et al 2013; Mack et al., 2008; He et al., 2011). In some cases, the decrease in concentration of substrate or generation of product allows the identification of active mutants by macroscopic observation, measuring of absorbance or fluorescence with the use of a plate reader. Microtiter plates are also useful as a well-developed alternative to shake flasks that allow screening of mutants, with different profiles of cell growth, substrate uptake and product formation (Xiao et al., 2015; Duetz et al., 2007; Weis et al., 2004).

HTS can be divided into four main steps. The first step is the preparation of samples and compounds using directed evolution to produce mutagenic libraries. Samples must be prepared in an arrayed format. The use of microplates with 96, 384 or 1536 wells are some of the most common plates used to achieve this. These enable replicates of the plates to be made, with different copies used for cell regrowth, different enzyme assay conditions, and for storage of cells at -80C. In practice, all variants tested must have a stock plate or plate copied in case some of the analysed samples show enhanced properties. The second step is to develop a suitable, preferably automated and stringent assay for the desired activity. The stringency of the assay is required to enable a clear distinction between negative and positive controls, while also minimising false negative and positive results. Depending on the nature of the sample and detection method for analysis, the colour, fluorescence, or absorbance of the solution in the microplate could change. The third step is the configuration of a robotic workstation to manipulate multiple plates simultaneously to speed up the data acquisition and throughput. Robotic platforms use mechanical arms that can manage the reagent addition, mixing, incubation and detection of samples in microplates. The last step of HTS is the product detection, this is usually performed by optical measurements (fluorescent or luminescent detection, colourimetry, light scatter, absorbance) that have a correlation with the concentration of a specific product

(Wiggelsworth, 2020).

The most common HTS selection method is either in solid phase (e.g. micro-organisms growing on a L-Agar plate) or microtiter well plates (e.g. micro-organisms growing individually in 96-well plates). Solid-phase screening relies on product solubilization following an enzymatic reaction that produces a halo, a fluorescent product or a strongly absorbing (chromogenic) product. It is possible to detect the enzyme product directly or maybe coupled to a second enzyme whose product can be easily detected and monitored as it was reported with the coupling of cytochrome P450 with horseradish peroxidase (Joo et al., 1999). When the production of the enzyme is intracellular, it is necessary to transfer a portion of the cells from a colony onto a filter membrane to perform the assay. However, the use of digital imaging has been impacting the evaluation of mutagenic libraries to achieve higher throughput and quantitative signal analysis. It is possible to use digital imaging to estimate the maximum rate of reaction (rate of reaction when the enzyme is saturated with substrate) Vmax of putative enzymes expressed by individual colonies. Screening of up to 105 colonies per day has been reported using advanced digital imaging systems (Bylina et al., 2000). When solid-phase format screening is not possible, individual clones must be grown and assayed in microtiter wells. Nevertheless, these assays are more time consuming than solid-phase assays. Although the use of robotic automation systems is an incredible advantage, one of the main drawbacks is the significant cost of that kind of equipment (Olsen et al., 2000).

More recently, there is a need of more sensitive high-throughput systems, like screening for specific product concentration that can leverage single cell, high-throughput analysis techniques such as fluorescence-activated cells sorting (FACS) (Mattanovich & Borth, 2006) or droplet-based microfluidic sorting (Agresti et al., 2010). The need for new technologies has led to a proliferation of different biosensor modalities and approaches that are each tailored to specific classes of molecules (Lin et al., 2017). Transcription factors (TFs), engineered fluorescent proteins and G protein-coupled receptors are the most common types of protein-based biosensors. TFs are proteins that can naturally react and then transduce an easily measurable transcriptional response through a reporter gene. Different studies have demonstrated that TFs can be used as a genetically encoded, in vivo and in vitro biosensor, being TFs one of the most popular and valuable biosensor types developed for high-throughput screening (Mahr & Frunzke, 2016; De los Santos et al., 2016; Li et al., 2017).

TFs detection includes acids, alcohols, sugars, amino acids, aromatics antibiotics, flavonoids, fatty acids, and others. These new high-throughput methodologies to detect small molecules and enzymes functions are the next generation techniques that allow automated detection, measurement of rapid analyte concentration and easy measurable absorbance of fluorescence signal (Lin et al., 2017).

Fluorescence-activated cells sorting (FACS) is a strategy based upon the fluorescent signal of individual cells or emulsion droplets that exhibits different fluorescent signals and allows to differentiate cells into two or more containers at rates up to 30,000 cells/s (Becker et al., 2004). For this methodology it is necessary to take advantage of the differences in physical/chemical properties between the substrate and the product. Generally, it uses a fluorescent substrate that can be transported both into and out of the cell and react with the enzyme of interest. Variants could be identified due that the fluorescent product cannot get out of the cell due to its physical or chemical properties (Xiao et al., 2015; Yang et al., 2009). Using this strategy, it was possible to identify a variant of glycosyltransferase with 400-fold enhanced activity for the fluorescent selection substrates (Aharoni et al., 2006).

Digital imaging (DI) integrates single pixel imaging spectroscopy for screening of variants in solid phase. A classic example of the application of this method was the screening of transglycosidases variants where a covalent glycosyl-enzyme intermediate was involved in the screening. This method was able to identify variants with significant higher activity (70-fold improvement) on transglycosidase/hydrolysis in the presence of the acceptor cellobiose and X-glycosyl as donor substrate. Mutants with non-hydrolytic activity gave light colored colonies, while mutants with high transferase activity resulted in a colour increase of the colonies (Kone et al., 2009).

The use of cell surface display is another strategy where the fluorescent product is enzymatically linked onto the cell surface and it directly reacts with substrates (Lee et al., 2003). This technology has been applied in bacteria, yeast and mammalian cells (Kim et al., 2000; Chen et al., 2008; Boder 1997; Yi et al., 2013) and it has been reported as a general strategy for evolution of bond-forming enzymes. Once the enzymes bound to the substrate, it is possible to subject them to fluorescence excitation and FACS screening (Xiao et al., 2015). This method has shown an impressive 6,000-fold enhanced activity after a single round of screening for the sortease (bacterial transpeptidase) (Chen et al., 2011). *In vitro* compartmentalization

(IVTC) uses artificial compartments such as water-in-oil (W/O) emulsion droplets or water-in-oil-in-water (W/O/W) double emulsion droplets to compartmentalize a DNA library within it, expressing a protein. Within each droplet, a mechanism to link the function of an enzyme to the gene is employed such that when the emulsion is broken, functional genes can be recovered and consequently amplified by PCR to finally be subjected to additional rounds of selection and mutagenesis until the desired characteristic is obtained.

In this context IVTC holds promise for screening enzymatic activity by combining with FACS of uniform polymer particles with diameters of 0.5-500 μm (microbeads) (Xiao et al., 2015). This approach can screen libraries of up to 10^{12} protein variants by allowing beads to be overloaded with up to 10^4 unique mutants (Gianella et al., 2016). IVTC has been proven to be a powerful tool for directed evolution allowing the use of extremely large libraries and has been successfully employed for the improvement of specific activities of several enzymes including phosphotriesterases (Griffths & Tawfik 2003), galactosidases (Mastrobattista et al., 2005) and restriction enzymes (Doi et al., 2004). A successful example is the identification of *Taq* polymerase variants capable of showing activity in the presence of heparin sulfate, a known inhibitor of this enzyme (Ghadessy et al., 2001). Another example is generation of a variant of β-galactosidase that exhibits 300-fold higher $kcat/K_M$ values compared with the wild type enzyme (Giepmans, 2006). Nevertheless, one of the main disadvantages of this strategy is the incompatibility with many enzymes because of the different conditions between transcription-translation and screening. Table 1.5 summarizes the advantages and disadvantages of the HTP screening methods previously discussed.

Table 1.5. Summary of HTP screening methods **(Xiao et al., 2015).**

HTP Screening Methods	Screening/Library size	Advantages	Disadvantages	Application
Microtiter Plates	$<10^4$/day	Applicable to many assays	Laborious	Enzyme reactions leading to change in color, fluorescence, pH, cell growth, etc.
Digital Imaging	Limited by transformation efficiency	High sensitivity	Not generally applicable	Restricted to colorimetric activity assays
Fluorescence-Activated Cells Sorting	up to $3x10^4$/s	High sensitivity and extremely high throughput	Target enzymes activity has to be couples with the expression level of fluorescent proteins	Enzyme reaction leading to change in fluorescence
Cell Surface Display	Limited by transformation efficiency	Avoids possible cell lysis during enzymatic reaction	Protein expressed as fusion proteins	Screening for bond-forming enzymes
***In Vitro* compartmentalization**	Limited by the throughput of the detection method (e.g. FACS) but not transformation efficiency	High sensitivity and efficiency	Lack of posttranslational modifications; not suitable for screening enzymes with incompatible conditions between transcription-translation and screening	Screening enzymatic activities combined with FACS

1.11 Transketolase, Structure and Reactions *In vivo*

Transketolase (TK) is an intracellular thiamine diphosphate (ThDP) dependent enzyme that is widespread in nature in all organisms. It has been isolated from a wide range of sources including: pig liver (Philippov et al., 1980), rat liver (Paoletti et al., 1986), rabbit liver (Masri et al., 1988), *Candida utilis* (Kieley et al., 1969), mouse brain (Blass et al., 1982), human leukocytes (Mocali et al., 1989) erythrocytes (Abedinia et al., 1992) and for synthetic purposes it has been isolated from yeast and spinach (Demuynck et al., 1991). It has also been isolated from *E. coli,* and it has been over-expressed to produce substantial quantities of the protein with a potential use in industrial processes (Hobbs et al., 1993).

TK was discovered by Racker and Horecker in 1953, however it did not receive much attention until 1992, when the X-ray crystal structure of the enzyme from *Saccharomyces cerevisiae* was reported (Lindqvist et al 1992). After that, other researchers reported the TK X-ray crystal structures from other organisms, such as maize, *Escherichia coli, Leishmania Mexicana, Bacillus anthracis, Francesilla tularensis* and human tissues (Mitschke et al 2010; Asztalos et al., 2007; Veitch et al., 2004). TK from *Saccharomyces cerevisiae* is referred to as '*Sce* TK'. TKs from different organisms exhibit a wide range of substrate specificities, giving this enzyme considerable scope and potential for the enzymatically catalysed synthesis of a wide range of chemicals and pharmaceutical intermediates (Costelloe 2006)

In vivo, TK is crucial in metabolic regulation, providing a link between glycolysis and the non-oxidative branch of the pentose phosphate pathway. The enantioselective carbon-carbon forming action of TK makes it significantly interesting for biocatalysis (Costelloe 2006). It catalyses two processes, the cleavage of the C-C bond in a ketose (the donor substrate) and the reversible transfer of the double-carbon fragment thus formed (active glycolaldehyde) onto an aldose (the acceptor substrate). *In vivo* transketolase reaction transfers a two carbon ketol group from D-xylulose-5-phosphate to D-ribose-5-phosphate thereby, generating D-sedoheptulose-7-phosphate and glycolaldehyde-3-phosphate in the presence of cofactors Mg^{2+} and ThDP (Figure 1.10) in the cytoplasm.

TK is generally considered as a promiscuous enzyme and it can use substrates such as xylulose-5-phosphate (X5P), fructose 6-phosphate (F6P), erythrulose-4-phosphate, and sedoheptulose-7-phosphate as typical donor substrates whereas

ribose 5-phosphate (R5P), glyceraldehyde-3-phosphate (G3P), and erythrulose-4-phosphate are typical acceptor substrates. Donor substrates must include an available oxo group adjacent to the C-C bond cleaved, the presence of a hydroxyl at the first carbon atom (C1), and the D-threo configuration of hydroxyls at the asymmetrical third and fourth carbon atoms (C3 and C4) (see Figure 1.10).

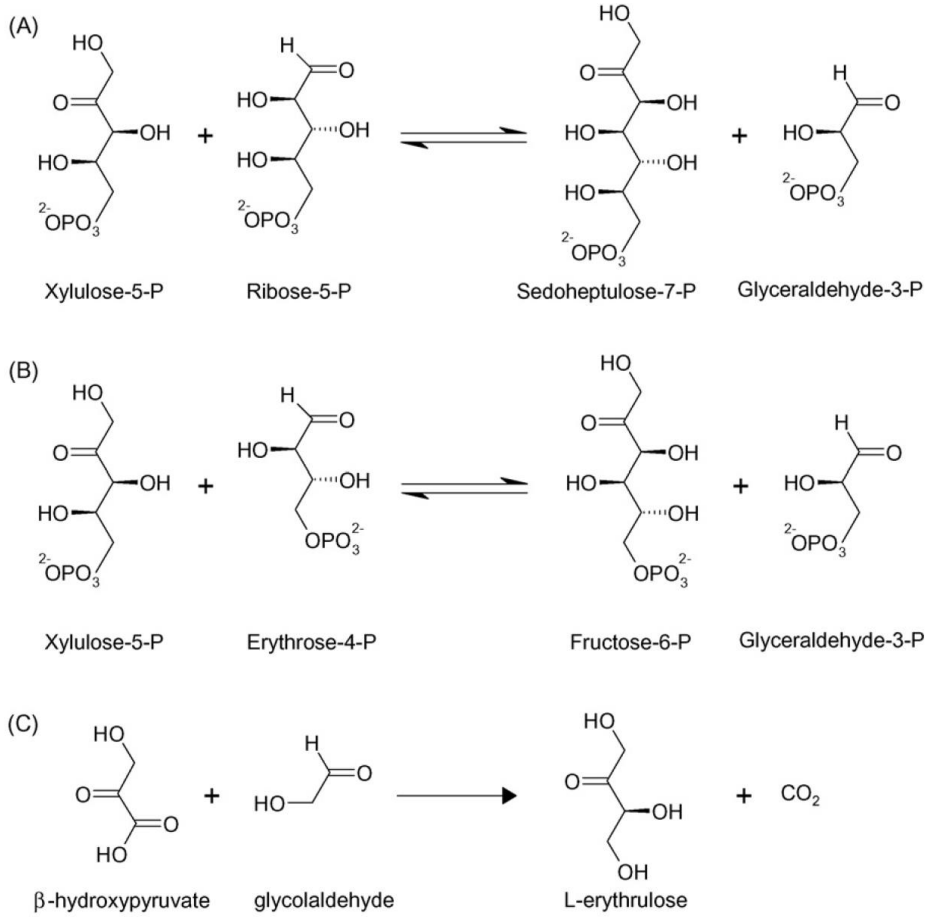


Figure 1.10. TK reactions. (A and B) Natural reactions of TK in the reductive pentose phosphate pathway. (C) Non-natural biocatalytic TK reaction in which Li-HPA is used to make the reaction irreversible **(Hibbert et al., 2007).**

Lithium hydroxypyruvate (Li-HPA) and dihydroxyacetone, which both lack of asymmetrical carbon atoms, are the exception to the rule. Li-HPA is the ideal donor substrate for TK as it yields CO_2, making the reaction irreversible. TK and transaldolase enable the interaction of the pentose phosphate pathway and glycolysis for the formation of coenzymes, vitamins, aromatic aminoacids (phenylalanine, tyrosine, and tryptophan) and generate precursors for nucleotide synthesis. Moreover, TK is also a key enzyme of the photosynthetic Calvin cycle, to catalyze the reactions of F6P and

sedoheptulose-7-phosphate with G3P leading to the formation of X5P and other products that are involved in the photosynthetic process. TK is considered the simplest representative of the ThDP-dependent enzymes, and it has served as a good model for these kind of enzymes (Kochetov et al., 2014).

The general structure of TK (from *S. cerevisiae*) appears as a homodimer of two identical subunits of 74.2 kDa. A homodimer is defined as two identical proteins whereas a dimer is considered a macromolecular complex formed by two protein monomers or single proteins which are usual non-covalently bound. As it can be seen in Figure 1.11, each subunit comprises three α/β-type domains: N-terminal or PP domain (residues 3-322), the Pyr domain or middle (residues 323-538), and the C-terminal (residues 539-680). (Kochetov et al., 2014). The PP domain consists of a five strand β-sheet (parallel) flanked with helices. The Pyr domain contains a six strand β-sheet (parallel) with α-helices. The TKC domain is composed of a mixed β-sheet with four parallel and one antiparallel strand. The first two domains are involved in binding the cofactors (ThDP), whereas the C-terminal contains no catalytic residues and its function is still unclear, however, some researchers have shown that TK is active in the absence of this domain (Costelloe et al., 2008).

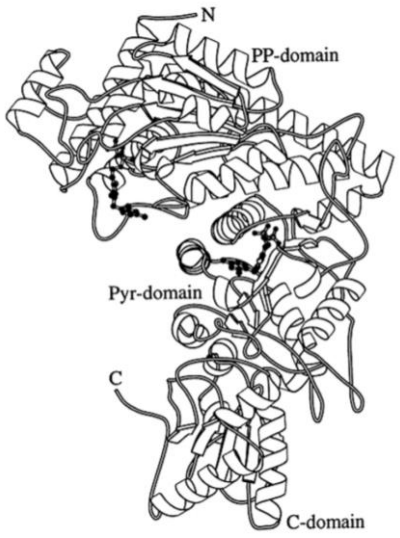


Figure 1.11. Schematic view of the subunit of transketolase. The two thiamine diphosphate molecules (which interact with PP and Pyr domain of the subunit, respectively) are shown as ball and stick models **(Schneider et al., 1998).**

The dimeric structure of TK is mainly formed for the PP domain and Pyr domains (see Figure 1.12). ThDP is embedded into the deep cleft between both

subunits and the C2 of the thiazole ring and 4'-amino group of the aminopyrimidine ring remains fully accessible to the solvent. Figure 1.13 shows a schematic representation of the ThDP interaction with the functional groups of the active site of TK, ThDP binds at the interface of the subunits, interacting with residues from PP domain of one subunit and the Pyr domain of the second subunit. The interface between TK subunits has several cavities where a large solvent-filled channel is formed running between the two ThDP molecules. The channel contains several glutamate residues and a hydrogen bonding network, where the aminopyrimidine ring can form a series of hydrogen bonds, being the one formed between the N1'-atom of the ring and the side chain of the Glu-418, a key bond in the mechanism of ThDP-dependent catalysis (Costelloe et al., 2008; Nikkola et al., 1994).

Some amino acids have been reported to be crucial to keep the TK activity, for example the conserved residues Glu411, Glu160 and Glu165. Mutation of these amino acids disrupt the hydrogen-bonding network and in consequence destabilizes the dimer (Meshalkina et al., 1997). The function of the hydrogen-bonding network has been considered to have a role in the negative cooperativity of ThDP binding to the two active sites, however this postulation has not been confirmed (Esakova et al., 2004).

The pyrophosphate group of the coenzyme interacts with the PP domain of the apoprotein both directly (by forming hydrogen bonds with His-69, His-263, and Lys-158) and indirectly (via the bound calcium ion) (see Figure 1.13). Since the active site of the TK is formed by two subunits, the dimeric form of the enzyme could be seen as a functional catalytic unit (Kochetov et al., 2014).

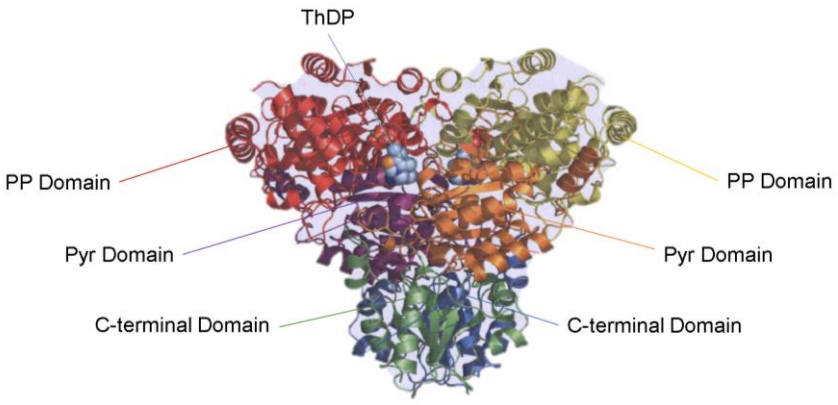


Figure 1.12. Schematic view of the dimer of transketolase. The PP domain is coloured in red and yellow, Pyr domain in purple and orange and C-term domain in green and blue. ThDP is symbolized in blue spheres. **(Costelloe et al., 2008).**

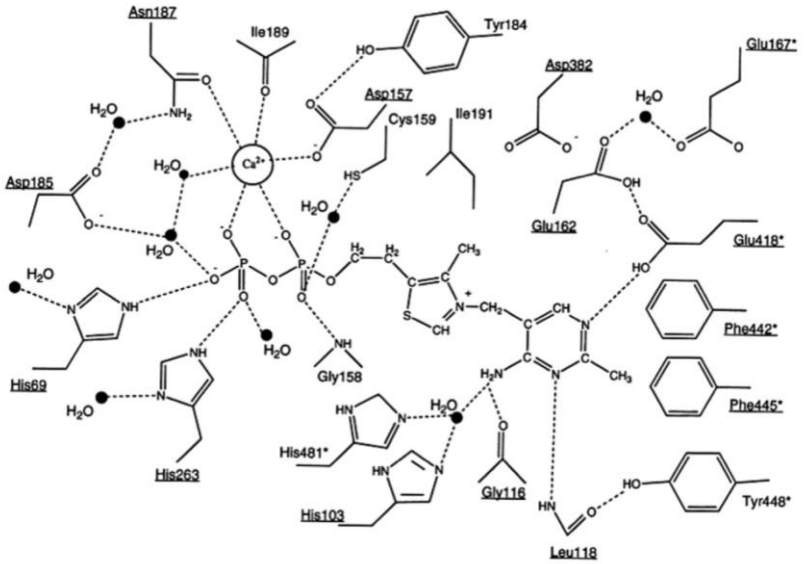


Figure 1.13. Cofactor-protein interactions in the thiamine diphosphate (ThDP) binding sites of TK. Conserved residues are underlined and residues from the second subunit are marked by * **(Schneider et al., 1998).**

TK has been also used for production of L-erythrulose and in process screening for the synthesis of D-xylose 5-phosphate (Smith et al., 2008). However, the tolerance of TK for a range of non-α-hydroxylated aldehydes is interesting to explore and expand the use of the enzyme for the preparation of chiral synthons and their application in the pharmaceutical industry. Previous results (Sprenger & Pohl, 1999) have shown that *E. coli* TK is deactivated at >40 °C, for that reason it is necessary to develop new variants with higher stability than the WT TK, as well as high specific activity and tolerance to new aldehydes as substrates for industrial application (Smith et al., 2008).

1.12 Application of TK

TK is a very interesting enzyme due to its ability to utilise a wide variety of substrates such as non-phosphorylated 2-hydroxyaldehydes, aromatic aldehydes, such as benzaldehyde and hydroxy benzaldehyde, among others. The conversion rate of *EcoTK* using Li-HPA as a ketol donor is around 60 U/mg (1U = the amount of enzyme activity that will catalyse the transformation of one micromole of the substrate per minute under standard conditions), compared with the spinach and yeast TK that convert only 2 U/mg and 9 U/mg respectively (Sprenger et al., 1995; Demuynck et al.,1991; Rohmer et al., 1993). In this context, *EcoTK* is seen as having an advantage over the other two TKs for industrial chemical syntheses. Recent studies have shown that industrial use of transketolase is possible. Process engineering and immobilisation techniques have extended the use of TK; however, the main progress have been achieved by protein engineering. One of the main disadvantages for the use of TK at industrial scale is the fact that the affinity of TK for nonphosphorylated substrates remains very low, compared to the natural phosphorylated substrates. In this sense, for a successful application of TK on a larger scale, this problem needs to be resolved (Ranoux et al., 2012).

The application of TK has been seen in different products such as the production of 6-deoxy-L-sorbose, a precursor of the caramel flavoured furanol with a 45% yield. Another example is the production of 4-deoxy-L-threose from the whole cells of *Comebacterium equi* or *Serratia liqufaciens* using 4-deoxy-L-erythrulose and Li-HPA as substrates (Hecquet et al., 1996) Other biotransformations using TK include the synthesis of the non-natural sugar 4-deoxy-D-fructose-6-phosphate with a 52% yield using *SceTK;* X5P with 82% yield, and 3-O-benzyl-D-xylulose, used in turn for the synthesis of the potential glycosidase inhibitor *N-hydroxypyrrolidine* (Humphrey et al., 2000).

1.13 Thiamine Diphosphate (ThDP) the main Cofactor of TK

Thiamine diphosphate (ThDP) is a thiamine (Vitamin B1) derivative which is produced by the enzyme thiamine diphosphokinase and acts as an essential cofactor that participates in the carbon-carbon bond breaking and formation in a wide range of biochemical reactions. ThDP works as a coenzyme in many enzymatic reactions such as transketolase, pyruvate dehydrogenase complex, pyruvate decarboxylase, alpha-ketoglutarate dehydrogenase complex and many others (Voet et al., 2008). Chemically, ThDP consists of an aminopyrimidine ring which is connected to a thiazole ring, which is turn is connected to a pyrophosphate (diphosphate) functional group (see Figure 1.14). The part of the ThDP that normally is involved in enzyme reactions is the thiazole ring, which contains nitrogen and sulphur. The C2 of this ring, can act as an acid by donating a proton and forming a carbanion (Pavia, 2006).

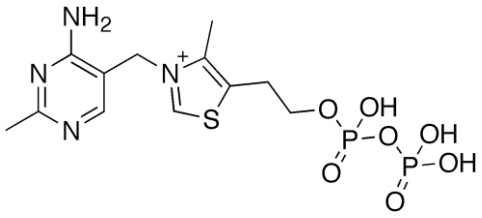


Figure 1.14. Structure of thiamine diphosphate (ThDP)

ThDP-dependent enzymes are involved in various types of reaction *in vivo,* for example in the nonoxidative decarboxylations of α-keto acids, oxidative decarboxylations of α-keto acids as well as cleavage of C-C bonds (Costelloe et al., 2008). ThDP enzymes have been found in all kingdoms of life; they form a diverse protein family that have been classified into nine super families (Vogel & Pleiss, 2014). These enzymes are promising to catalyse reactions that could be useful for the biotechnology industry because they can catalyse a broad range of reactions including C-C, C-N, and C-O bond cleavage and formation (Muller et al., 2013). The cofactor ThDP is bound at the interface between the PYR and the PP domain.

1.14 Development of TK Variants

TK from yeasts, plants and bacteria have shown a broad range of substrate specificities, that is defined as the preference of an enzyme for one substrate over other competitor substrates. In this context TK has been shown to have activity with sugars donors as the ones mentioned in Figure 1.10, as well as dihydroxyacetone phosphate (DHAP), dihydroxyacetone (DHA) and propanaldehyde (PA). In this context, TK could be a suitable candidate for protein engineering, since a small number of amino acid variations can modify the substrate specificity.

In 2008, Smith et al., showed that active site single point mutation is an effective technique to enhance the stereoselectivity of TK for non-α-hydroxylated acceptors. Jahromi et al., (2011) reported a relation between the initial specific activity and temperature-time for holo-TK, where it was shown that the enzyme lost the activity at 60 °C for 30 minutes. The same authors reported the effect of pH on holo-TK activity where the protein is more tolerant to deactivation at high pH than a low pH, with 50% activity remaining even at pH 11, while at pH 5, TK losses more than 50% of the activity.

Hibbert et al., (2007) used directed evolution to improve the activity of *E. coli* TK towards the non-phosphorylated substrates glycolaldehyde and HPA and they found several variants with great improvement in activity. In the same study it was removed the C-terminal domain of the R520Stop mutant and surprisingly the enzyme kept its activity. This mutation removes the C terminal and 18 residues of Pyr domain. In other words, the function of the C-terminal domain remains unclear as it does not contribute to the active-site formation, or cofactor binding in any way. Another strategy for improvement of desirable characteristics is the combination of different mutations that could have a synergistic effect in terms of activity as it was reported by Strafford et al. (2012). However, there is a possibility that the proteins lose their functionality, solubility, or stability (Strafford et al. 2012). As it has been shown there are several advances in protein engineering, nevertheless, one point that must be improved is the conversion yields with different aldehyde acceptors (Cázares et al., 2010).

1.15 Possible Uses of Hydroxyketones

The asymmetric α-hydroxyketone obtained from the TK catalysed reaction between Li-HPA and PA is a valuable building block that is difficult to prepare chemically, moreover they are versatile building blocks for the pharmaceutical and chemical industry for production of many biologically active compounds, drugs and fine chemicals (Ranoux et al., 2012). The first and most famous commercial application of 2-hydroxyketone as a building block in the pharmaceutical industry was in 1930's to produce (R)-phenylacetyl carbinol ((R)-PAC) by fermentation of yeast in presence of benzaldehyde (Neuberg & Hirsch, 1921). This product is also a key intermediate for the synthesis of norephedrine, adrenaline, amphetamine, methamphetamine, phenylpropanolamine and phenylamine (Shukla & Kulkarni, 2000).

An interesting potential application of hydroxyketones may be in the pharmaceutical industry, through the combination of their formation using transketolase, with the subsequent formation of 2-amino 1,3 diol using transaminase. This may in future be useful after a final conversion into chloramphenicol analogues, (see Figure 1.15) that would represent an important route to diversification of this antibiotic, and hence a wide range of antimicrobial applications.

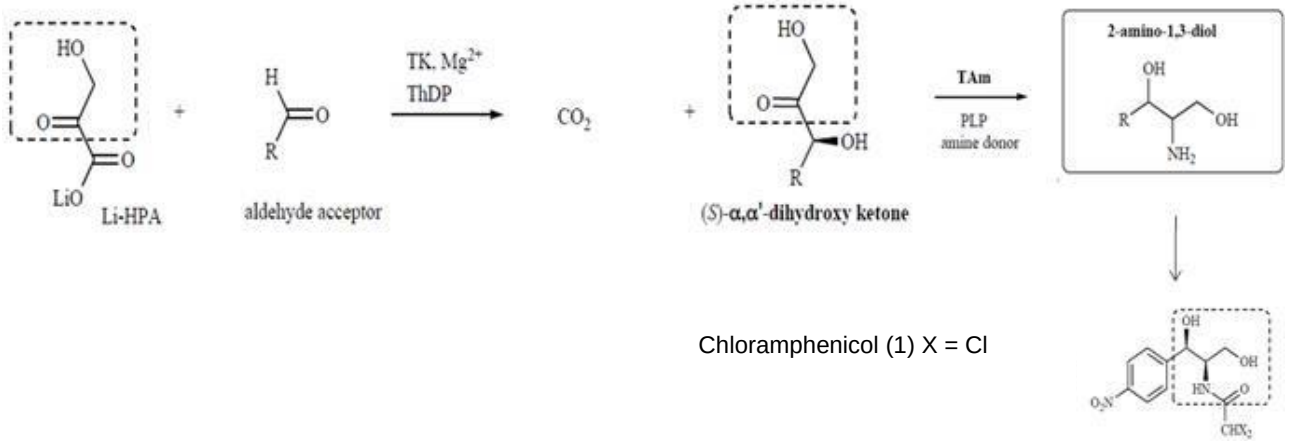


Figure 1.15. TK reaction within lithium β- hydroxypyruvate (Li-HPA) and an aldehyde acceptor to produce an α, α'-dihydroxy ketone product. Which it is used by the transaminase (TAm) to produce 2-amino 1,3-diol that is a building block for production of antibiotics such as chloramphenicol **(Hailes et al., 2010).**

Hydroxyketones have also been reported, with different properties as

antidepressants, inhibitors of amyloid-β protein production (used in the treatment of Alzheimer's), and antitumor antibiotics. They also have been considered as fine chemicals, because of their ability as building blocks to produce larger molecules. They have seen as a highly valuable building blocks for many applications for the fine chemistry sector for example the production of structures like amino alcohols and diols (Hoyos et al. 2010).

1.16 Colorimetric Assays and Tetrazolium Salts

Colorimetric methods have been widely used in high-throughput screening of enzymes, and also for basic measurements of reaction kinetics. In the specific case of TK, activity is normally monitored by an HPLC assay that measures the concentration of the bioconversion products in a protocol that takes 1-15 mins per sample. This method is very sensitive and shows high accuracy results. However, the number of samples that could be analysed is limited by the HPLC and auto sampler equipment available. An alternative assay reported by Smith et al., (2006), measured the 2-hydroxyketone from the reduction of red tetrazolium (colourless), that leads to its corresponding formazan with an intense red colour that can be measured at 485 nm (see Figure 1.16). The tetrazolium red (tetrazolium chloride) assay was used previously for identification and screening of mutants that produce 1,3-dihydroxypentan-2-one from the bioconversion with TK using β-HPA and propionaldehyde as a substrate (Smith et al., 2008; Cázares et al., 2010).

There are several kinds of tetrazolium salts available, such as MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide), XTT (2,3-bis(2-methoxy-4-nitro-5-sulphophenyl)-5-carboxanilide2H-tetrazolium, monosodium salt), tetrazolium red (2,3,5-tripenyltetrazolium chloride) and WST-1 (2-(4-iodophenyl)-3-(4-nitrophenyl)-5-(2,4-disulfophenyl)-2H-tetrazolium, monosodium salt) among others. However, their toxicity to users, solubility, and stability, must be considered for each specific case and intended use. Tetrazolium chloride is a redox indicator commonly used in biochemical experiments such as the indicator for cellular respiration. It is a white crystalline powder, soluble in water, ethanol, and acetone but insoluble in ether. Tetrazolium chloride reacts with hydroxyl ketones, forming the corresponding tetrazolium formazan. Nevertheless, there have to date been no reports on either the limits of detection or the quantitative constraints of the method as applied to TK-catalysed reaction monitoring or screening.

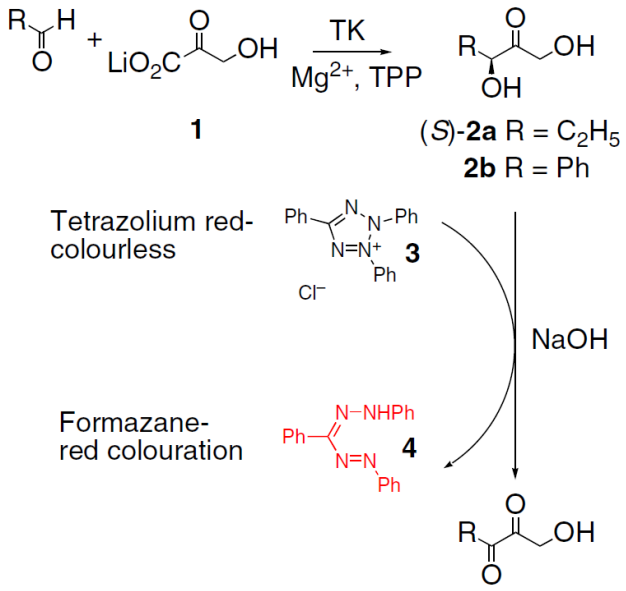


Figure 1.16. Schematic reaction of the TK bioconversion that produces a 2-hydroxyketone and the colorimetric reaction between tetrazolium red (which is colourless) and the 2-hydroxyketone to form the corresponding formazane which has an intense red colour and is measured at 485 nm **(Smith et al., 2006).**

This assay has the potential to be used for screening chemical reactions or bioconversions rapidly, and with a relatively low cost compared to HPLC. However, red tetrazolium salt, decrease its water solubility once the formazan is formed, which causes a practical problem, as it precipitates giving errors in the measured values. In this case, it is not yet known how well this assay correlates to analytical HPLC measurements over a wide range of hydroxyketone concentrations likely to be encountered in different enzyme screenings. It is also not known how stable the assay and readout will be over these ranges. This is important to assess and understand the quantification limits as well as the reliability of the assay. An alternative to solve this problem is the selection of a tetrazolium salt that generates a water-soluble formazan that avoids absorbance errors during measurements. Water-soluble tetrazolium salts (WSTs) were developed by introducing positive or negative charges and hydroxy groups to the phenyl ring of the tetrazolium. Positive charges, such as trialkylammonio groups, improve the water-solubility of the formazan dye. However, a large cation is easily precipitated out with organic anions such as carboxylate or phosphate.

The use of tetrazolium salts is diverse, for example the MTT-based assay is a common method to detect the level of cell metabolism. Meanwhile, it has also been used to develop a new colorimetric method to screen and measure transaminase

activities using aldehydes and ketones as substrates among others (Yin, et al., 2012; Baud, 2015). In this context, colorimetric methodologies have demonstrated their potential as enzymes screening methods in directed evolution strategies. Other tetrazolium salts such as tetrazolium blue (3,3'-[3,3'-Dimethoxy-(1,1'-biphenyl)-4,4'-dyl]-bis [2,5-diphenyl-2H-tetrazolium] dichloride) can be reduced by acyloins and other compounds with similar reduction potential. However, tetrazolium blue turned out to be inappropriate due to the instability of the formazane that precipitates. The precipitation dramatically affected photometric quantification, making unsuitable for reliable measurements (Breuer et al., 2002).

Tetrazolium salt WST-1 could be reduced through an electron mediator (see Figure 1.17). WST-1 is a highly sensitive tetrazolium reagent that produces a water-soluble and very stable formazan (compared to 2,3,5- Triphenyl tetrazolium chloride, tetrazolium violet and nitro tetrazolium blue chloride) with a solubility of 10 mg/mL in water and 0.65 mg/mL in 50 mM Tris buffer, pH 8.0 (Miyazono et al., 2000; Yin et al., 2012).

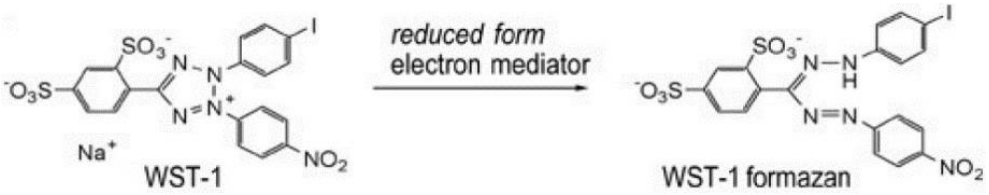
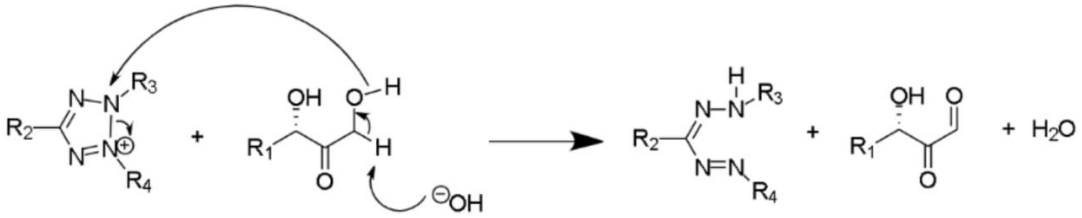


Figure 1.17. Formation of the WST-1 formazan from WST-1 tetrazolium **(Ishiyama et al., 1993).**

The WST-1 assay proposed in this project, can be used to identify and quantify transketolase activities against Li-HPA and PA in a high-throughput format that enables the rapid selection of new TK variants with specific properties such as thermostability. This method shows a low level of background coloration and does not generate a precipitate. An interesting property of tetrazolium salts is their ability to receive a hydrogen, via an intermediate electron transport such as reduced nicotinamide-adenine dinucleotide (NADH) or a hydroxyketone which produces the corresponding formazan. In most of the cases a nucleophile is necessary such as ^{-}OH that donates a pair of electrons to form a chemical bond. In this project NaOH [3 M] was used as a nucleophile (also known as a Lewis base) that is attracted to a full or partial positive charge of the hydroxyketone that activates the mechanism of reaction where the electron pair makes a double oxygen bond, and a H^{+} is liberated to continue

the mechanism of reaction with WST-1 as it is shown in Figure 1.18. The mechanism of reaction of this colorimetric method, shows the oxidation of the 2- dihydroxyketone in the presence of WST-1 and NaOH, resulting in the reduction of the WST-1 to the corresponding water-soluble coloured formazan. The TK activity can be determined by the spectrophotometric quantification of 1,3-dihydroxypentan-2-one at 600 or 450 nm. The assay steps of this proposed colorimetric method are easy to perform and the assay time is significantly shorter, allowing the measurement of thousands of samples per day compared with HPLC that would take much longer times to analyse that many samples.



(1) (2) (3) (4) (5)

Figure 1.18. Reaction mechanism of the colorimetric assay between WST-1 (1) and 2 hydroxyketone (2) conversion in alkaline condition using NaOH [3M]. NaOH is the nucleophile that starts the reaction and in consequence WST-1 is reduced to the corresponding formazan (3) with a specific colour that changes depending on its concentration, the corresponding oxidized hydroxyketone (4) and water (5).

Objectives

This thesis explored several objectives to provide the basis for production of mutant TKs using different growth conditions, the generation of an effective screening method, the development of mutagenic libraries and truncated TK versions.

Chapter 3 aims to explore the parameters and conditions for different volumes and media for *E. coli* fermentation as well as the determination of the best harvest point for maximizing TK expression and to analyse the pros and cons of different protein extraction methods.

The objective of Chapter 4 is to develop a stable colorimetric high-throughput (HTP) screening method able to identify and quantify the activity of transketolase variants over a wide range of product concentrations. The HTP method aims to be suitable to screen thermostable variants from mutagenic libraries, in a simple, efficient, and rapid way.

The objective of Chapter 5 is to develop the conditions to generate mutagenic transketolase libraries with low mutation rate via directed evolution, using different techniques of error-prone PCR and MEGAWHOP.

The objective of Chapter 6 is to explore the possibility of generating truncated and active transketolase variants by gradually removing the Pyr-domain from the C-terminal end utilizing designer stop codons.

2 Materials and Methods

2.1 Laboratory equipment

The water used in all experiments was purified using an Elix 5 water purifier (Millipore, UK) and de-ionised afterwards. This water is called MilliQ Water (MQ Water). Growth Media and items that must be sterilised were sterilised using an autoclave (Thermo Electron, UK) under 15 psi for 20 minutes. All buffers and solutions were pH adjusted using solutions of NaOH and HCl 1M and measured using a pH meter (Thermo Scientific, UK) previously calibrated. Experiments that were carried at room temperature (RT) refer to a temperature controlled between 22.5 °C and 23.5 °C inside the laboratories.

2.2 LB Media

LB media was prepared into a final volume of 1 L. Tablets or components were added to MQ water solution and mixed thoroughly before autoclaving. pH was adjusted to 7.0 when necessary. Before autoclaving, the solution was mixed thoroughly. This media is commonly used for cultivation, maintenance, and high-density growth of *E. coli.* Two different LB compositions were used in this project and their components are described in Tables 2.1 and 2.2.

Table 2.1. LB Broth composition (Tablets).

Component	g/L
Enzymatic digest of casein	10
Yeast extract (low sodium)	5
NaCl	5
Inert agents that are necessary for the tableting process	2

Table 2.2. LB-Broth composition.

Component	g/L
Peptone from casein	10
NaCl	10
Yeast extract	5

2.3 Terrific Broth Media

This media was used for higher density growth, protein, and plasmid DNA production in *E. coli.* Its composition is described in Table 2.3.

Table 2.3. Terrific Broth composition (TB).

Component	g/L
Tryptone	12
Yeast Extract	24
Potassium Hydrogen/Phosphate Dibasic	9.4
Potassium Phosphate Monobasic	2.2

2.4 Microbial Medium Powder

This is the most common growth medium for phage production and *E. coli* cultivation. Table 2.4 shows the composition of it.

Table 2.4. Microbial Medium Powder composition.

Component	g/L
Tryptone	16
Yeast Extract	10
NaCl	5

2.5 Magic Media *E. coli* Expression Medium

Magic Media from (MM) was designed to dramatically increase the yield of recombinant proteins in T7-regulated *E. coli* expression systems. This formulation enables *E. coli* growth to reach culture densities 3 to 10-fold higher than those achieved with traditional LB media.

2.6 Buffers

The next tables describe the composition of the buffers that were used in the project.

Table 2.5. Lysis buffer composition with lysozyme.

Component	mM	mg/mL
Tris buffer HCl pH 7.0	200	-
NaCl	500	-
Imidazole	9	-
TPP	2.4	-
Lysozyme	-	10

Table 2.6. Purification buffers for 6xHis-tagged Proteins using Ni-NTA columns (Superflow BioRobot). pH was adjusted to 8.0 using NaOH.

Wash Buffer
50 mM NaH_2PO_4
300 mM NaCl
10 mM imidazole

Elution Buffer
50 mM NaH_2PO_4
300 mM NaCl
250 mM imidazole

Table 2.7. Cofactors buffer, 10x.

Component	mM
Tris buffer HCl pH 7	50
ThDP	25
$MgCl_2$	90

Table 2.8. Buffers for protein purification using EGTA for Ni-NTA columns.

Tris Buffer	Elution Buffer EGTA
	500 mM EGTA
50 mM Tris HCl	500 mM NaCl
	10 M Tris HCl pH 8.5

Table 2.9. Tris-glycine-SDS (TGS) Buffer for SDS PAGE gel electrophoresis.

Component	Concentration
Tris buffer HCl 8.3	25 mM
Glycine	192 mM
SDS	0.1% (w/v)

2.7 Antibiotic stocks (1000x)

Antibiotic stock solutions were prepared using MQ water and ampicillin sodium salt (Sigma Aldrich) and sterilised using a syringe filter (Thermo Scientific Nalgene) with a pore size of 0.2 µm. Final concentration of the stock solution was 150 mg/mL. Finally, the solution was aliquoted in Eppendorf tubes of 1.5 mL, and stored at -20 °C.

2.8 Agar Plates

Agar plates with LB and 150 µg/mL of ampicillin as a final concentration were used to isolate colonies of *E. coli*, harbouring the pQR791 plasmid that contained the *tktA* gene expressed from its endogenous promotor.

Table 2.10. Agar LB-Broth.

Component	g/L
Peptone from casein	10
NaCl	10
Yeast extract	5
Agar	15

2.9 Microorganisms

Wild type transketolase, and mutants of transketolase introduced into the *tktA* gene in plasmid pQR791 (Ward, 1995; Martinez-Torres, et al., 2007), were expressed in *E. coli* XL10Gold (Stratagene, La Jolla, CA) with the protein containing an N-terminal His-6 tag for purification using Ni-NTA columns. All TK variants used on this project were inserted on the plasmid pQR791 (Ward et al., 1995), (see Figure 2.1). This plasmid pQR791 contains Bg1 II restriction site in the TK gene promoter region, and an N-terminal His_6 tag on the TK, to purify it using Ni-NTA columns. Because the TK plasmid (5.4 Kbp) in *E. coli* is not controlled by the *lac or tac* promotor/operator the addition of IPTG was not necessary to induce protein expression. Ampicillin-resistance gene that contains the pQR791 gene allows to selective growth of *E. coli*.

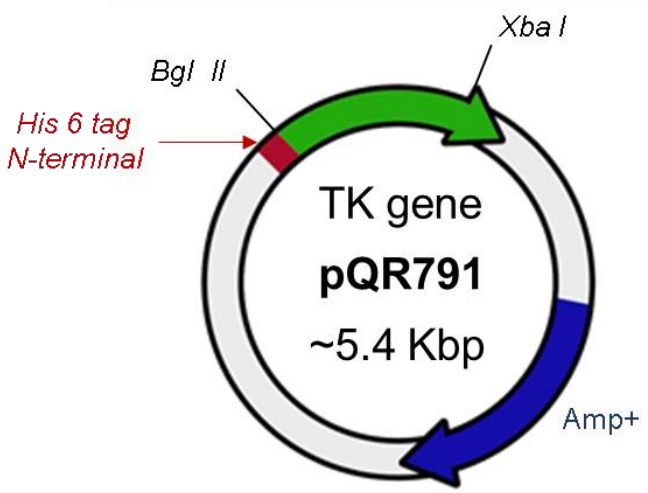


Figure 2.1. Schematic representation of plasmid pQR791. The ampicillin resistance (Amp^+) is represented in blue; the *tkt* gene is represented in green, and the His tag tail is coloured in red. The total plasmid size is 5.4 Kbp.

2.10 Glycerol Stocks

Glycerol stock cultures were prepared from a single *E. coli* colony of each TK variant from agar plates previously incubated for 12 hours at 37 °C. This was then inoculated into sterile 50 mL Falcon tubes containing 20 mL of LB media with the corresponding antibiotic (Ampicillin 150 µg mL^{-1}). TK cultures were taken off the incubator during the log phase of growth, when the culture reached an OD_{600nm} ~ 0.5, 500 µL aliquots of cell broth were mixed aseptically with 500 µL of a glycerol solution 40% (v/v) previously sterilised using a Syringe Filter 0.2 um (Thermo Scientific Nalgene) into a cryovial tubes, to a final volume of 1 mL. Tubes were mixed softly and then stored at -80 °C until required.

2.11 TK Production

Fermentations were started from glycerol stocks and 10 mL of LB Broth in a sterile Falcon tube of 50 mL, with 150 µg/mL ampicillin as antibiotic (constant antibiotic concentration for all dishes plates, pre- inoculum and fermentations) and incubated at 37 °C and 250 rpm (standard conditions) on an Innova 4330 orbital shaker (New Brunswick Scientific, Germany) for 16 hours (overnight) to produce the pre-inoculum with an OD_{600nm} of around 2.5. Fermentations in baffled flasks, were made in triplicate from the overnight pre-inoculum into LB media. Baffled flasks were immediately inoculated with pre-inoculum and an initial OD_{600nm} of ~0.1. The cultures were incubated at 37 °C and 250 rpm and were monitored by taking samples of 600 µL approximately every hour until the stationary phase was reached. Growth curves were made from different strains of TK glycerol stocks. All fermentations were carried out under aseptic conditions to avoid any contamination. Measurements of optical density at 600 nm were made using a Nanodrop 2000 (Thermo Fisher Scientific). For OD measurements, samples over 1 OD 600 nm were diluted with sterile LB medium to get values between 0.1 and 1 that is the optimal range of value for spectrophotometer measurements.

For high production of TK, Magic Media was used for fermentations of different variants including WT, D469T, H192P/A282P/I365L/G506A carried out in 1 L baffled flasks using with a working volume of 200 mL. All flasks with Magic Media were inoculated using 10 mL of pre-inoculum with an OD_{600} ~ 0.75. Fermentation was carried out at 37 °C, 250 rpm, for 17 hours. The total volume of fermentation was divided into four Falcon tubes of 50 mL, and centrifuged at 5000 rpm, 4 °C for 1 hour using a Hettich Universal 320R Centrifuge (Buckinghamshire, UK), and the supernatant was discarded. Around 3.5 g of wet pellet were obtained from each tube and resuspended in 20 mL of Tris buffer (50 mM, pH 7.0), with cofactors solution (25 mM ThDP, 90 mM $MgCl_2$), and sonicated as described in Section 2.12. Moreover, TK was purified using Ni-NTA columns as described in Section 2.15. Total protein was measured using the Bradford method as described in Section 2.19.

Finally, TK was also produced in 96 and 24 well plates using different growth media, LB media, LB media with 7.5% of glycerol, Terrific Broth and MM A single colony from the D469T TK variant was picked and inoculated in 1 mL of the different growth media in 96 well plates. Fermentation was carried out at 37 °C, 500 rpm, for 24 hours. Moreover, the same process was carried out in 24 well plates with a total volume

of 3 mL. Samples of 100 µL were taken from the plates every hour, to measure their optical density at 600 nm using a Nanodrop 2000 (Thermo Fisher Scientific).

2.12 Protein Extraction by Sonication

After bacterial fermentation, cells were recovered by centrifugation (5000 rpm, 4 °C, 60 minutes). The cell pellet was resuspended in cofactor buffer, using only 1/10 of buffer with respect to the initial sample volume. Resuspended cell solution was placed on ice for sonication. Samples were sonicated with MSE Soniprep 150 sonicator (Sanyo, Japan) with 25 cycles (10 seconds ON – 15 seconds OFF) with a probe frequency of 20 kHz. Once the sonication process was finished, lysate was recovered by centrifugation at 13000 rpm, 4 °C, for 30 minutes to remove cell debris, and the clarified lysate was aliquoted into 1.5 mL Eppendorf tubes. Lysates were stored at -20 °C and used within 1 month. Before use, lysates were thawed in a water bath at 37 °C for 2 minutes.

2.13 Protein Extraction by Bug Buster Reagent

The pellet from fermentation was weighed and then resuspended at room temperature using 5 mL of Bug Buster per gram of wet cell paste. Mixed solution was incubated on a platform or rotating mixer for 20 minutes. Afterwards, the solution was centrifuged at 15,000 rpm for 30 minutes at 4 °C. The supernatant containing the TK protein was ready to use directly for biotransformation or purification with Ni-NTA columns.

2.14 Protein Extraction by Freezing-Thawing

Cell pellets were frozen by submerging the microfuge tube in a dry ice/ethanol bath for 2 minutes. The sample was then thawed by transferring it to an ice/water bath for 8 minutes. Thawing was done at 0 °C rather than at elevated temperatures to minimize the possible degradation of sensitive samples. The cycle was repeated 2 additional times (Johnson et al., 1994).

2.15 Protein Purification by Ni-NTA Columns

Transketolase was purified using QIAGEN Ni-NTA spin columns (Cat. No. 31014) following the conditions from the supplier. Each Ni-NTA spin column was equilibrated with 600 µL of wash buffer (see Table 2.6) and centrifuged for 2 minutes at 2900 rpm in a bench-top centrifuge. After that, the column was loaded with up to 600 µL of clarified lysate containing the 6xHis-tagged protein, into the pre-equilibrated Ni-NTA spin column, and then centrifuged for 5 minutes at 1600 rpm. The flow-through was

collected for further measurement of protein concentration and analysis by SDS-PAGE electrophoresis. Subsequently, the Ni-NTA spin column was washed two times with 600 µL of wash buffer, with the spin column centrifuged each time for 2 minutes at 2900 rpm.

The final elution of the protein was carried out with 300 µL of elution buffer (see Table 2.6), two times, with centrifugation at 2900 rpm for 2 minutes. An alternative elution buffer containing EGTA (see Table 2.8) was also used for protein purification. Different concentrations (5 to 500 mM) of EGTA at pH 8.5 were prepared. These buffers were used for the elution of the TK instead of using the standard elution buffer containing imidazole.

2.16 SDS-PAGE Protein Gel

SDS-PAGE analysis was performed using Mini-PROTEAN Gels 12% Tris-glycine acrylamide gradient gels (BIORAD). The culture samples were prepared by mixing 21 µL of each cell suspension and 7 µL of 4x Laemmli Sample Buffer (BIORAD), with an additional 0.2 M dithiothreitol (DTT) (Sigma-Aldrich), to denature the protein by reducing disulphide linkages within the tertiary and quaternary protein structure. Solutions were mixed well and heated at 100 °C for 5 minutes. Samples of 20 µL of the dye-sample mixture were loaded in each gel well, and electrophoresis was performed in 1x Tris/Glycine/SDS (TGS) buffer (see Table 2.9), at 200V for 45 minutes. In each gel 5 µL of marker was loaded in the first or last lane. The protein bands were stained with Instant Blue (Eppendorf) for 20 minutes, and then washed with MQ H_2O three times. The gels were finally left in water overnight. The amount of protein loaded in each lane was between 0.024 and 0.03 mg of protein. SDS gels were imaged using an Ultraviolet TransIluminator (Bioimaging system, UVP, Kansas EE. UU).

2.17 Synthesis of Li-HPA

Li-HPA was synthesised using LiOH and bromo pyruvic acid following the method described by Morris et al., 1996. The method involved addition the reaction of LiOH (1M) with bromo pyruvic acid 5% until it reached a pH of 9.5. Once the desired pH was obtained, concentrated glacial acetic acid was added drop by drop until a pH of 5 was reached. The solution obtained was transferred to a 500 mL round bottom flask and concentrated in a rotary evaporator under vacuum at 30 °C until it reached half of the initial volume. Concentrated solution was stored in the fridge overnight. After that, the solution was washed using an Erlenmeyer flask and ethanol under vacuum. Filtrate was recovered and washed again with ethanol. The obtained product was dried in a drying chamber during 48 hours at room temperature. Li-HPA was stored at -20 °C until used for bioconversions.

2.18 Conditions of Bioconversions

Different conditions for the biotransformation using transketolases were tested on this project. In this section general conditions of bioconversion are described.

2.18.1 Short Bioconversions

The overall volumes of bioconversion were often changed. However the volume ratio of reagents was kept constant in the majority of the bioconversions, with: 30% of TK cell-free lysate [~ 1 mg/mL], 30% of Li-HPA (100 or 200 mM in 50 mM tris HCl, pH 7.0), 30% of propionaldehyde (100 or 200 mM in tris-HCl, pH 7.0) and 10% of 12 x cofactor stock solution (25 µL of 28.8 mM TPP, 108 mM $MgCl_2$ in 50 mM tris-HCl, pH 7.0). Reactions were mixed and incubated in Eppendorf tubes at room temperature, for 24, 48 or 72 hours, at 500 rpm at room temperature using a thermomixer.

2.18.2 Long Bioconversions

Long bioconversions were tested with lysate and pure TK from H192P/A282P and measured by HPLC. Concentration of substrates for different bioconversions are shown in the following tables.

Tables 2.11. Bioconversion volumes and concentrations.

Reactive or substrate	[Initial]	Volume [µL]	[Final]
TK Lysate (Total Protein)	0.74 mg/mL	800	0.22 mg/mL
Li HPA	200 mM	800	60 mM
PA	200 mM	800	60 mM
Cofactor solution	12 x	280	1x

Reactive or substrate	[Initial]	Volume [µL]	[Final]
TK Lysate (Total Protein)	5.8 mg/mL	800	1.72 mg/mL
Li HPA	200 mM	800	60 mM
PA	200 mM	800	60 mM
Cofactor solution	12 x	280	1x

Reactive or substrate	[Initial]	Volume [µL]	[Final]
Control no TK, H_2O	-	800	-
Li HPA	200 mM	800	60 mM
PA	200 mM	800	60 mM
Cofactor solution	12 x	280	1x

All bioconversions were run at room temperature, for 90 hours, 800 rpm, in 2 mL Eppendorf tubes. After 90 hours, reaction was quenched with TFA (0.1%) and measured in HPLC as it was described in Section 2.21. Other bioconversions with TKs from WT and H192P/A282P/I365L/G506A were performed for 72 hours, at room temperature and 200 rpm, using the concentrations and volumes described in Table 2.12.

Table 2.12. Bioconversion volumes and concentrations.

Reactive	[Initial]	Volume [µL]	[Final]
TK Lysate	1 mg/mL	800	0.3 mg/mL
Li HPA	200 mM	800	60 mM
PA	200 mM	800	60 mM
Cofactor solution	12 x	260	1.2x

2.19 Protein Quantification by Bradford Assay

To determine the protein concentration of samples, a calibration curve of standard BSA at different concentrations (0, 0.002, 0.004, 0.006, 0.008, 0.01 mg/mL) was made for all the protein measurements. 300 µL of each concentration (by triplicate) was mixed with 300 µL of Coomassie Reagent and mixed well. Absorbance measurement at 595 nm was taken after 5 minutes of reaction, when the solution colour became stable. This calibration curve was used to determine the total protein concentration of TK samples. The same method was used, except that all samples were required to have an absorbance value in the range of the standard curve between 0 and 0.4 (at 595nm) to calculate the correct protein content. For samples with absorbance value >0.4, serial dilutions were necessary.

2.20 Colorimetric Method

The colorimetric method using WST-1 salt can identify the product 1,3-dihydroxypentan-2-one from the bioconversion between Li-HPA and PA. Once the reaction was finished, 50 µL of sample was mixed with 20 µL of 0.2% (w/v) WST-1 solution and 10 µL of 3 M NaOH. The reaction was mixed and measured in a plate reader at 450 nm, for orange-red coloured samples that indicated (3S)-1,3-dihydroxypentan-2-one (HK) presence, and at 595 nm after 10 minutes for blue coloured samples with concentration of HK <5mM. Samples were followed and measured every 5 minutes to determine the reaction behaviour.

2.21 High-pressure Liquid Chromatography (HPLC) Analysis

Samples from bioconversion of were analysed by HPLC (Ultimate 3000, ThermoFisher Scientific, US) with an Aminex HPX-87H ion exclusion column (300 x 7.8 mm) (Bio-Rad Laboratories, Hertfordshire, UK) at a constant temperature of 60 °C. Trifluoroacetic acid (TFA) 0.1% (v/v)) was used as a mobile phase at a constant flow rate of 0.6 mL/min. The UV absorption at 210 nm was monitored by Ultimate™ 3000 VWD variable wavelength detector (ThermoFisher Scientific, US). 180 µL of TFA 0.1% (v/v) and 20 µL of the bioconversion solution were mixed in a 96-well microplate for quenching the reaction before testing by HPLC. Reaction between Li-HPA and propionaldehyde has a product, 1.3-dihydroxy pentan-2-one (HK). Retention times for Li-HPA and HK, were 8.5 and 16 minutes, respectively. All samples were previously centrifuged at 13,000 rpm, 4 °C, for 30 minutes, to remove any precipitate.

2.22 Generation of Truncated Transketolases

Truncated versions of TK were generated using the stabilising TK double mutant H192P/A282P as a template for the PCR's. Different stop codons were introduced: 453Z, 425Z, 403Z, 375Z, 349Z, retaining the PP-domain, N-terminal 6-His tag (position 2393 from master sequence), but gradually cutting the PYR-domain from the C-terminal end, (see Table 2.13).

Table 2.13. Truncated versions and sequence showing the stop codons.

Truncated Variant	Sequence 5' → 3'
Q453Stop	CTGATGAAA**TAG**CGTCAG
G425Stop	TCCCTGCAC**TGA**GGCTTC
N403Stop	GCTGCGGGT**TAG**TACATCCAC
F375Stop	TTGCCGGAA**TAG**CTCGGCGGT
Q349Stop	GCTAAACTG**TAG**GCTAATCCG

Mutagenesis was carried out using the Kit Site-Directed Mutagenesis (Agilent Technologies) and truncated versions were transformed into XL10 competent cells. Digestion was carried out with 1.5 µL of enzyme Dpn-I for 2 hours at 37 °C. After transformation, NYT media was used to recover the cells for 1 hour at 27 °C, 250 rpm. 50 µL of each truncated variant was inoculated into LB agar with ampicillin (150 µg/mL) and incubated at 37 °C overnight. Three individual colonies of each truncated variant were selected and fermented in 5 mL of LB with ampicillin for further DNA extraction using Mini Prep columns. DNA was sequenced to confirm the mutations. The same fermentation was used to prepare glycerol stocks of generated mutants. Glycerol stocks were used for fermentations in 150 mL Magic Media at 37 °C, 250 rpm overnight. Extraction of total protein was made by sonication (see Section 2.12) and quantification measurements were performed using the Bradford method (see Section 2.19).

2.23 PCR conditions for Truncated TK variants

Components and conditions of PCR for generation of truncated TK variants are shown in the following tables. All primers used in this study were purchased from Eurofins Genomics, and they were designed about 20 – 30 bp in length with a predicted *Tm* between 56 to 65 °C.

Table 2.14. Mix of components for PCR.

µL	Component
5	10x Rxn buffer
1	DNA template H192P-A282P (15 ng/µL)
2	Primer Forward (10 pmol/ µL)
2	Primer Reverse (10 pmol/ µL)
1	dNTP mix
3	Quick solution
36	H_2O MQ

Table 2.15. PCR conditions for generation of truncated TK's.

Condition	Cycles	T [°C]	Time
Initial denaturation	1	95	1 minute
Denaturation	18	95	50 seconds
Annealing	18	60	50 seconds
Extension	18	68	5 min, 30 seconds
Final Extension	1	68	7 minutes
Final Hold	1	4	∞

2.24 Agarose Gel Preparation

Agarose gels were necessary to observe the PCR results. 0.3g of agarose was dissolved in 30 mL of TBE buffer 1x for an agarose solution of 1% (w/v). 3 µL of SYBR 10000x stain was added and mixed well. The solution was heated for 10 seconds in a microwave until the agarose was completely dissolved (4 to 5 cycles approximately). The solution was poured in a gel caster and left at room temperature until solid. When the gel was ready, 1x TBE buffer was poured into the electrophoresis chamber until the gel was completely covered. PCR samples were loaded into the gel mixing the following volumes, and run at 80 volts, for 45 minutes.

Table 2.16. Mix for PCR products and marker in agarose gel.

Sample	Sample [µL]
H_2O	8
6x loading dye	2
PCR or maker	2

All gels were also loaded with a standard solution of 1.1 kb that was mixed in the following ratio.

Table 2.17. Mix for standard DNA in agarose gel.

Sample	Sample [µL]
Standard 1.1 kb	5
6x loading dye	1

2.25 Mutagenic Libraries

The MEGAWHOP method consists of two PCR steps. In the first step, error-prone PCR was used for the generation of the megaprimer (<800 bp) with random mutations in the interested gene. The second PCR used the Megaprimer previously generated, and the whole pQR791 vector as the template. This second PCR (MEGAWHOP) makes an elongation of the Megaprimer and the whole plasmid, resulting in creation of large random mutagenesis libraries (Copp et al., 2014.). MEGAWHOP reactions were transformed into XL10 Gold competent cells as described in the QuickChange Kit protocol. Colonies were obtained in LB-agar with ampicillin, and three colonies of each mutagenic library were randomly selected for further sequencing. Ten primers were designed at different positions of the TK gene to generate the first Megaprimers with random mutations. The following tables described the Megaprimers used and PCR conditions for the mutagenic libraries generation.

Table 2.18. Position and sequence of Megaprimers.

Megaprimer	Position in TK sequence	Position in whole sequence	Sequence 5' → 3'	Note
Forward 1	-21 to 0	2365 to 2386	CATCATCAGATCT GGAGTCAAA	Starts outside TK
Reverse 1	573 to 595	2959 to 2981	CGTGACCATCGA TAGAAATACCG	-
Forward 2	573 to 595	2959 to 2981	CGGTATTTCTATC GATGGTCACG	Reverse complement of Reverse 1
Reverse 2	1001 to 1022	3387 to 3408	GACGGCATTTCG CCTTTCATAC	-
Forward 3	976 to 996	3362 to 3382	CAGGAAGCCGCT GAATTTACC	-
Reverse 3	1688 to 1708	4074 to 4094	CCAGTTCAACTTC TGAACCGG	-
Forward 4	1608 to 1629	3994 to 4015	AGAGCAACTGGC AAACATCGCG	-
Reverse 4	2053 to 2075	4439 to 4461	CCGCAAACGGAC ATATCAAGGTA	Finishes outside TK
Forward 5	-21 to 0	2365 to 2386	CATCATCAGATCT GGAGTCAAA	Same than Forward 1
Reverse 5	2053 to 2075	4439 to 4461	CCGCAAACGGAC ATATCAAGGTA	Same as Reverse 4

Table 2.19. PCR mix for Megaprimers generation.

µL	Component
5	10x Mutazyme II buffer
1	DNA template H192P-A282P (50 ng/ µL)
1	Primer Forward (25 pmol/ µL)
1	Primer Reverse (25 pmol/ µL)
1	40 mM dNTP's mix
1	Mutazyme II DNA polymerase (2.5 U/ µL)
40	H_2O MQ

Table 2.20. PCR conditions for Megaprimers generation.

Condition	Cycles	T [°C]	Time
Initial denaturation	1	95	2 minutes
Denaturation	30	95	30 seconds
Annealing	30	55	30 seconds
Extension	30	72	1 minute
Final Extension	1	72	10 minutes
Final Hold	1	4	∞

Table 2.21. PCR mix for MEGAWHOP generation.

µL	Component
10	PCR buffer HF 5x buffer
1	DNA template H192P-A282P (50 ng/ µL)
1	Megaprimer from first PCR
1	DMSO [4% final]
1	dNTP's mix (10 mM each)
1	Polymerase (2.5 U/ µL)
34	H_2O MQ

Table 2.22. PCR conditions for MEGAWHOP generation.

Condition	Cycles	T [°C]	Time
Initial denaturation	1	95	30 seconds
Denaturation	24	95	1 minute
Annealing	24	65	3 minutes
Extension	24	70	21 minutes
Final Extension	1	70	10 minutes
Final Hold	1	4	∞

3 Results: TK expression, Purification and Analysis

3.1 Introduction

Development of a high-throughput screening (HTS) method for screening of enzymes variants, requires efficient methods for enzyme production, purification and analysis that includes methods for protein purification, measurements of kinetic parameters, analysis of different media, among others. In this chapter evaluation of different growth conditions to improve TK production were tested providing an overview of key aspects for a reliable HTS method. In addition, different methods of protein extraction were evaluated to determine the most suitable methodIn addition, some methods for quantification and identification of TK were tested, including basic methodologies such as Bradford Method, SDS-PAGE gel and His-tag protein purification. These results provided parameters and conditions that must be considered to ultimate achieve robust and sensitive assays suitable for use throughout this project.

3.2 Importance of Media, Buffers and Reagents

Each microorganism requires a specific medium for its optimal growth and protein expression. In this sense, different growth media were used for specific purposes. It is important to have in mind that even small differences in the composition of the medium (pH, salts concentration, nitrogen, and carbon sources, etc.) can make dramatic changes mainly in protein production. Growth media must have a source of carbon that is required for biomass production (glucose or other organic compounds) and serves to supply energy. Similarly, a source of nitrogen (ammonium ions, nitrate ions, proteins, peptones, amino acids, etc.) is also required, and can help as a source of energy and microbial growth as well. In some cases, it is necessary to add phosphates, trace metals like magnesium, and vitamins or additional growth factors, that can increase cell density (Atlas, 2010).

One of the most essential factors in growth media is the pH; this must be controlled for each microorganism to get the required growth and protein production. During shake-flask fermentations, the pH can increase quickly to 9 or higher. At this pH, *E. coli* is unable to grow efficiently, or express protein, and will eventually lead to cell lysis. Moreover, some high-density growth media such as Magic Media (MM) can support growth to *E. coli* cell densities of OD_{600nm} ~ 40. This media normally contains trace metals, minerals, and vitamins, and are buffered by phosphates. The main advantage of this high-density growth media is that protein production can be 10 times more than in commonly used media such as LB broth. Buffers contain weak acid and bases, and their corresponding salt, and have a specific pH range that maintains the ideal conditions for optimal growth of microorganisms.

3.3 TK Growth Curves and Protein Extraction

Liquid media is the most common lab method to growth bacteria from a colony or a glycerol stock to any kind of fermentation for production of a specific protein such as transketolase. All *E. coli* fermentations require some basic nutrients for survival including vitamins and minerals (normally provided by yeast extract), as well as a source of nitrogen to aid in nucleic acid and amino acid production (usually provided by tryptone). Media used *for E. coli,* normally includes substantial concentration of glucose due that is an inexpensive and readily utilizable carbon source. During aerobic fermentation (with excess of glucose) the pH could decrease due to the formation of acid by-products, mainly acetate. In consequence aerobic acidogenesis or aerobic fermentation is a major factor in the limitation of high cell density growth (Luli et al., 1990).

On the other hand, composition difference in growth media allows scientists to optimize the growth rate, plasmid yield, transformation efficiency or protein production. Each strain has specific behaviour in terms of growth and protein production, in biochemical engineering laboratories, the most used growth medium for *E. coli* is LB 'lysogeny' or Luria broth. It is an excellent choice specially at small-scale because it promotes fast growth. Moreover, another medium used for *E. coli* growth is Terrific broth (TB), it is richer than LB and includes glycerol as an energy source, leading to faster growth and higher cell densities than for LB. Additionally, TB contains potassium phosphates that keep the optimum pH for a longer time, decreasing the chance of cell death during fermentation (Lessard et al., 2013). Other growth media such as the Magic Media (MM) promotes the high-yield growth of *E. coli,* this medium allows the regulation of protein expression in any expression system even if it is solely inducible by IPTG from *E. coli* strains. In this case, protein expression starts automatically after inoculation with a single colony or starter culture.

The most common solid agent that is used for growth of microorganisms and *E. coli* is agar, which is a polysaccharide extracted from marine algae. It has a melting point of 84 °C and solidifies at 38 °C (Atlas, 2010). Moreover, glycerol stocks provide an effective method to store and keep the bacterial strains for long term. All *E. coli* strains containing plasmids of TK variants were conserved using this technique. *E. coli* is widely used for production of proteins (enzymes). Because naturally *E. coli* lacks the ability to secrete recombinant protein into the extracellular space, the use of an extra step to lyse the cells is necessary to release the recombinant product. Nevertheless,

cell lyses can influence the downstream process, for example, increasing viscosity due to cytoplasmic DNA. The use of different methods of TK extraction were evaluated in this section to determine the pros and cons of each method. However, before the protein extraction, the growth kinetics must be known to harvest just before the maximum OD is reached, this allows to avoid premature cell lysis and consequently lower protein production (Voulgaris et al., 2016).

Results from this chapter compare the effect of different growth media using different strains of *E. coli* and their effect in the TK production, influence of extraction methods and quantification with the goal of determining the best conditions to harvest and extract TK. These results are fundamental to determine the best conditions of fermentation, extraction, and quantification of TK for further experiments in this thesis and future projects. WT TK was used and tested on these experiments as a control. All TK variants analysed on this chapter were previously obtained in the laboratory by other colleagues. These variants have shown interesting characteristics, for example R520Z does not have the C terminal and showed 3.6 times more activity for erythrulose production than WT (Hibbert et al., 2007). Meanwhile, D469E showed activity with different substrates such as benzaldehyde, 2-furaldehyde and 2-thiopene-carboxaldehyde. D469T was found to improve the specific activity towards propionaldehyde, and decrease it towards glycolaldehyde, when compared to WT. Moreover it has been shown to enhance and reverse the stereoselectivity of TK and improve the reaction towards non-hydroxylated substrates (Hibbert et al., 2008; Panwajee et al., 2012). Finally, this variant was found to have the highest activity towards the carboxylated aromatic substrates such as 3 formylbenzoic acid (3FBA) and 4 formylbenzoic acid (4FBA) (Panwajee et al., 2015).

3.4 Fermentations of *E. coli* for TK Production

Shake flasks are the most applied laboratory cultivation vessels for production of recombinant proteins due to their very easy set-up and operation. Erlenmeyer flask is now widely used to culture microorganisms, plant, and animal cells. Baffled shake flask culture is usually used to optimise culture conditions for aerobic microorganisms, it improves the oxygen transfer rate and helps to promote mixing. Some studies have shown a yield improvement of some specific recombinant protein by using baffled shake flasks compared with standard shake flasks (Takahashi et al., 2020; Ukkonen et al., 2011).

All fermentations were followed for the first 8 hours, taking samples approximately every hour. After reaching the highest OD, the fermentation was continued overnight, and the final OD measured after 24 hours from the initial time to have a reference and some points within the death phase. The objective of this initial experiment was to determine the best point to harvest, and according with Figure 3.1, it is around 7 hours of fermentation, using the conditions mentioned above. This growth curve was taken as a reference for further fermentations and WT harvest.

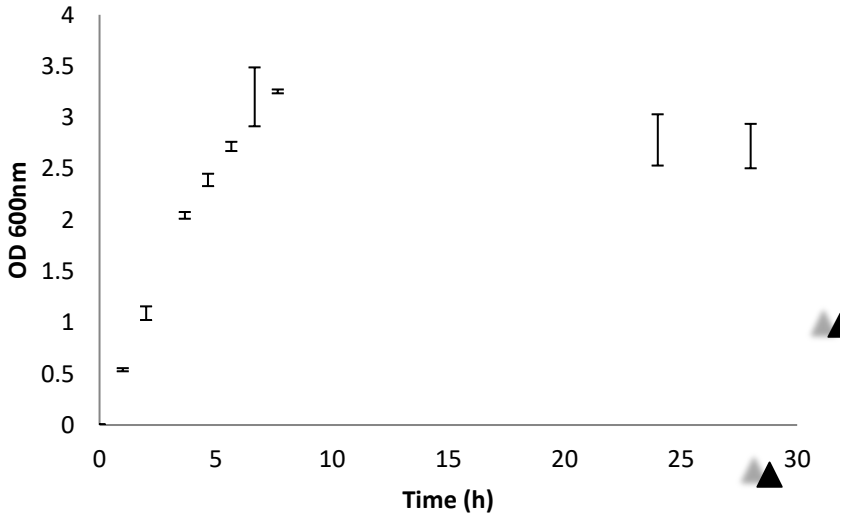


Figure 3.1. TK Fermentation growth curve of *E. coli* WT samples in LB media, in baffled flasks at 37 °C and 250 rpm over 28 hours. Fermentations were performed as described in section 2.11. Data were averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3).

3.5 Growth Curves of *E. coli XL10Gold* expressing TK variants G425Z, R520Z, D469E and D469T in LB

Variants with interesting characteristics such as G425Z and R520Z, with a truncated C-terminal, have been reported to have higher activity than WT using PA as a substrate. These variants were tested in standard conditions (see Section 2.11) in triplicate to determine their behaviour and the best harvest point during expression of both variants. Figure 3.2 shows that the behaviour of the cell cultures expressing both variants were very similar in all measured samples. These 2 variants reached around 0.5 higher OD values compared with WT. The best harvest point for expression of these variants was between 7 and 8 hours of fermentation, just before the stationary phase. Other interesting variants with higher activity compared with WT are D469E and D469T. These variants showed higher activity using PA and Li-HPA as a substrate instead of glycolaldehyde (Hibbert et al., 2008). As can be seen in Figure 3.3, the cells expressing variant D469E grew faster and got a higher final OD at 600 nm compared to those expressing D469T. Its best harvest point is also between 7 and 8 hours.

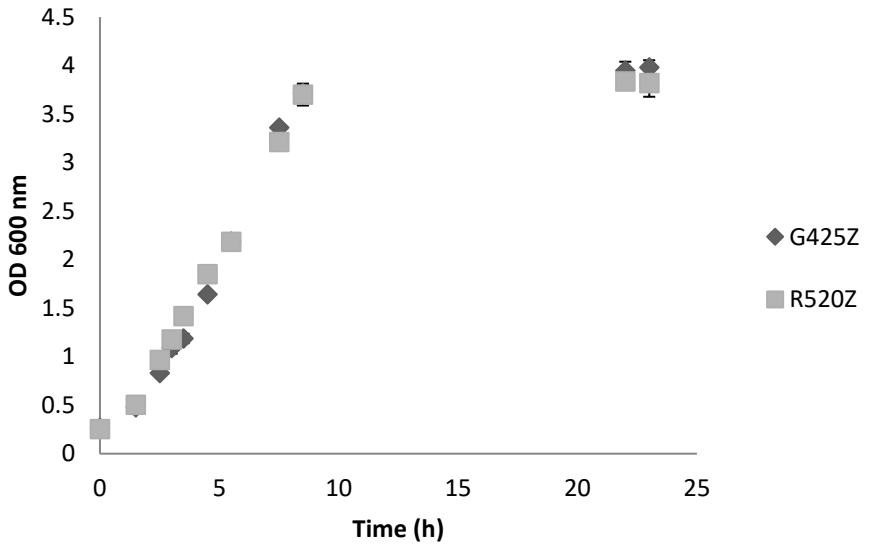


Figure 3.2. TK Fermentation curves of *E. coli* XL10Gold expressing TK variants G425-stop and R520-stop in LB media, baffled flasks at 37 °C and 250 rpm over 23 hours. Data points are averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3).

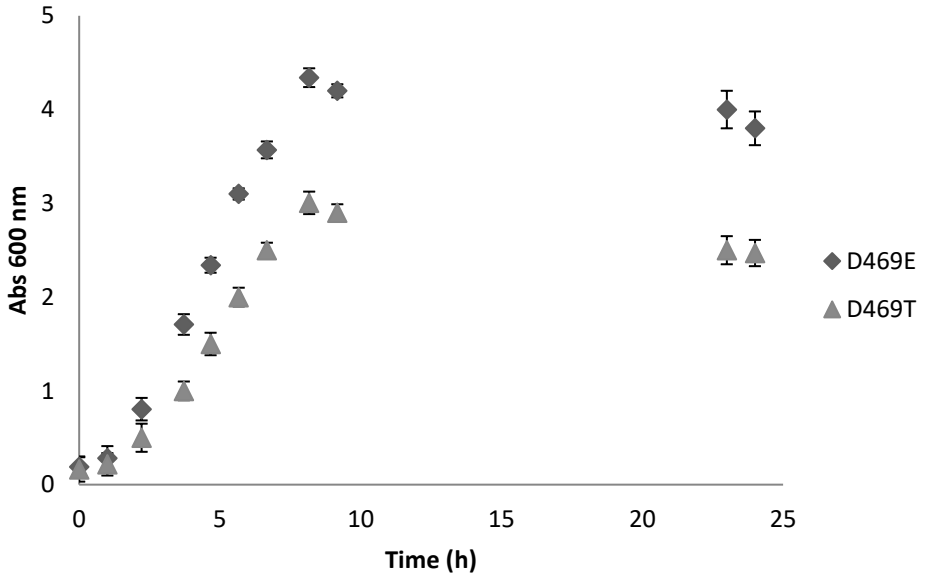


Figure 3.3. Fermentation curves of *E. coli* XL10Gold expressing TK variants D469E and D469T in LB media. Fermentation in baffled flasks at 37 °C and 250 rpm over 24 hours. Data points are averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3).

3.6 Calibration Curves of Total Protein Using Bradford Method

A simple and common method for total protein quantification is the one described by Bradford in 1976 (Bradford, 1976). This method involves the binding of Coomassie Brilliant Blue G520 to the proteins, causing a shift in the colour solution and in consequence in the absorption of the dye at 595 nm. This method has several advantages, including high reproducibility, low cost, fast results, and stability of the colouration for measurements even after 1 hour from the beginning of the reaction.

After fermentation and harvest total protein was quantified. Quantification of total protein was performed as described in Section 2.19. Before quantification of total protein for each sample, a calibration curve was made using five different concentrations of Bovine Serum Albumin (BSA) as standard protein (0.002, 0.004. 0.006, 0.008, 0.01 mg/mL). To know if the time of incubation of the Coomassie reagent affects the results, three times were taken (5, 30 and 60 minutes) after its addition and the results are shown in Figure 3.4.

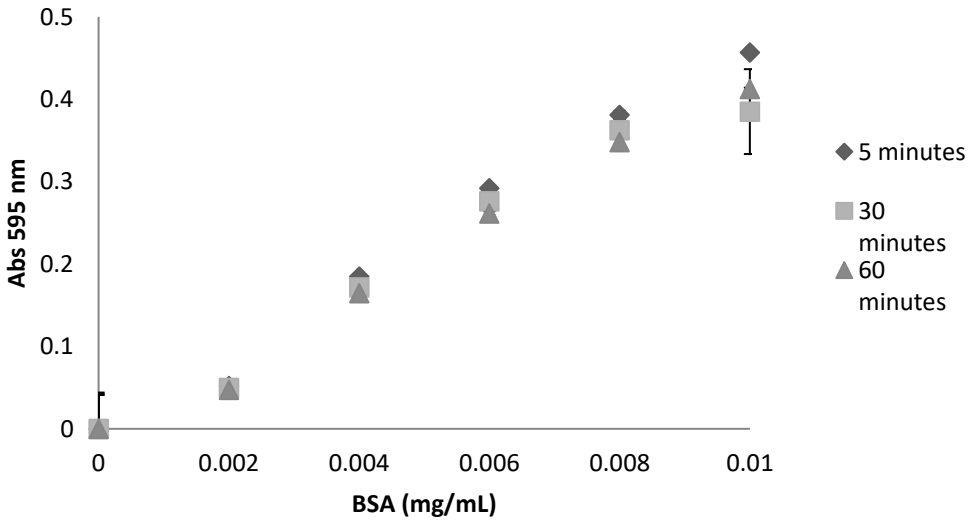


Figure 3.4. Calibration curve of protein using Bradford method and BSA as a standard protein. The samples were measured at different times after the addition of Coomassie reagent. Data were averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3).

Figure 3.4 shows that time had a low impact in the absorbance measurements

for the three stablished times. However, according with the method and the previous results, 5 minutes of reaction with Coomassie reagent was enough to quantify the total protein. Based on these results, all measurements for the Bradford assay were performed after 5 minutes of reaction as this also gave a good correlation with them known BSA standard concentration, with a R^2 of 0.99 (Figure 3.5). Figure 3.5 is an example of a standard protein curve however, to get reliable results, a new calibration curve was made for each experiment for protein quantification.

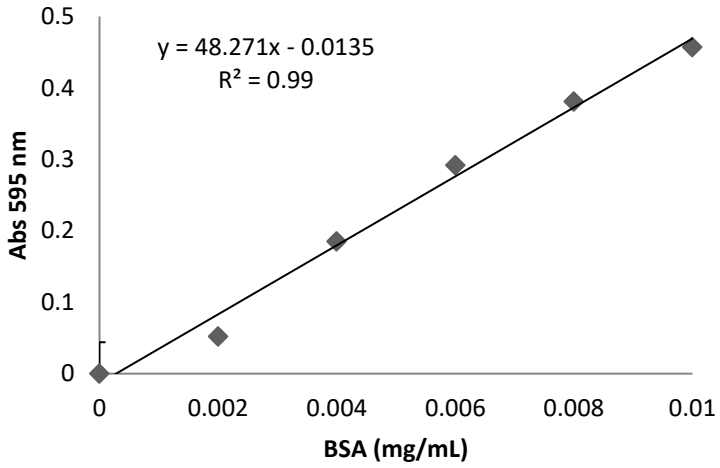


Figure 3.5. Calibration curve of protein using Bradford method and BSA as a standard protein, measured after 5 minutes of Coomassie reagent addition. Data were averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3).

3.7 WT-TK Quantification Using Different Disruption Methods

A common laboratory method to extract proteins from bacteria is sonication. This technique is based on the application of sound energy to the cells, by using ultrasonic pulses at a frequency of 20 kHz or higher. This will disrupt the cell membranes and release the intracellular content. Besides physical methods for cell disruption as sonication, alternative reagent-based methods also have certain advantages, such as not requiring expensive equipment. Moreover, they are compatible with high-throughput approaches and small volumes.

Bug Buster reagent (Sigma Aldrich) is formulated for the gentle disruption of the cell wall of *E. coli,* for the extraction of the soluble protein. This extraction method is very convenient because it is a simple, rapid, and low-cost alternative compared to other methods such as sonication or homogenisation. Formulation of the reagent utilizes a mixture of non- ionic detergents that can perforate the cell wall without denaturing the soluble protein. However, one of the main disadvantages is the high viscosity produced after cell disruption, mainly due to the release of DNA. Alternatively, freeze-thaw method is commonly used to lyse bacterial and mammalian cells. This causes formation of ice on the cell membrane which helps in breaking down the cell membrane. This method is time consuming and is not suitable for extraction of cellular components sensitive to temperatures.

WT-TK was harvest as described in Section 2.11 and extracted using three different techniques: Bug Buster reagent, freezing thawing, and sonication to determine the best method for protein extraction. Specific conditions for TK extraction are described in Section 2.12, 2.13 and 2.14. Bradford assay was used to measure total protein in lysates and samples after purification of TK with Ni-NTA columns (see Section 2.15). Protein was visualised using an SDS-PAGE gel with the conditions described in the Section 2.16. Table 3.1 describes the samples that are shown in figure 3.6 B.

Table 3.1. Samples TK-WT from the figure 3.7 (B).

Lanes	Samples
M	Marker
1	Lysate TK (Bug Buster)
2	Pure TK (Bug Buster)
3	Lysate TK (Freezing – Thawing)
4	Pure TK (Freezing – Thawing)
5	Lysate TK (Sonication)
6	Pure TK (Sonication)

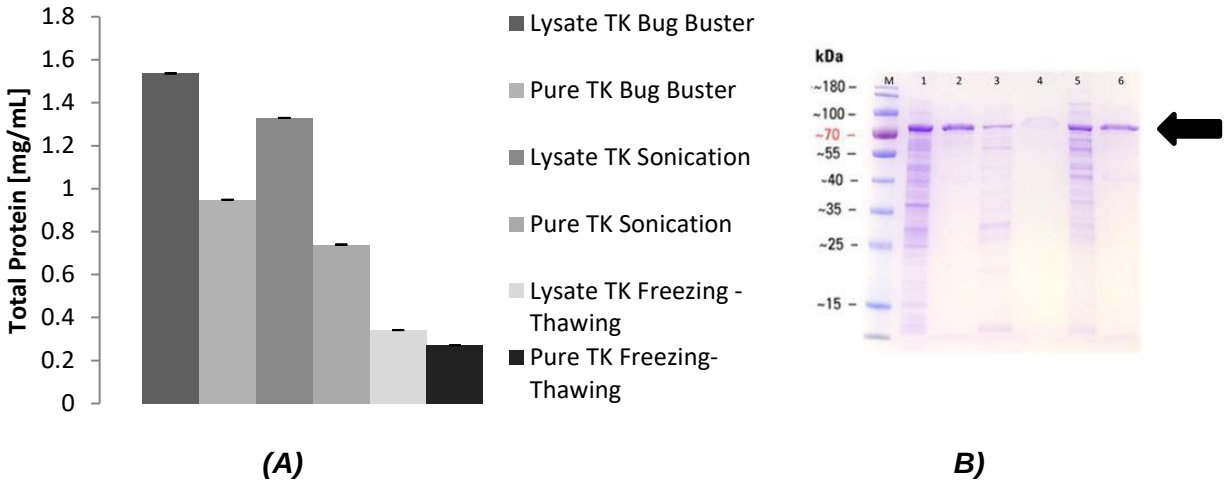


Figure 3.6. (A) Quantification with Bradford method of WT-TK (TK lysate) and purified TK using Ni-NTA columns for three different methods of extraction: Bug Buster, Sonication and Freezing – Thawing. Data were averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3). (B) shows SDS PAGE gel of transketolase (WT-TK) lysate and purified protein using different methods of TK extraction that are mentioned in Table 3.1. Arrow indicates the expected size of TK (~74 kDa).

Results in Figure 3.6 (A) shows a significant difference between TK lysates and purified TK samples, with a decrease of around 35% of protein concentration for purified TK samples for protein extraction with Bug Buster and by sonication. The same graph also shows that the use of Bug Buster achieves the highest extraction of TK protein. However, the use of this reagent increased the viscosity of the lysate, and in consequence the manipulation of these samples got very difficult to work with. Centrifugation and dilution of TK lysate could be an option; however, this extra step must be considered for each specific experiment or bioconversion.

On the other hand, the method of freezing-thawing using dry ice was not very efficient for protein extraction, with a yield of only 25 and 22% compared to the sonication and Bug Buster methods, respectively. Nevertheless, if it is not necessary to obtain a high concentration of protein, this could be a good low-cost alternative. Finally, sonication is the most common method for laboratory-scale protein extraction, it showed a high yield for protein extraction (1.32 mg/mL in TK lysate), and contrary to Bug Buster, it does not increase the viscosity of the samples. Figure 3.6 (B) shows the visualization of TK proteins using an SDS-PAGE. It shows a comparison of lysates of TK-WT with TK purified using Ni-NTA columns, for the different methods of protein extraction. TK extracted using Bug Buster and sonication showed similar protein content, and consequently the intensity of the bands in the gel were also similar. On the other hand, samples processed by freezing thawing gave the lowest protein content, and after purification the content of protein was ~ 0.2 mg/mL, which was not visible on the SDS gel. Finally, it also can be seen that the purification process using Ni-NTA columns was successful as it removes essentially all the unwanted proteins.

3.8 D469E and D469T Protein Quantification

In this project, two different reagents were used, *BugBuster Protein Extraction Reagent (Novagen),* and *B-PER Bacterial Protein Extraction Reagent (ThermoFisher)* for TK extraction from variants D469E and D469T (to compare efficiency and difference of both reagents) after 8 hours of fermentation at standard conditions. Both reagents enable mild extraction of the proteins from bacteria (*E. coli*), providing a simple, rapid, and low-cost alternative to mechanical methods such as French Press or sonication for releasing expressed target proteins in preparation for functional studies or purification, moreover sonication it is not suitable for HTP screening methods. The formulations are ready to use and contain a mixture of non-ionic detergents that are capable of cell wall perforation without denaturing soluble proteins. For this experiment, two different TK variants were analyzed following the instructions of the supplier and finally the protein content was measured using the Bradford assay (see Section 2.19). Table 3.2 shows that the B-PER reagent was two times more effective for extraction of total protein compared to the Bug Buster reagent. The results also correlate with the viscosity of the protein lysate that is considerably higher for the B-PER. Because it is very difficult to accurately pipet high viscosity samples, the use of B-PER was discarded for further experiments. D469E and D469T were chosen due to its similarity protein expression level, for SDS-PAGE gel (Figure 3.7), 0.03 mg of protein was loaded in each lane following the conditions described in section 2.16. These TK proteins were used as example to get a good point of reference to determine the efficiency of both reagents.

Table 3.2. Protein content of D469E and D469T using two different protein extraction reagents by triplicate.

Extraction reagent	D469E [mg protein/mL]	D469T [mg protein/mL]
BugBuster	3.5 ± 0.28	4.68 ± 0.25
B-PER	7.13 ± 0.56	8.89 ± 0.69

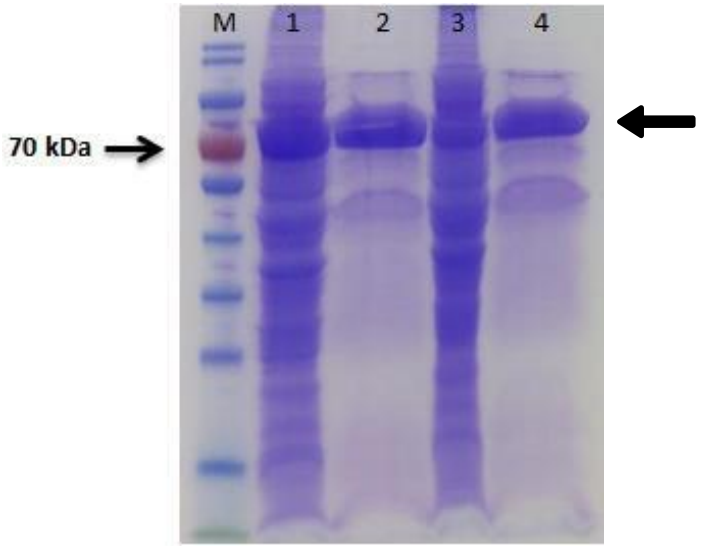


Figure 3.7. SDS-PAGE gel of TK lysate extracted with Bug Buster and pure TK samples from Ni-NTA columns. M: Marker, 1: lysate D469E, 2: pure TK D469E, 3: lysate D469T, 4: pure TK D469T. Left arrow indicated the reference of molecular weight and right arrow indicates the band that corresponds to TK (~74 kDa).

3.9 Magic Medium as an Alternative for High Biomass Production

Magic Media is an alternative media for high biomass production compared with LB and TB. Fermentations of *E. coli* expressing the variant D469T were performed in 250 mL baffled flasks with working volume of 50 mL at 37 °C, 250 rpm for 25 hours (Figure 3.8). The final biomass production using MM, as measured by OD_{600nm}, was six times higher than that obtained when TB or LB were used.

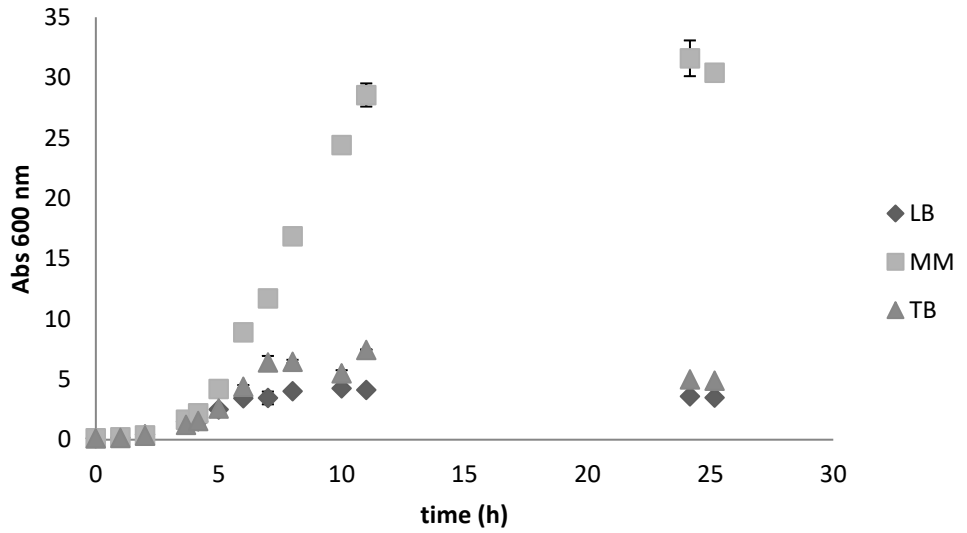


Figure 3.8. Fermentation curves of *E. coli* XL10Gold expressing TK variants D469T (50 mL) using three different growth media: Luria Bertani, Magic Media, and Terrific Broth in baffled flasks at 37 °C-500 rpm for 25 hours. Data points are averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3).

Magic Media was therefore found to be very useful for further small fermentations (~ 1mL) in 96 or 24 well plates, due to the high biomass production and TK expression. This medium may be particularly useful where variants with low enzyme expression should be screened, or when the protein needed to be purified in the high-throughput format for other biophysical measurements such as stability. Due to the surprisingly high OD reached with variant D469T using Magic Media, fermentations at standard conditions were also made in the same media for *E. coli* expressing TK-WT, D469E and D469T, giving the results shown in Figure 3.9. The three different bacterial strains had no significant difference in their growth profiles as measured by OD600nm over 24 hours of fermentation. At this point it could be concluded that Magic Media was the best medium for high biomass production, regardless of the protein variant.

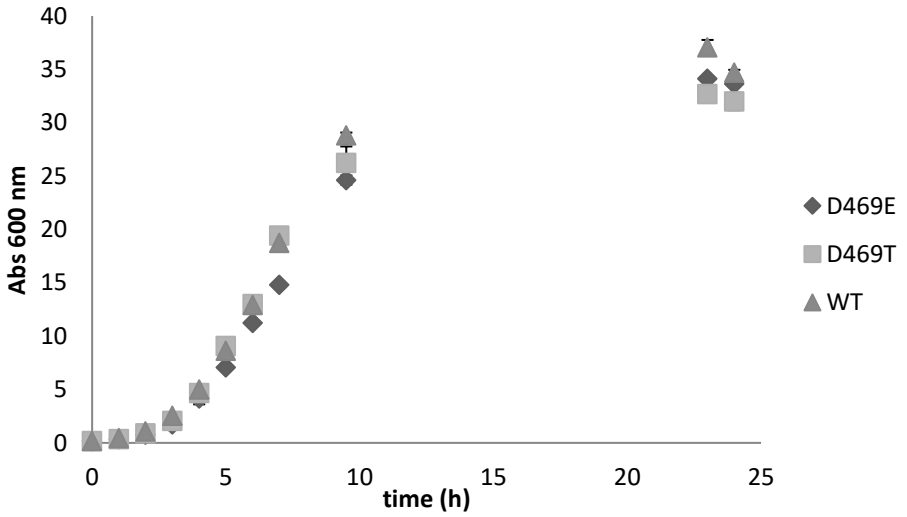


Figure 3.9. Fermentation curve of *E. coli* XL10Gold expressing TK variant D469T, D469E and WT-TK (50 mL) in Magic Media using baffled flasks at 37 °C and 500 rpm for 24 hours. Data points are averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3).

3.10 TK Production and Quantification in Different Media

Production of total protein is related to TK production, as for well-expressed variants this is typically 40% of the total protein, in Magic Media the total protein was approximately 6 times higher than in Terrific Broth or LB.

Table 3.3. Results of total protein from TK lysates.

Sample grown in Magic Media	Total protein [mg/mL]
WT	14.16
D469T	12.22
H192P/A282P/I365L/G506A	5.08
R35B D10	20.17
R35BD11	15.73
R520Y B12	12.79

To compare the TK production of variant D469T in different media, and recovery at each step, the total protein was measured using the Bradford assay for the samples. Total protein was quantified in the lysate and after each purification step (Figure 3.10). SDS-PAGE gels were made for each sample as it is described in section 2.16. As an example, a gel from the TB media is shown in Figure 3.10, using imidazole (250 mM) elution buffer for Ni-NTA columns (see Table 2.6).

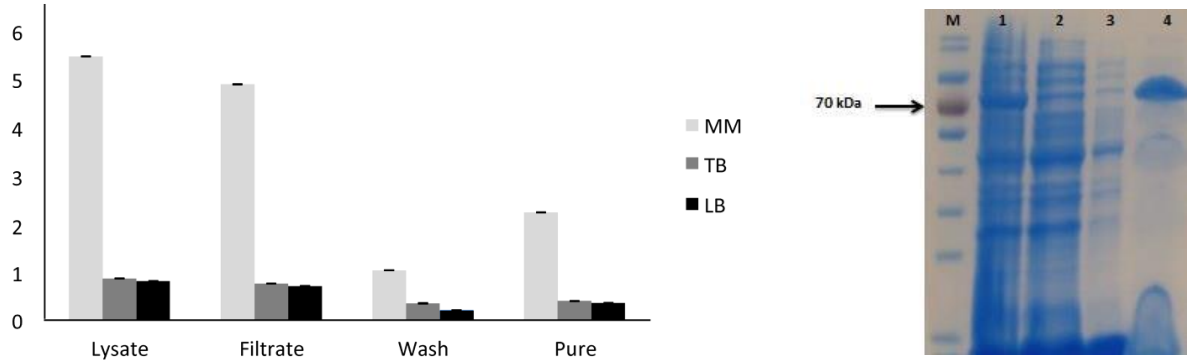


Figure 3.10. (A) Total protein quantification using Bradford assay after 24 hours of fermentation at standard conditions, using TK from D469T in three different growth media, LB, Magic Media, and Terrific Broth. Data averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3). (B) SDS-PAGE gel of protein samples from TB, M: Marker, 1: lysate D469T, 2: Filtrate, 3: Washing Step, 4: Purified TK.

Figure 3.10, does not show a significant difference in terms of TK production between TB and LB. However, the total concentration of protein in Magic Media was around six times higher, making this media not only a good alternative for high biomass production and for the corresponding production of protein. Moreover, as expected, total protein content decreased after each purification step, and achieved a final purified TK that was around 40% of the initial total protein concentration.

3.11 TK Fermentation in 96 and 24 Well Plates

Fermentations in 96 and 24 well plate were performed considering that after the generation of the mutagenic library, hundreds of colonies must be grown in 96 and/or 24 well plates for later protein extraction, and finally the screening for thermostable variants. In order to determine the conditions of fermentation, *E. coli* expressing TK variant D469T was grown in 96 and 24 well plates using three different growth media: Magic Media (MM), LB Miller Merck (LB), LB Miller Merck + 7.5% glycerol, Terrific Broth (TB) (Figures 3.11 and 3.12).

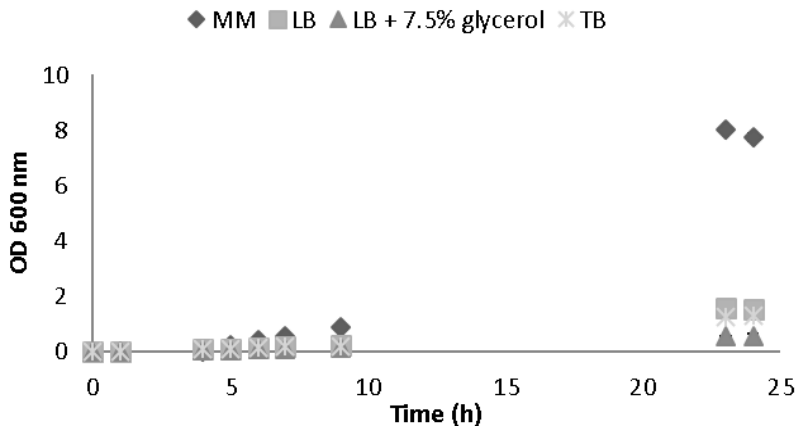


Figure 3.11. 96 well plate fermentation curve of *E. coli* XL10Gold expressing TK variant D469T in 1 mL of different growth media in 96 well plates at 37 °C and 500 rpm over 24 hours. Data were averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3).

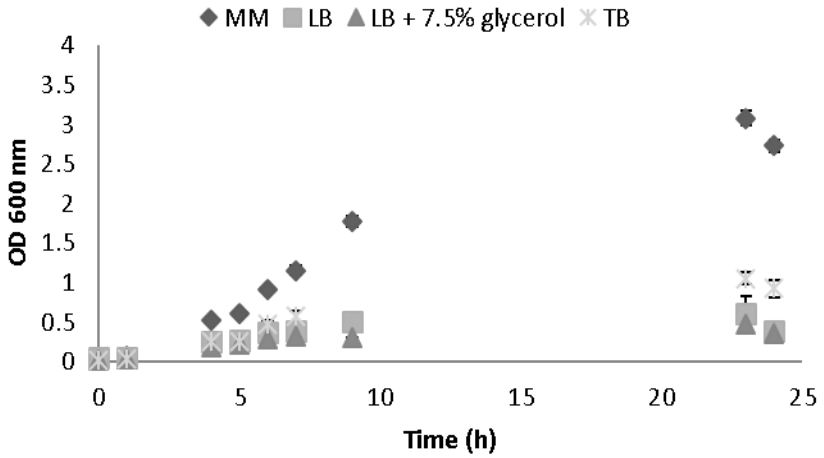


Figure 3.12. Fermentation curve of *E. coli* XL10Gold expressing TK variant D469T in 3 mL of different growth media in 24 well plates at 37 °C and 500 rpm over 24 hours. Data were averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3).

All fermentations were carried out at 37 °C for 24 hours. Statistical analysis (ANOVA) between growth media was made; the result showed that there is no significant difference between biomass concentrarion using different growth media (LB, LB + 7.5% glycerol, and TB) at the same growth conditions.

3.12 EGTA Buffer as an Alternative for TK Purification

One objective of the project was to develop a high-throughput purification in microplates that could be compatible with biophysical analysis, such as intrinsic fluorescence for stability measurements, using UV at 260-340 nm, without the need for further buffer exchange. The standard elution buffer contains imidazole which absorbs strongly in this range, and so it would interfere with measurements. An alternative to imidazole for the elution of TK (variant D469T) when purifying using Ni-NTA columns is ethylene glycol, tetra-acetic acid or EGTA, which has low absorbance in the UV-Visible range. This is a common buffer ingredient, due to its chelating activity with a high specificity for calcium ions, instead of magnesium ions that EDTA chelates.

Different concentrations of EGTA solutions were prepared (0, 5, 10, 15, 20, 25, 50, 75 and 100 mM) in buffer Tris-HCl 20 mM and NaCl 500 mM, pH 8.5. These different buffer concentrations were used as elution buffer for the final step of purification using the standard Ni-NTA protocol, results are shown in Figure 3.13. This result showed that 75 mM of EGTA was necessary to recover TK from the Ni-NTA columns.

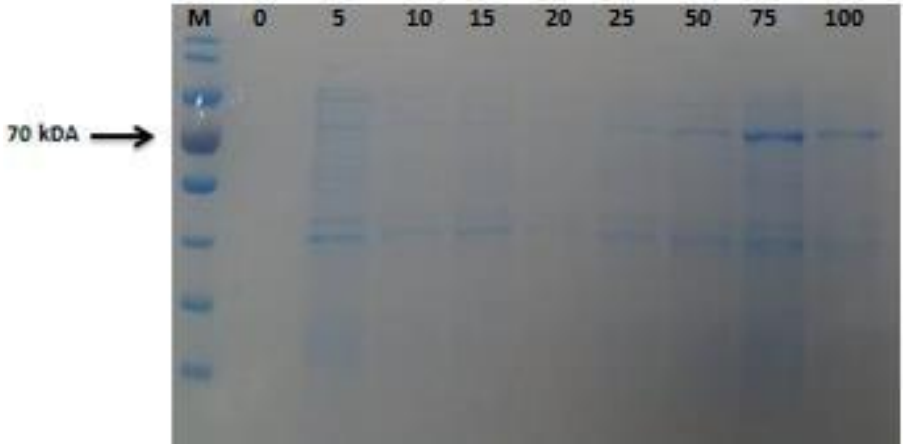


Figure 3.13. SDS-PAGE gel of TK samples eluted at different concentrations of EGTA: 5, 10, 15, 20, 25, 50, 75 and 100 mM in tris buffer HCl 20 mM and NaCl 500 mM, pH 8.5.

Moreover, higher concentrations of EGTA were tested to determine the best concentration for extraction of pure TK (see Figure 3.14). The buffers were prepared as described in Section 2.6 and the EGTA concentrations tested were 200, 300, 400 and 500 mM, pH 8.5. Figure 3.13 shows that above 75 mM of EGTA, the elution of TK was the same, and did not show a significant difference for buffers with between 75 mM and 500 mM of EGTA for TK extraction.

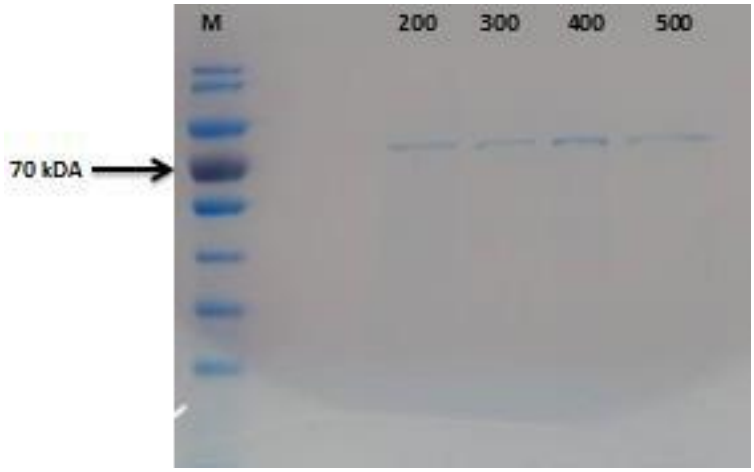


Figure 3.14. SDS-PAGE gel of TK samples eluted at different concentrations of EGTA: 200, 300, 400 and 500 mM in tris buffer HCl 20 mM and NaCl 500 mM, pH 8.5.

TK samples eluted with EGTA and imidazole buffers were compared on an SDS-PAGE gel (Figure 3.15) to determine the difference between protein content. Elution with buffer containing 250 mM of imidazole (see Table 2.6) was made using the same column after the first elution with 500 mM EGTA. The Bradford method could not be used as EGTA also reacts with the Coomassie reagent, giving an overestimated result.

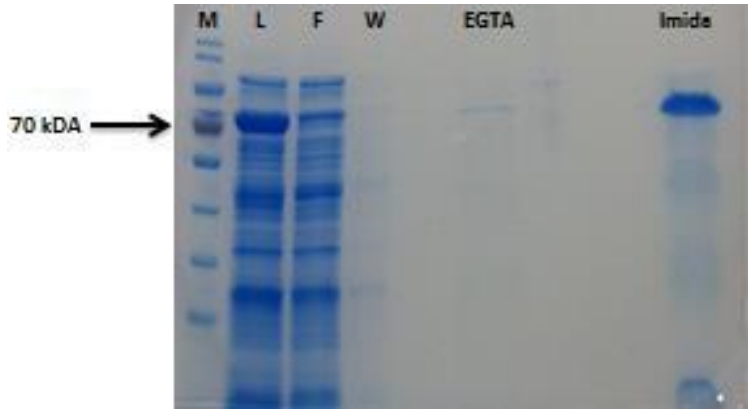


Figure 3.15. SDS-PAGE gel of TK samples: L: lysate, F: filtrate, W: washing, EGTA: EGTA buffer 500 mM, Imida: elution buffer 250 mM of imidazole.

As it can be seen in Figure 3.15, the content of TK after elution with EGTA was much lower than that eluted using the imidazole buffer. *Ultra 4 Centrifugal Filter Devices (Amicon)* can be used as an alternative to increase the protein concentration in purified samples. These filters have the capacity to concentrate protein samples seven times, in a relatively short time (10 minutes) due to a cellulose membrane in a

range of molecular weight cut-offs. However, this would not be useful in the development of a high-throughput assay.

3.13 Conclusions

This chapter provided basic but important conclusions that could be used to guide future experiments with transketolase. A key area with potential for the improvement of both biocatalytic screening and enzyme production, by impacting the production of biomass, was the choice of growth media. Different media were tested, and it was found that “Magic Media” gave the highest biomass production, reaching around six times higher biomass and total protein levels compared to the more commonly used Terrific Broth and LB. In this sense, we can say that Magic Media could be an effective alternative for specific experiments that require a high production of TK or other recombinant proteins. In the same context, it was proven that the addition of glycerol to LB media did not show a statistically significant difference in biomass concentration compared with LB without glycerol.

For example, the best harvest points of fermentation for TK-WT and different variants i.e., D469E, D469T, G425Z and R520Z in baffled flasks in 250 mL, were between 7 and 8 hours, where the bacteria had the maximum metabolic activity and therefore, the highest production of transketolase. Once the fermentation reached the ideal time for harvest, three TK extraction methods were used: Bug Buster reagent, sonication, and freeze-thawing. This comparison was performed because it was important to know quantitatively the difference between them in terms of protein extraction, as well as their advantages and disadvantages in the context of library screening. The chemical method using Bug Buster showed the highest protein extraction. However, the viscosity of the samples increased significantly, making the following steps difficult to work with. Sonication extracted around 86% of the protein compared to Bug Buster and the viscosity did not increase, so this method provided a solution that was easier to work with for further experiments like assays and biotransformations. Finally, the freeze-thawing technique extracted the least protein, with a yield of 22% compared to the chemical method. However, this method can be considered as the best alternative for extraction of a large number of samples in small volumes, for example in 96 well plates, because it does not require special equipment and is compatible with microtiter plates.

Another key technique that was evaluated for ensuring reproducibility and robustness, while attempting to minimise sample processing time, was protein quantification using the Bradford Method. It was concluded that the time required to incubate the samples with Coomassie reagent, prior to measurement of the protein samples at 595 nm, did not have a significant impact on the assay, and so the shortest

time of 5 minutes for reaction with the Coomassie reagent, was considered to be sufficient to get a reliable measurement.

One aim of this Chapter was to consider modifications to the existing high-throughput assays for TK that could enable the direct measurement of protein stability using biophysical assays such as intrinsic fluorescence. This would require high-throughput purification, and so would most likely use the existing approach of Ni-NTA capture of the His_6-tagged TK, but in microplates instead of usual purification column. A major issue with that approach is that the 250mM of imidazole used to elute the protein, strongly interferes with the quantitation of the protein by absorbance at 280 nm, intrinsic fluorescence, or by its activity. Therefore, the alternative eluant of EGTA was proposed, and used for the elution of purified TK. However, the total protein concentration extracted was only 10% compared with the standard elution buffer with imidazole. One possible advantage of the elution buffer with EGTA is that it is not necessary to take the extra step of dialysis prior to quantitative assays of the protein.

4 Colorimetric Method

4.1 Introduction to High-Throughput Screening Methods

Several methods are currently used to detect enzyme activity, including ultraviolet (UV) spectroscopy, high-performance liquid chromatography (HPLC), and colorimetric assays (Rice et al., 2013). Identification of TK variants with new characteristics such as thermostability, rate of reaction or acceptance of new substrates have been normally identified and quantified using HPLC. However, even when this method is very accurate, the use of this methodology is very laborious and time-consuming technique that limits the number of samples to analyze, and consequently delays the identification of new improved variants. Such limitation is solved with a well established HTP screening assay, able to identify variants with specific characteristics from a large number of samples.

Dye-based methods are the most common choice for viability assays and screening methods, due that it is possible to read by an automatic plate reader in a fast, simple and practical manner (and in some cases it is possible to identify by eye according with a color scale) samples with desirable characteristics. Another advantage of the HTP is that it requires, short time to identify and analyse large number of samples compared with HPLC and finally HTP methods do not need highly qualified training to execute compared with HPLC (Aguiar et al., 2017). Sensitivity, ease of use, and rapid data acquisition are also key characteristics that these methods must have for the identification of new variants (Smith et al., 2006). Moreover, it is crucial that the method could be able to screen hundreds, thousand and even millions of samples, and identify accurately, those that have desirable characteristics, in a relatively short time.

A colorimetric assay using 2,3,5-triphenyltetrazolium chloride (tetrazolium red) was tested to verify previous reported results and the efficiency of the method reported by Smith et al., (2006). Tetrazolium red is able to oxidize the α-hydroxyketone (HK) that is produced from the TK bioconversion between Li-HPA as a donor and an aliphatic aldehyde as the acceptor. Tetrazolium reaction (in a basic pH) involves the oxidation of the HK into the corresponding α-diketone and the reduction of the tetrazolium moiety into the corresponding formazane which displays an intense red color that can be measured at λ~450nm (Hecquet et al., 2014). However, this method is not reliable and neither accurate for screening measurements due to the poor stability of the formazane.

4.2 Colorimetric Method Using 2,3,5-Triphenyltetrazolium Chloride

In 2006 Smith, M. et al., reported a colorimetric assay for screening transketolase mutants by identifying (3S)-1,3-dihydroxypentan-2-one from the TK bioconversion between Li-HPA and PA using 2,3,5-triphenyltetrazolium chloride (TZC). The principle of the method is the oxidation of 2-hydroxyketone to the corresponding formazane that produces an intense red colour that can be measured in a spectrophotometer or microplate reader at 485 nm (see Section 1.15). In this project, Li-HPA was used as a control at different concentrations (1.5, 3, 6, 15, 50 and 100 mM) due that it also reduces the tetrazolium salt and generates the formazane with intense red colouration. However, the main disadvantage of this method is its poor stability of the colour over time. Figure 4.1 shows that for all samples, red formazane was very insoluble and precipitated within the first minute of the reaction, making the method not suitable to screen samples and neither convenient to use instead of HPLC, contrary to the conclusions reported by Smith et al., (2006).

(A)

(B)

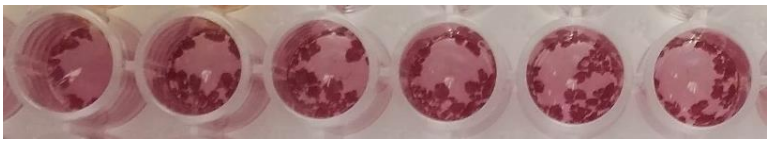


Figure 4.1. Colorimetric assay between Li-HPA and TZC 0.2%. (A) Shows the colouration immediately after addition of NaOH. (B) Shows the same reactions after 1 minute.

In order to analytically determine the effect of poor stability when using TZC, absorbance of the reactions were followed at different Li-HPA concentrations (50 µL) and mixed with TZC solution at 0.2% (20 µL) and NaOH 3M (10 µL). Samples were measured at λ = 450 nm, every 22 seconds for ten minutes in the plate reader (Figure 4.2).

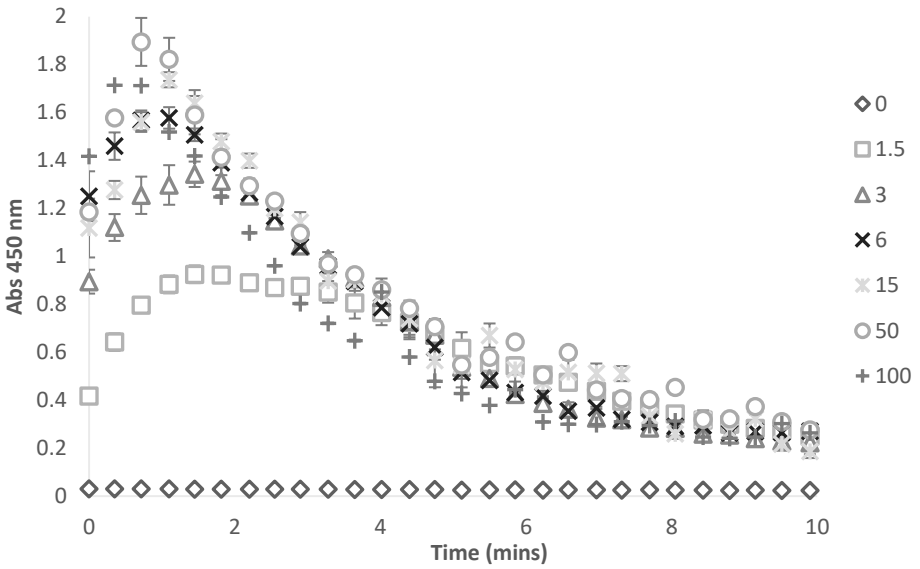


Figure 4.2. Reaction between different concentrations of Li-HPA between 0 and 100 mM, TZC 0.2% and NaOH 3M measured at 450 nm every 22 seconds for 10 minutes at room temperature. Data points are the averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3).

As it can be seen in Figure 4.2, the formazan produced is very unstable and the results depend on the moment when the sample is measured. The method also does not have a clear correlation between absorbance and the concentration of Li-HPA. In the first minute of reaction, the samples reached the maximum peak or absorbance value. However, the absorbance was very similar at higher concentrations, and hence it was not possible to distinguish these different concentrations through colour intensity. The absorbance then decreased rapidly after one minute, resulting in the same absorbance values for all the samples after 3 minutes. Considering this analysis, the use of TZC was completely discarded for further experiments, and instead the use of WST-1 was set as the best option due to the formation of stable formazans and no formation of precipitate compared with TZC.

4.3 Conditions for the Colorimetric Method Using WST-1

Colorimetric reactions were tested using a Tris buffer control, Erythrulose (Ery), Li-HPA, glycolaldehyde (GA), and propionaldehyde (PA), in order to determine any differences in terms of their impact on colour change and intensity, towards the overall effect of the substrates and products involved in the bioconversions. Different controls were made using the solutions and volumes described in Table 4.1. Reactions were carried out at room temperature with a constant volume of WST-1 and NaOH (standard conditions) for all the samples. For these reactions, the NaOH was added at the end, as this can affect the results if the reaction is not initiated immediately.

Table 4.1. Solutions and volumes used for colorimetric reaction with tetrazolium WST-1.

Solution	Abbreviation	µL
MQ H_2O	H_2O	50
Tris buffer HCl 50 mM, pH 7.0	Buffer	50
Li-HPA 200 mM in tris buffer HCl 50 mM, pH 7.0	Li-HPA	50
100 mM Glycolaldehyde in tris buffer HCl 50 mM	GA	50
100 mM Propionaldehyde in tris buffer 50 mM (100)	PA	50
100 mM L-Erythrulose in tris buffer 50 mM (100)	L-Ery	50
WST-1 (0.2% in MQ H_2O)	WST-1	20
NaOH (3M)	NaOH	10

Table 4.2. Controls of reaction to determine the colouration after mixing different solutions with WST-1 (0.2%) and NaOH (3M) at room temperature.

H_2O	Buffer	Li-HPA	GA	PA	L-ery	Initial colour	After 5 minutes	Conclusion
YES	-	-	-	-	-	Transparent yellow	Light blue	No reaction WST-1
-	**YES**	-	-	-	-	Transparent yellow	Light blue	No reaction WST-1
-	-	**YES**	-	-	-	Dark blue	**Orange**	**Reaction WST-1**
-	-	-	**YES**	-	-	Dark blue	**Orange**	**Reaction WST-1**
-	-	-	-	**YES**	-	Transparent yellow	Light blue	No reaction WST-1
-	-	-	-	-	**YES**	Dark blue	**Orange**	**Reaction WST-1**

Table 4.2 shows that Li-HPA, glycolaldehyde and L-erythrulose reacted with WST-1, forming a formazan complex with an orange colouration. Water, buffer, and PA did not result in the final reaction colour. For the samples that reacted with WST-1, after addition of NaOH, the solution colouration turned to dark blue and continued changing to purple, dark red, dark orange and finally orange after five minutes of reaction at room temperature. An advantage of WST-1 compared with tetrazolium red is that WST-1 does not form a precipitate. The orange colouration formed was very stable, although after 1 hour at room temperature some changes in terms of colour intensity could be detected using a spectrophotometer. This characteristic is fundamental to generate not only a reliable but also a fast and quantitative method that can be utilised for screening of new variants using directed evolution approaches.

4.4 MP Resin and Absorption Capacity

As the 2-hydroxyketone moiety expected in the product of TK bioconversion of PA and Li-HPA is also present in the Li-HPA substrate, tetrazolium salts react with it to give false positive background results. MP resin is suitable to absorb the residual Li-HPA after the bioconversion for avoiding this (Smith et al., 2006). The resin used for these experiments was obtained from *Biotage*, who report a maximum capacity of Li-HPA absorption of 3.12 mmol/g resin. Nevertheless, the maximum binding capacity was determined experimentally using a constant amount of resin (100 mg) and different Li-HPA concentrations as described in Table 4.3. MP carbonate resin was equilibrated with 200 µL to avoid liquid absorption and consequently loss its absorption capacity. Determination of the maximum absorption of Li-HPA was based also on a colorimetric method, considering that the presence of Li-HPA will react with tetrazolium WST-1 producing an orange colouration, while a blue colour indicates the absence of Li-HPA.

Table 4.3. Conditions for determination of maximum binding capacity of MP-Carbonate resin for Li-HPA at room temperature after 1 hour.

Sample	H2O (µL)	Li-HPA [200 mM] (µL)	Mili moles Li- HPA/g MP carbonate resin	Colour
1	400	0	0	Dark blue
2	350	50	0.1	Dark blue
3	300	100	0.2	Dark blue
4	250	150	0.3	Dark blue
5	200	200	0.4	Dark blue
6	150	250	0.5	Dark blue
7	100	300	0.6	Dark blue
8	**50**	**350**	**0.7**	**Red purple**
9	**0**	**400**	**0.8**	**Red purple**

According to Table 4.3, sample 8 changes its colour from blue to red and purple indicating presence of Li-HPA. Contrary to what was reported from *Biotage*, the maximum binding capacity of the resin was much lower under experimental conditions, achieving just 0.7 mmol/g resin. In order to absorb the remaining Li-HPA after the bioconversion and before the colorimetric reaction with tetrazolium WST-1 it is important to consider this result to make sure that the formation of a formazane and consequently the orange colour was only due to the presence of (3S)-1,3-dihydroxypentan-2-one and not the remaining Li-HPA.

4.5 Calibration Curves of Li-HPA Using the Colorimetric Method

Knowing that Li-HPA, GA, and L-ery are able to react with WST-1, calibration curves were made in order to get results that could be used as a control. After addition of WST-1 and NaOH [3M] into the plate, the reagent was kept in the 96 well-plate and incubated for 1 hour at room temperature (standard incubation time) taking measurements at 450 nm and 600 nm every 10 minutes. Three replicates were used at each concentration, using tris HCl buffer [50 mM] as a control. All data are expressed as the mean ± SD that corresponds to the standard deviation between measurements of three samples for the same analysis. The appropriate incubation time after addition of all reagents depends on the individual experiment setup. Moreover, absorbance measurements should be taken at different times, this allows to determine the optimal incubation period for the experiment setup used. It is also recommended to use a blank control in the same experiment. Measurements from the control allows to use this background absorbance as a blank data for the samples (Roche, 2007).

Calibration curves were divided into two different categories, from 0 to 1.2 mM (low concentration) and from 10 mM and above (high concentrations) as these ranges resulted in different spectroscopic properties. For low concentrations of product, the formazane generated gave a blue colour that could also be quantified at 600 nm. From 1.2 to 10 mM of product, the colour varied from dark blue to orange, and finally over 10 mM the colour remained orange but with different tonalities that could be measured and quantified by the spectrophotometer at 450 nm. Li-HPA was used as a model to study the effect of time on the colour formed at room temperature.

Figure 4.3 shows that low concentrations of Li-HPA (< 1.2 mM) have a linear correlation with the absorbance at 600 nm, and the known concentration of the product. However, the method required will focus on screening of samples with higher product concentrations > 10 mM. Therefore, concentrations between 10 and 50 mM of Li-HPA were monitored for 60 minutes at room temperature, taking measurements every 10 minutes (Figure 4.4). The blank contains cofactors buffer, NaOH (3M) and WST-1 and does not go the origin due that these ‘impurities’ can absorb light in the selected wavelength that produces an absorbance value.

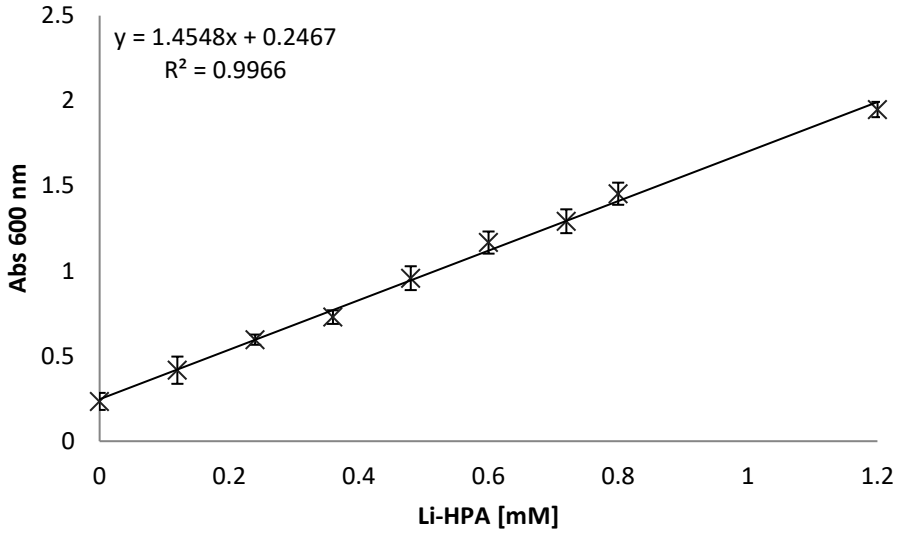


Figure 4.3. Calibration curve of Li-HPA reaction with WST-1 (0.2%) and NaOH 3M at 600 nm. Error bars represent one standard deviation of the mean (n=3). R^2 refers to the strength of the degree of correlation between the *y* and *x* values, the closer is to 1, the stronger correlation.

According to Figure 4.4, the absorbance and the intensity of the orange colour changes over time, and after 60 minutes at room temperature it is possible to get a direct correlation between the absorbance at 450 nm and the known concentration of the Li-HPA. One advantage of this method is that it can be used qualitatively to distinguish between samples with low product concentration <5 mM with a blue colour, and those with concentrations between 10 and 50 mM having an orange colouration. It could even be used quantitative by measuring the samples that present orange colour after 1 hour at 450 nm.

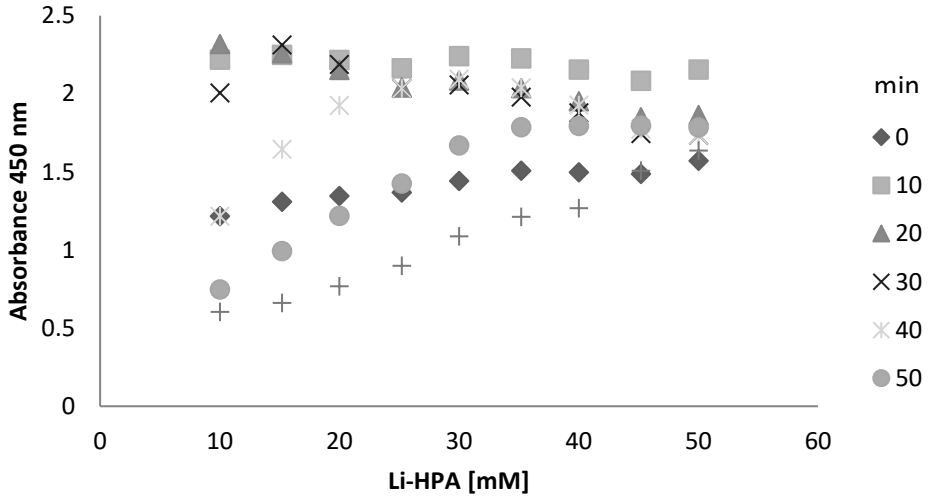


Figure 4.4. Calibration curve of different concentrations of Li-HPA (mM) after reaction with WST-1 (0.2%) and NaOH 3M measured at 450 nm. Measurements were taken at room temperature every 10 minutes for 60 minutes. The last measurements showed a linear correlation between concentration of Li-HPA and absorbance at 450 nm in the range of 10 to 50 mM of Li-HPA.

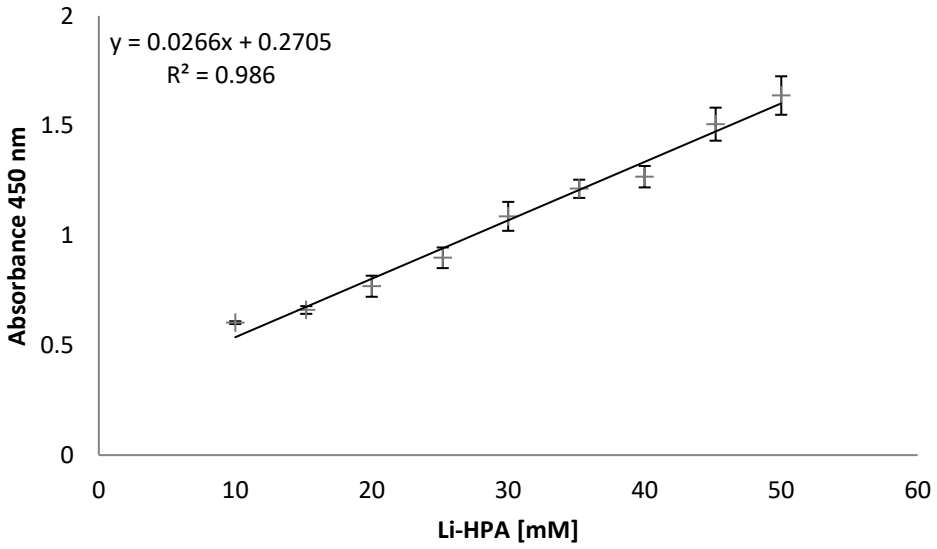


Figure 4.5. Calibration curve of Li-HPA after reaction with WST-1 (0.2%) and NaOH 3M incubated at room temperature for 60 minutes at 450 nm. Data points are averages from triplicate experiments and error bars represent one standard deviation of the mean (n=3).

In conclusion, it was possible to get a quantitative method instead of only a qualitative method, over a specific range of product concentrations. Sixty minutes of reaction at room temperature was selected and plotted in Figure 4.5, where the correlation between absorbance and product concentration is shown, with R^2 of 0.986.

4.6 Calibration Curve of (3S)-1,3-Dihydroxypentan-2-one (HK) Using the Colorimetric Method

Previous calibration curves were made with Li-HPA as a model substrate. However, the most important standard curve to use is the one obtained from the product (3S)-1,3-dihydroxypentan-2-one (HK) that will identify the most active variants in bioconversions that use PA and Li-HPA as the substrates. From previous analysis (Figure 4.4) it was seen that the absorbance changed over time, and that after 60 minutes at room temperature it was possible to get a linear correlation between the absorbance and Li-HPA concentrations between 10 and 50 mM. To analyse the product, calibration curves were made, using pure 1,3- dihydroxypentan-2-one (HK) synthesized in the UCL Chemistry Department with the method reported by Cazares et al., (2010). However, since these results were obtained using pure HK, it is important to determine if these curves have some differences compared to the HK that it is produced from the TK bioconversion.

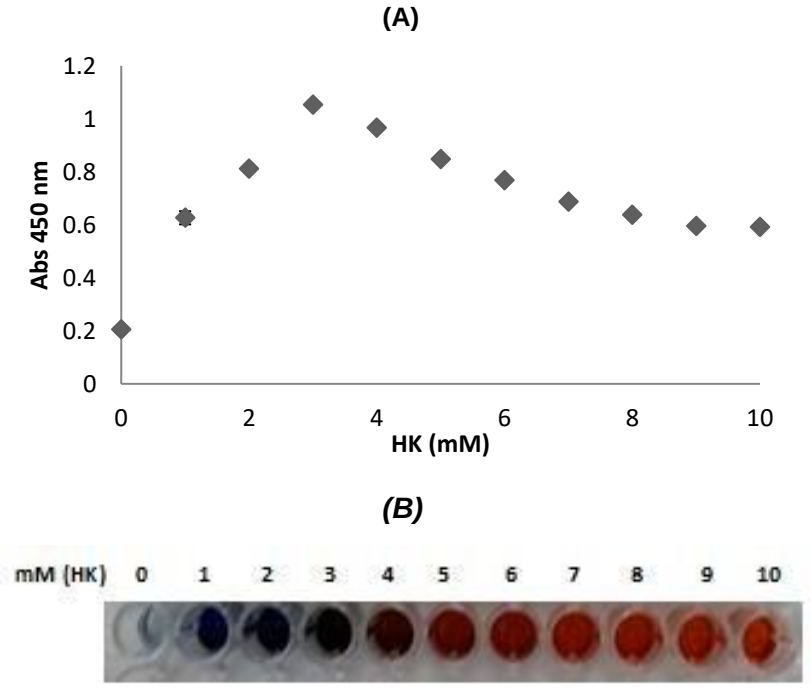


Figure 4.6. (A) Calibration curve of (3S)-1,3-dihydroxypentan-2-one (HK) at low concentrations (1 – 10 mM) after reaction with WST-1 (0.2%) and NaOH 3M incubated at room temperature for 10 minutes and measured at 450 nm. (B) Represents the colouration of samples. Error bars represent one standard deviation of the mean (n=3). Most error bars were smaller than the symbol and so cannot be seen.

The calibration curve with low concentrations of HK (0 – 10 mM) is shown in Figure 4.6, this curve can be used for HK quantification. However the standard conditions for bioconversion gives higher HK concentrations and for that reason this curve was not used for further experiments in this project. The results from HK were very similar as the samples tested with Li-HPA where concentrations of 8 mM or higher gave an orange colouration. High concentrations of HK were also tested and measured at 450 nm after 60 minutes of reaction.

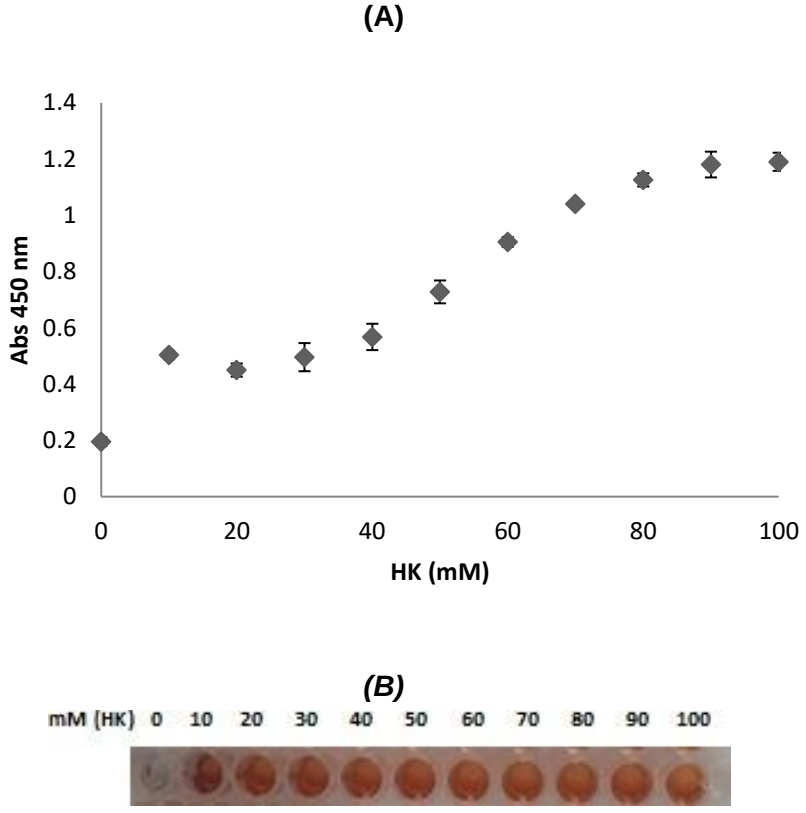


Figure 4.7. (A) Calibration curve of (3S)-1,3-dihydroxypentan-2-one (HK) at high concentrations (10 – 100 mM) after reaction with WST-1 (0.2%) and NaOH 3M incubated at room temperature for 60 minutes and measured at 450 nm. (B) Physical samples. Error bars represent one standard deviation of the mean (n=3).

As can be seen in Figure 4.7, formazane is very stable and does not a precipitate, in addition measurements show a linear correlation between 20 and 70 mM of HK. This gives the possibility to use this method for quantification of HK from TK bioconversion. The linear correlation using HK is shown in Figure 4.8. The next step was to compare these results with standard quantification method using HPLC.

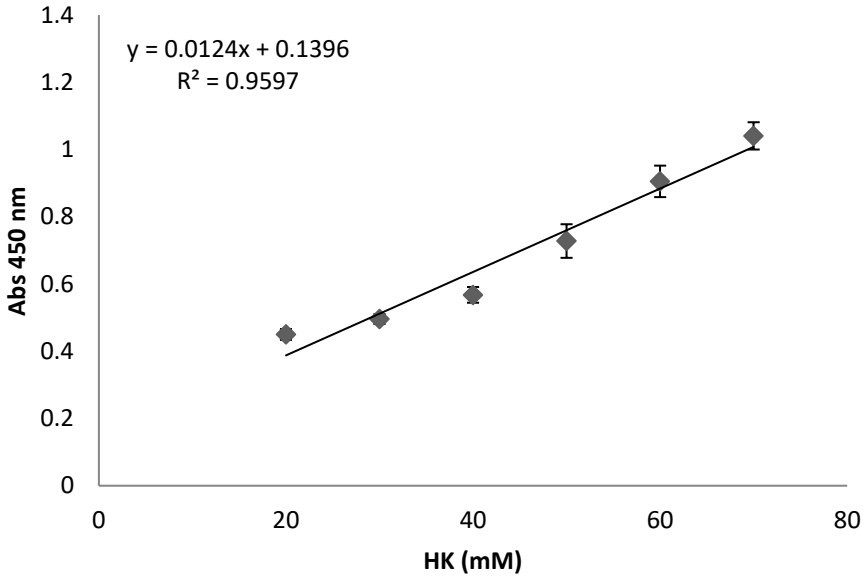


Figure 4.8. Calibration curve of (3S)-1,3-dihydroxypentan-2-one (HK) at high concentrations (20 – 70 mM) after reaction with WST-1 (0.2%) and NaOH 3M incubated at room temperature for 60 minutes. Error bars represent one standard deviation of the mean (n=3).

4.7 Calibration Curve of Li-HPA and (3S)-1,3Dihydroxypentan-2-one (HK) in HPLC

Low concentrations of HK (0 to 10 mM) were measured by HPLC (Dionex, CA, USA) as described in Section 2.21. Figure 4.9 shows correlation of the area (mAU*min) of the peak at 15.5 minutes with the concentration of pure (3S)-1,3-dihydroxypentan-2-one.

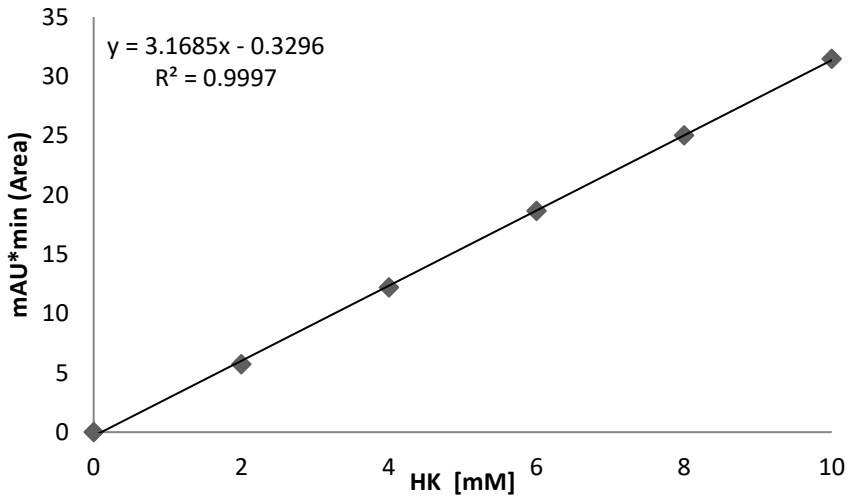


Figure 4.9. Calibration curve of (3S)-1,3-dihydroxypentan-2-one (HK), 0 – 10 mM in HPLC using UV detection at 210 nm, retention time of the product is 15.25 minutes.

Figure 4.9 shows a high correlation (R^2 = 0.9997) for HK samples between 0 and 10 mM. A new calibration curve using higher concentrations from 0 to 50 mM of HK was then made (Figure 4.10).

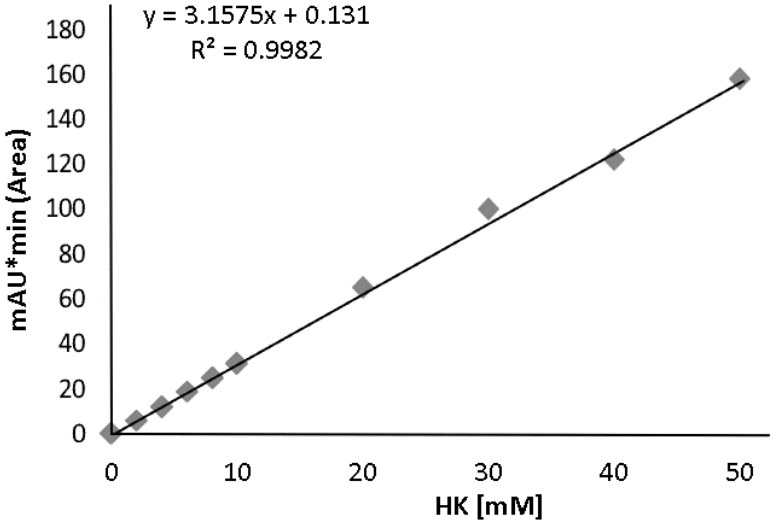


Figure 4.10. Calibration curve of (3S)-1,3-dihydroxypentan-2-one (HK), 0 – 50 mM in HPLC using UV detection at 210 nm, retention time of product 15.25 minutes. R^2 refers to the strength of the degree of correlation between the y and x values, the closer is to 1, the stronger correlation.

The calibration curve of Li-HPA was also obtained (Figure 4.11) to have analytical measurements of the substrate consumed during the bioconversion.

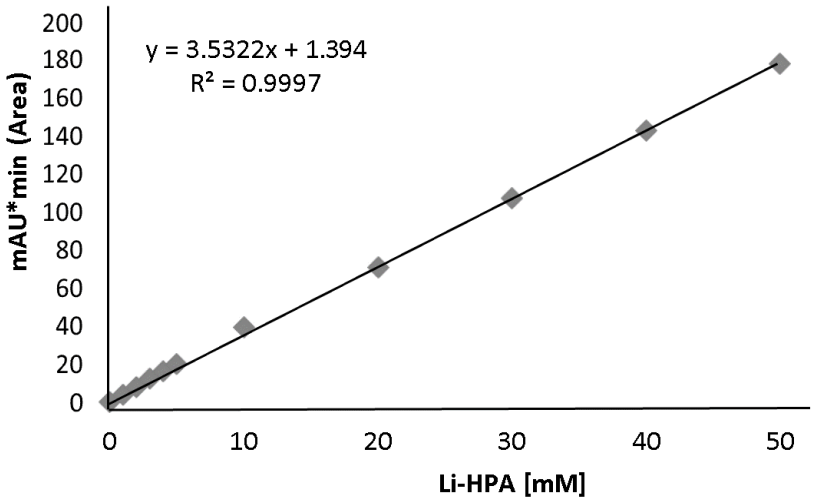


Figure 4.11. Calibration curve of Li-HPA, 0 – 50 mM in HPLC using UV detection at 210 nm, retention time of the product is 8.26 minutes. R^2 refers to the strength of the degree of correlation between the y and x values, the closer is to 1, the stronger correlation.

4.8 Application of Colorimetric Method in TK Bioconversion

After having determined the conditions for the colorimetric method, the resulting protocol was tested in reactions with TK lysates from the variant D469T, using Li-HPA and PA as substrates. Bioconversion was performed over 48 hours using the substrate concentrations described in Table 4.4.

Table 4.4. Solutions and volumes used for TK bioconversion.

	[Initial]	Volume [µL]	[Final]
TK Lysate	1 mg/mL	200	0.3 mg/mL
Li HPA	200 mM	200	60 mM
PA	200 mM	200	60 mM
Cofactor solution	12 x	65	1.2 x

For the colorimetric assay, 100 µL of reaction product was transferred into an Eppendorf tube containing 100 mg of MP-carbonate resin *(Biotage)* previously equilibrated with 200 µL of 50 mM buffer Tris-HCl, pH 7.0, for 30 minutes. The addition of this resin proved successful as a scavenger to remove Li-HPA before the addition of tetrazolium solution. Once that MP resin had absorbed the remaining Li-HPA, 50 µL of the supernatants for each bioconversion was mixed with 20 µL of WST-1 (0.2%) and 10 µL of NaOH (3M), and mixed by aspiration with a pipette. After 10 minutes of reaction (when the colour was relatively constant) a picture was taken and is shown in Figure 4.12.

1 2 3 4

Figure 4.12. Colorimetric reactions of 1: Li-HPA 200 mM, 2: Product of bioconversion after MP resin, 3: mix of Li-HPA, cofactors and PA after MP resin treatment, 4: sample 3 with no MP resin treatment.

Figure 4.12 shows the application of the colorimetric method qualitatively. The experiment had three different controls. Sample 1 was the first control, containing only Li-HPA that produced an orange colour. Sample 3 contained substrates (Li-HPA and PA) and cofactors without the TK lysate but with MP resin treatment. Sample 4 was mixed during 30 minutes with MP-resin to absorb all Li-HPA giving a blue colouration, this indicated that the amount of MP resin was enough to absorb all Li-HPA. Sample 4 corresponded to sample 3 but with no treatment with MP-carbonate, and finally Sample 2 corresponded to the bioconversion product (3S)-1,3-dihydroxypentan-2-one (HK) after MP resin treatment. Using these controls, it was possible to confirm that the orange colouration in Sample 2 corresponded to HK with concentration ≥ 8 mM.

Using the method described in this section, it is possible to screen large quantities of TK variants per day in a considerably shorter time (around 5,000 samples in three hours) compared with HPLC that needs 20 minutes to measure a single sample. In addition, results from HPLC and spectrophotometry (see Sections 4.10) suggested that it is possible to make a linear correlation between absorbance at 450 nm and the concentration of HK [mM]. Another advantage of the HTP method developed in this section is that depending on the objective of the experiment, the colorimetric method can be used either qualitatively to identify the most active variants, or quantitatively for just identifying active TK mutants by using the correlation between colour and product concentration.

4.9 Is MP Resin Necessary in the Colorimetric Method?

Considering that the MP resin addition adds another step in the process for the colorimetric method, which as a consequence can introduce variability, cost and time in the screening of several thousand of samples, an experiment was designed to determine whether it was possible to remove this step while keeping reliable results. Bioconversions using TK (in pure and lysate forms) from WT were performed using the conditions shown in Table 4.5. The objective of this experiment was to quantify the Li-HPA after 72 hours of bioconversion at room temperature and 800 rpm, in Eppendorf tubes of 2 mL. The reaction was quenched with TFA (0.1%) centrifuged for 1 hour at 15 000 rpm, 4 °C, and analysed by HPLC. The results are shown in Table 4.6.

Table 4.5. Solutions and volumes used for TK bioconversion using pure and lysate forms.

Reaction	Catalyst, Reactive or Substrate	Initial Concentration	Volume [µL]	Final Concentration
1	TK pure	0.74 mg/mL	800	0.22 mg/mL
2	TK lysate (total protein)	1 mg/mL	800	0.33 mg/mL
3	Control no TK H_2O	-	800	-
	Li HPA	200 mM	800	60 mM
	PA	200 mM	800	60 mM
	Cofactor solution	12 x	280	1x

Table 4.6. Results of bioconversion after 72 hours.

Bioconversion sample	Remaining Li-HPA (mM)
Pure TK	2.5 ± 1.02
Lysate	2.19 ± 0.78
Control no TK	28.71 ± 0.18

The results from Table 4.6 show that after 72 hours of bioconversion, more than

95% of Li-HPA was consumed. Considering that less than 5 mM of Li-HPA does not show an orange colouration, using the colorimetric method it can be concluded that using these reaction conditions, it is not necessary to use the MP resin in the colorimetric method. This is an advantage in terms of time and manipulation of the samples. However, it is necessary to have some previous control, and try the same experiment using different measurement times in order to know the exact time where the TK converts more than 95% of the Li-HPA; and in case that Li-HPA remains in concentrations >10 mM it would be necessary to use of MP resin for achieving Li-HPA complete absorption.

4.10 Relationship Between Colorimetric Method and HPLC

Two different TK lysates were used to establish the correlation between the colorimetric and HPLC methods, the first one using WT and the other using the quadruple variant H192P/A282P/I365L/G506A. The experiments were performed in triplicate for 72 hours of bioconversion at room temperature and 800 rpm (see Table 4.7). The quadruple variant was selected because it has been reported as the most stable mutant until now. It retains 66.2% activity after incubation at 60 °C for 1 hour, that represents a 10.2-fold improvement over WT and it has the highest half-life of 82.5 minutes, representing a 21-fold improvement over that of WT (4 min) (Yu & Dalby, 2018).

Table 4.7. Solutions and volumes used for TK bioconversion.

Reactive	[Initial]	Volume [µL]	[Final]
TK Lysate	1 mg/mL	800	0.3 mg/mL
Li-HPA	200 mM	800	60 mM
PA	200 mM	800	60 mM
Cofactor solution	12 x	260	1.2 x

After 72 hours of bioconversion, the same samples were measured by HPLC and with the colorimetric method using a plate reader at 450 nm and after 60 minutes of reaction with WST-1, as described in Sections 4.5 and 4.6. This provided the correlations between the results from HPLC and the colorimetric method (CM) presented in Figures 4.13 and 4.14. Considering the results from Section 4.9 it was not necessary to use the MP for the CM.

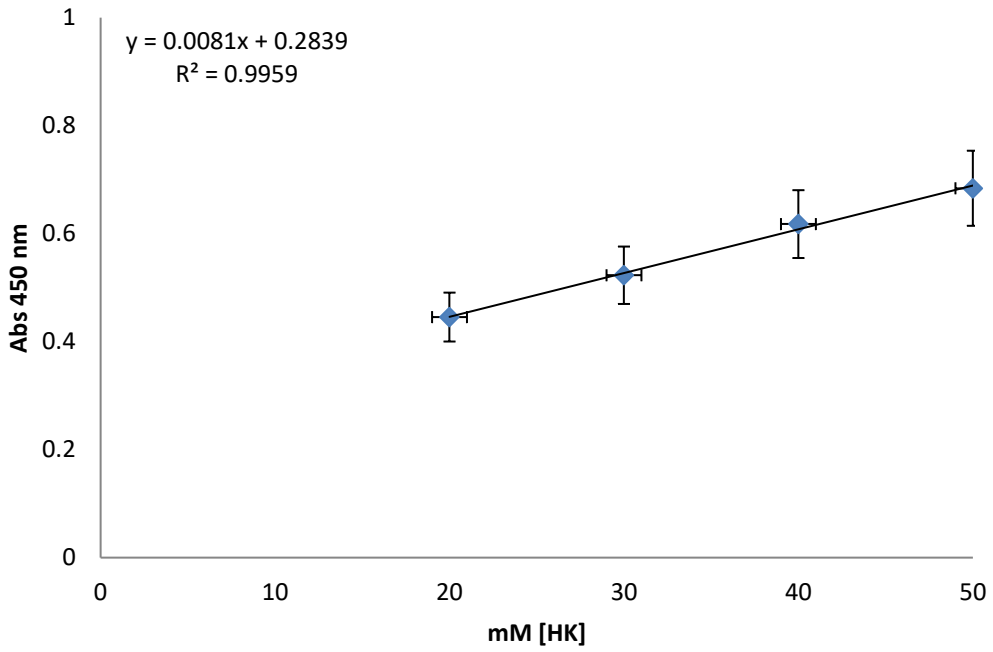


Figure 4.13. Correlation curve between HPLC (x axis) results and absorbance (y axis) after 60 minutes of reaction using the colorimetric method for WT-TK lysate. Dilutions were made with TFA 0.1%, retention time of product 15.25 minutes. Error bars represent one standard deviation of the mean (n=3). Correlation of colorimetric method and HPLC is only valid in the range of hydroxyketone shown in the graph. R^2 refers to the strength of the degree of correlation between the y and x values, the closer is to 1, the stronger correlation.

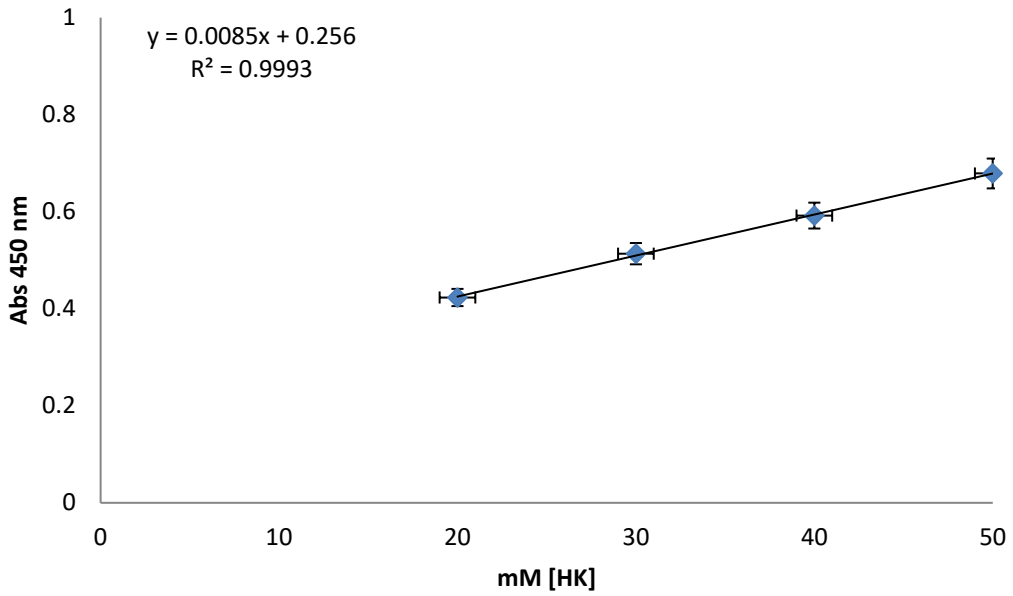


Figure 4.14. Correlation curve between HPLC (x axis) results and absorbance (y axis) after 60 minutes of reaction using the colorimetric for quadruple variant (H192P-A282P-I365L-G506A). Dilutions were made with TFA 0.1%, retention time of product 15.25 minutes. Error bars represent one standard deviation of the mean (n=3). Correlation of colorimetric method and HPLC is only valid in the range of hydroxyketone shown in the graph. R^2 refers to the strength of the degree of correlation between the y and x values, the closer is to 1, the stronger correlation.

As can be seen in Figures 4.13 and 4.14, both variants show an R^2 above 0.99 indicating a a linear correlation between HPLC and absorbance (for the colorimetric method) measurements in the range of product between 20 to 50 mM of HK. The new colorimetric method could also therefore be used as a quantitative method for following the reaction under different reaction conditions as time, and temperature, giving a reliable quantification of the product (3S)-1,3-dihydroxypentan-2-one. The new method was considerably simpler than the previous colorimetric approach, it is quantitative, and it is reduced in time and cost, which represents a major advantage compared with HPLC measurements. Depending on the laboratory automation, it is possible to analyse more than 5,000 TK variants per day.

The colorimetric assay using WST-1 has demonstrated a high sensitivity for two different ranges of product concentrations, the first one from 0 to <2 mM of HK, making measurements at 600 nm that corresponds to a blue colouration and from 20 to 50 mM of HK at 450 nm, corresponding to an orangcoloe-yellow formazan coloration within 60 minutes of reaction. Standard solutions of HK were previously measured with

both methods (see sections 4.6 and 4.7). Once that was calibrated and stablised, the results of the colorimetric method (absorbance at 450 nm) were compared with HPLC measurements, showing a good linear correlation between both methods. This results shown that absorcance of the colorimetric method can be used as a reference to get the concentration of HK (between 20 to 50 mM) from TK bioconversion. As conversion yields can be directly estimated based on the colour intensity of the formazan, this assay offers a simple, rapid, and sensitive screening approach to use in a high-throughput platform for the evaluation of large enzyme libraries. Moreover, this simple screening method has enormous potential to discover new TK variants because it does not require expensive equipment, saves time and it is suitable to identify and quantify the activity of variants without the use of analytical instruments such as HPLC (Baud et al., 2015). Finally, this screening technique provides the possibility to perform a first selection for a large number of variants in short times. These results set the basis to apply this method for the selection of TK variants from mutagenic libraries.

4.11 Conclusions of Colorimetric Method

The method proposed on this chapter using another tetrazolium salt such as WST-1 has shown promising results, because it allows to identify active TK variants qualitatively and quantify based on the product of bioconversion (HK) using water soluble tetrazolium salt (WSTs) that changes the colour when reduced, using a spectrophotometer or plate reader. The conditions of the high-throughput method were established for identification and quantification of (3S)-1,3-dihydroxypentan-2-one (HK) from TK bioconversions between Li-HPA and PA. One of the main advantages of this method, is the stability for measurements using WST-1, contrary to the 2,3,5-triphenyltetrazolium chloride (TZC) reported by Smith et al., 2006 that produce a very unstable formazane that cannot be readily quantified with accuracy. Depending on the HK concentration the method could be divided in two different concentration ranges: from 0 to 1.2 mM that gives different blue tonalities, and the range between 20 and 50 mM, that shows different orange tonalities that can be detected at 450 nm.

In this chapter a novel colorimetric assay for determining the activity of TK was demonstrated. The method was firstly proven qualitatively using the product reaction from TK lysate of the variant D469T and the use of MP resin that absorbed the remaining Li-HPA substrate from the bioconversion, with a maximum capacity of Li-HPA absorption of 0.7 mili-mole per g of resin, previously established. However further experiments showed that using the standard conditions for bioconversions: TK (0.33 mg/mL), Li-HPA (60 mM), PA (60 mM) and 72 hours or more of reaction, the final concentration of Li-HPA is below 5 mM and in consequence the use of MP resin is not necessary.

Reaction parameters such as optimal enzyme/substrate ratio, incubation time, lower and higher detection and quantification limits were determined. The assay was then performed by determining the accuracy and the repeatability of the results obtained using different TK samples. To demonstrate the practicality and accuracy of the microplate method, the assay was tested and compared with HPLC measurements. The correlation between HPLC and colorimetric method was established, using TK from WT and the quadruple variant H192P-A282P-I365L-G506A. Firstly, HK was measured in HPLC using UV detection at 210 nm and identifying the peak product at 15.25 minutes, with the same samples the colorimetric method was performed under the conditions previously described and measured after one hour of reaction in the spectrophotometer at 450 nm. This correlation provided the

possibility to use the CM not only to screen variants qualitatively, but also for quantitative measurements of HK, saving time for the measurement of high number of samples. Taking into account the conditions in our laboratory, it is possible to analyse around 2,500 samples in less than three hours and potentially more than 6000 samples per day compared with ~100 samples per day that could be analysed using the standard HPLC method for this specific reaction.

5 Development of Mutagenic Libraries

5.1 Development of Mutagenic Libraries Using MEGAWHOP Method

Advance in molecular biology allows to introduce sufficient genetic diversity by several methods like random mutagenesis, site-directed mutagenesis, and semi-rational design (Xiao et al., 2015). We can say that the growth in the development of enzyme technology is still in the dynamic phase and it is expected to give astonishing outcomes in the next years (Madhavan et al., 2017). The classical method for cloning of a DNA fragment requires the digestion of the corresponding DNA sequence with restriction endonucleases and ligation into a vector having compatible ends (Sambrook & Rusell, 2001). Although this method is one of the most common and efficient for generation of mutagenic libraries, it is laborious, and the resulting libraries could be often contaminated with unwanted plasmids that have no inserts or multiple inserts. These undesirable plasmids are problematic specifically for random mutagenesis as they lead to false positive/negative after subsequent functional screening. In this sense, MEGAWHOP (megaprimer PCR of whole plasmid) is an alternative that provide a ligation-independent cloning method (Miyasaki, 2011).

The MEGAWHOP method, was firstly introduced in 2002 by Miyasaki and Takenouchi (2002), and it has been successfully applied for the generation of mutant libraries of a wide range of genes, for example to improve the solubility of fungal glycosyl hydrolase when overexpressed in *E. coli* (Yaoi et al., 2007). For that specific project, three rounds of directed evolution were enough to convert the aggregation-prone enzyme to a fully soluble one. Another example is the one reported by Kirschner and Bornscheuer (2008), that used the MEGAWHOP method in the Baeyer-Villiger monooxygenases and allowed to improve its enantioselectivity. The random mutagenesis library consisted of over 3500 clones with two variants identified for demonstrating improved selectivity. In this sense, the number of clones or cycles for the generation of an enzyme with desirable characteristics cannot be defined.

In MEGAWHOP, the DNA fragment to be cloned serves as a set of complementary primers, that are much longer than common oligonucleotide primers, that is the reason of the name, ‘megaprimer’. The MEGAWHOP cloning method, consists of four steps: preparation of the DNA fragment to be cloned using for example PCR error-prone; whole plasmid PCR (high fidelity) using the megaprimer, which serves as a set of overlapping primers; *DpnI* digestion of the template plasmid; and finally the introduction of the *DpnI*-treated mixture into competent *Escherichia coli* cells

to yield a plasmid with a mutated insert. Moreover, error-prone PCR (ep PCR) was first reported by Leung, D. W., et al., (1989). *DpnI* is a restriction enzyme which digest methylated DNA, that could be formed during PCR process. Methylation can change the activity of a DNA segment without changing the sequences.

PCR error-prone generates point mutations during PCR amplification of a gene using a low fidelity DNA polymerase under certain conditions like higher concentration of Mg^{2+}, addition of Mn^{2+} and the use of unbalanced dNTP concentrations that reduce the base-pairing fidelity and increase mutation rates to 10^{-4} ~10^{-3} per replicated base. The mutations are accumulated during each PCR cycle of amplification; in consequence the average number of mutations can be increased by increasing the number of PCR cycles (Packer & Liu, 2015). The use of MEGAWHOP and ep PCR as a method for generation of mutagenic libraries have shown to be an effective method for creation of random mutagenesis libraries. Libraries produced using this technique are virtually free of contamination by plasmids without any insert or with multiple inserts. In this chapter we focused on development of the conditions of PCR to generate mutagenic TK libraries. In this chapter, the TK gene was divided in four different sections for the generation of mutagenic libraries. Finally, random variants were analysed to determine the number of changes of nucleotides and amino acids in each library.

5.2 Development of Mutagenic Libraries

Modification of the substrate specificity of TK by random mutagenesis promises to broaden their application range. The development of new enzymes, and the improvement of the known ones, is essential for further progress in biocatalysis (Hecquet et al., 2014). Molecular cloning is an essential prerequisite to test protein design and engineering ideas. Nevertheless, it is not a straightforward method for every experiment, it could be time consuming and unreliable. For example, ligation dependent cloning procedures require many steps such as PCR restriction digest, dephosphorylation and ligation. The addition of steps could decrease the overall yield, making the entire procedure inefficient. There are several methods reported as useful for protein engineers that include all the reaction conditions (time, temperature, number of cycles for PCR, etc.), however, sometimes these conditions must be adjusted for every specific case. (Speltz & Regan, 2013).

Although the MEGAWHOP technique is relatively easy and fast to perform, there are several technical considerations. Product yield is dependent on the amount of megaprimer and the number of cycles used during the whole-plasmid PCR (high fidelity). Size and concentration of megaprimers also must be considered. Some authors have suggested 0.2 µg of megaprimer (~750 bp) and 50 ng of template plasmid (~3.5 kb) with a PCR run of 24 cycles to get around 2000 transformants, however in some instances the number of transformants could be lower than expected and the number of cycles and/or concentration of megaprimer must be adjusted. On the other hand, the speed of a PCR depends on a number of factors, such as the extension rate (standard rate ~1 kb/min) of the polymerase used, the ramp speed of the thermocycler and the size and complexity of the DNA template.

After having determined the conditions for the screening method, the next objective of this project was the development of mutagenic libraries (ML) using H192P-A282P as a DNA template. This double mutant H192P/A282P was selected because the combination of A282P with H192P resulted in a variant with a 3-fold improved half-life at 60 °C. This variant was found to be more stable than both H192P and A282P separately, with a 7-fold improvement in residual activity compared with WT and retained 50% activity after 1 hour at 60 °C, making this variant a potential candidate to be used in bioconversions at elevated temperatures. H192P/A282P improved the half-life 3-fold relative to WT (Yu et al., 2017).

In this project, MEGAWHOP conditions from previous successful experiments (Miyasaki et al., 2002; Miyasaki, 2011) were tried using the same conditions of temperature and number of cycles for MEGAPRIMER and MEGAWHOP generation, and after several failed attempts the final conditions were established. Some of the failed variables are the following: 5, 10 and 25 ng of DNA template H192P-A282P. 1 and 5 minutes of initial denaturation at 95 °C. 15, 20, 24, 35 and 40 cycles of PCR. 68 °C for extension and only 5 minutes of final extension time. No addition of DMSO to the MEGAWHOP PCR, among others. Several PCR conditions were tried without positive results for the generation of mutagenic libraries. The conditions mentioned before were selected from several reported protocols without any positive result. Conditions mentioned in Section 2.25 are the result of several months of work that are not shown in this thesis.

PCR's were carried out following the instructions of the *GeneMorph II Random Mutagenic Kit* from Agilent with the components and conditions described in Section 2.25. The first step to generate the mutagenic libraries was to carry out a PCR on the whole plasmid template to generate the megaprimer containing random mutations. These reactions used specific primers to amplify four different sections that collectively spanned the whole TK gene. An extra PCR using random mutagenesis was carried out, but in this case the whole TK gene was involved, using primers at the beginning and at the end of the whole TK gene (see Table 2.18). Megaprimers were visualised using an 1% agarose gel and results are shown in Figure 5.1.

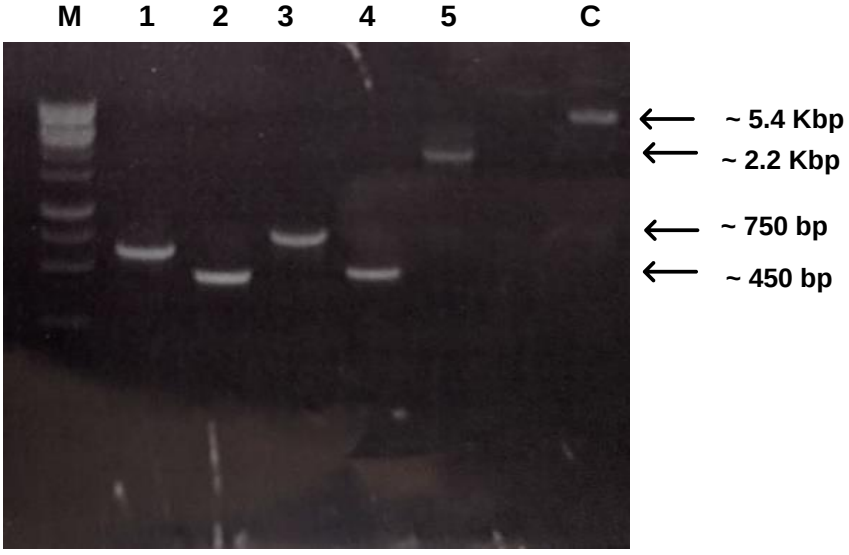


Figure 5.1. 1% Agarose gel of Megaprimers (MP) from the first (mutagenic) PCR step. M (Marker), 1, 2, 3, 4 and 5 (Megaprimers 1, 2, 3, 4 and 5) and C (Control) template H192-282P.

According to the results shown in Figure 5.1, DNA bands in the agarose gel were well defined in all the PCR reactions, showing successful results in all cases, with the expected sizes of MP 1 (616 bp), MP 2 (449 bp), MP 3 (732 bp), MP 4 (467 bp) and finally, the whole TK with 2096 bp, in lanes 1, 2, 3, 4, and 5 respectively. The first four megaprimers, covered segments of the full TK gene, while the fifth one was an error-prone PCR of the whole TK sequence.

After obtaining the correctly sized mutagenic megaprimers, a second PCR for each megaprimer was performed with the template plasmid, using the reagent mix shown in Table 2.21 and 24 PCR cycles before the final extension at 70 °C (see Table 2.22). This second PCR is designed for inserting the mutation-harbouring megaprimers back into the correct location of the template plasmid. This generates linearized whole plasmids containing the mutations, but with complementary overhangs at each end to aid recircularization. The products of the second PCR were visualized on a 1 % agarose gel, and the results are shown in Figure 5.2.

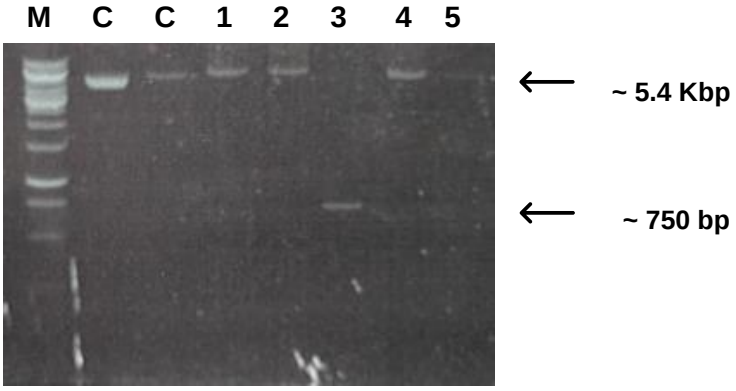


Figure 5.2. 1% Agarose gel of MEGAWHOP PCR products. M (Marker), 1, 2, 3, 4 and 5 (MEGAWHOP 1, 2, 3, 4 and 5) C (Control TK template).

As it can be appreciated in Figure 5.2 samples 1, 2, 4 and 5 were amplified correctly, however, sample 3 was not amplified, showing only the same band that corresponds to the Megaprimer seen in Figure 5.1. It was not clear why this specific PCR failed, although sample 3 is the second largest megaprimer, where the largest (sample 5) performed at least partially successfully. MEGAWHOP reactions (1, 2, 4 and 5) were transformed into *E. coli* XL10 Gold competent cells as described in the *QuickChange Kit protocol* and plated onto LB-agar with 150 µg/mL ampicillin. After incubation at 37 °C overnight, approximately 1000 colonies were obtained (see Figure 5.3).

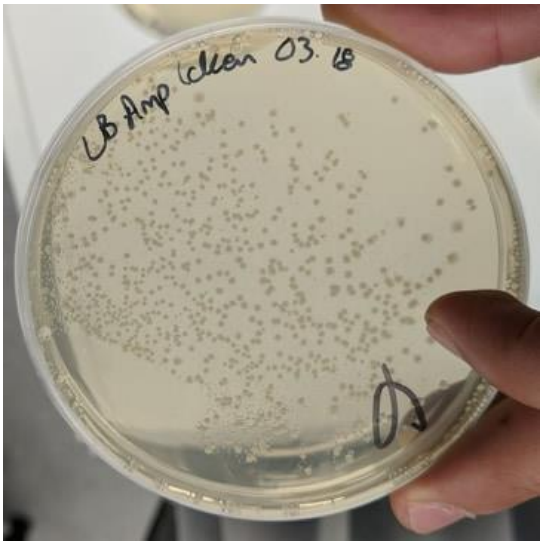


Figure 5.3. Sample of a mutagenic library after transformation. Around 1000 colonies can be observed on LB agar with ampicillin. Glycerol stocks were frozen to secure the library.

Although the transformation efficiency of XL10 Gold competent cells is reported as very high (≥5x10^9 cfu/µg pUC18 DNA), this efficiency could decrease dramatically by the storage conditions. On the other hand, ultracompetent cells are very sensitive to even small variations of temperatures. Transferring tubes from one freezer to another may result in a loss of efficiency. Additionally, transformation efficiency of the MEGAWHOP product is around 10^4 to 10^5 mutants/µg DNA, lower than that of the optimized transformation efficiency of 10^6 mutants/µg ligated DNA. In some cases much smaller-size protein libraries (10^{3-6} mutants) are usually generated during the subcloning step, this could be due to the low digestion efficiency of the vector and inserted fragment, low efficiency of ligation, low efficiency of *DpnI* digestion, and possible self-ligation of the digested plasmid. Efficiency could be improved by a series of optimizations during the cloning process. (You et al., 2012). Number of cycles could be changed in the whole plasmid PCR while fixing the concentrations of the megaprimer from 0.5 ng to 50 ng of template. Process optimization will maximize the yield of transformants to get transformation efficiencies of greater than 1x10^5 µg/megaprimers using 100 µL of ultracompetent cells (Miyazaki, 2002). Although the efficiency dropped under suboptimal conditions, it was enough to establish the first PCR and transformation conditions for the generation of TK mutagenic libraries using the MEGAWHOP technique.

Mutation rate is an essential parameter in directed evolution. The number of individual mutants gives information for a direct estimate of mutation rate. In some cases, mutant individuals in the same family can share the same mutations due to premeiotic mutation events, this can be tranduced to significant larger number of individual mutants than the number of mutations in amino acids observed. In this sense, counting mutant individuals is a correct approach for estimating mutation rate, whereas counting only mutation events could be considered as an underestimated result (Fu, Yun-Xin & Huai, 2003). Determination of the mutation rate of the mutagenic libraries are associated to different factors. The Taq polymerase used was the Mutazyme II DNA polymerase that is a novel error prone PCR enzyme blend, the formulation of it provides useful mutation rates with minimal mutation bias. This enzyme is a blend of two error prone DNA polymerases – Mutazyme I DNA polymerase and a novel *Taq* DNA polymerase mutant that exhibits increased mis-insertion and mis-extension frequencies compared to wild type *Taq.* The buffer used in the PCR contains Mn^{2+} and unbalanced deoxynucleotide concentration, which are mutagenic conditions for *Taq* DNA polymerase. This specific mix of polymerases produce a less biased mutational spectrum with equivalent mutation rates at A's and T's vs. G's and C's. In other words, libraries generated with Mutazyme II exhibit greater mutant representation compared to libraries generated with other enzymes.

The conditions used in the PCRs were designed to get mutation rates of 1-16 mutation per kb. Another factor that influences the mutation rate is the concentration of the initial amount of target DNA in the reaction. For a mutation frequency between 9-16 mutations/kb (considered high) the template concentration of DNA must be between 0.1 and 100 ng. Moreover, if a low mutation rate is desired between 0 and 4 mutations/kb, the initial concentration of DNA must be between 500 and 1000 ng. Mutation rates also can be decreased by lowering the number of cycles to achieve fewer target duplications. For example, targets that produce high product yield after 30 cycles, lower mutation rates can be achieved by amplifying lower target amounts for 20-25 cycles. The desired mutation frequency for protein structure-function relationship is one amino acid change (1-2 nucleotides) per gene (Vartanian et al., 1996). However, this is not a rule, other authors have shown improved activities from highly mutagenized libraries with 20 mutations per gene (Daugherty et al., 2000).

Diversity in mutagenic libraries is a key factor for success in directed evolution. The diversity of a library is the number of unique variants, these need to be sufficiently diverse to find improved variant 'winners'. However, mutations should not be high to

be swamped by non-functional protein. For best identifying the 'winning' variants in an error prone PCR library, the mutations need to be at an optimal frequency between 5 to 10 mutations/kb. In order to ensure an effective diversity of a mutagenized plasmid pool it is recommended to assess the diversity in a small test library, that allows to avoid heavily investing in a potentially suboptimal library further downstream (Ferla, 2016; Hanson-Manful, 2013).

Once the mutagenic libraries were generated and *E. coli* culture transformed and plated on LB media with ampicillin, 24 colonies in total, six of each mutagenic library (1, 2, 4 and 5) were randomly selected and sequenced to calculate the overall mutation frequency and the individual mutation frequencies. The gene sequences were used to determine the mutation frequency and the estimation of error associated with the values found. Sequences were compared using BLAST with the TK template (H192P-A282P) used for the generation of the mutagenic libraries (*https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins).* Average mutations of nucleotides with an average mutation frequency ± standard deviation (SD) values of three independent samples (n=3), are shown in Table 5.1, whereas amino acid mutations are indicated in Table 5.2.

Moreover, is important to clarify that the number of mutations in every single colony from the same library may vary. The number of mutations per sequence is a key indicator of the diversity of the library. In this thesis the mutational frequencies were calculated as it was mentioned before, however, there are some software such as *Mutanalyst* (available at *www.mutanalyst.com*) that allows to perform the same analysis automatically and simplifies laborious checking of sequence for mutations. This rough sampling is generally enough to estimate the main contributors of the mutational spectrum of the library. Nevertheless, if the objective is to identify the variety of sequence-specific mutation hotspots and coldspots, it is necessary to analyse a higher number of samples that will provide a highly accurate picture of the diversity of the library (Ferla, 2016; Zhao et al., 2014).

Table 5.1. Average number of nucleotide mutations of four different mutagenic libraries compared with template H192P- A282P. SD represent one standard deviation of the mean (n=6).

Mutagenic Library	Average of Nucleotides Mutations	SD
1	38	2.5
2	24	2.6
4	34	3.9
5	16	2.5

Using the previous PCR conditions proved that a variety of random mutants can be generated. In some cases, a single DNA mutation is not enough to translate the sequence to different amino acids. The table 5.2 shows the generated mutations and positions that were detected for the same colonies that were previously selected and sequenced.

Table 5.2. Observed amino acid mutations generated in each mutagenic library. H192P-A282P was used as a template.

Mutagenic Library	Mutation
1	A8S
	A8N
	K105Y
	Q136R
	Y182N
2	A202T
	M203K
	H261R
4	F325L
	A547T
	A595D
	K603E
	G631D
5	I11S
	R59C
	E289G
	L516P
	A537V

Having generated a small mutagenic library (ML), to establish a complete screening process, the colorimetric method was applied to 96 variants of each ML, with the aim of applying the screening methods developed in Chapter 4 to the mutagenic libraries generated to perform an initial screening of TK variants with possible altered thermostability. Thermostability was tested using 2 mL of microwell fermentation culture in 96 deep-well plates (DWP). For each variant 200 µL of cells were used for protein extraction by freeze-thawing method. Afterwards, lysates were resuspended in the cofactor mix and heated at 60 °C for 1 hour in a thermal cycler as described by Yu Haoran (2017). The bioconversion reaction continued for 24 and 48 hours before performing the colorimetric method.

After the heat treatment at 60 °C for 1 hour, variants with higher thermostability than the template must show activity that could be easily identified using the colorimetric method previously developed in this project. However, all the samples analysed did not show any significant change of colour (all of them showed a blue colouration after colorimetric assay) that indicates that the product of bioconversion is lower than 5 mM or in other words none of analysed variants showed a significant

higher thermostability. Therefore, while both methods (conditions for development of mutagenic libraries and HTP screening method) were now established, the next steps for further projects will be the screening of mutagenic libraries with a larger number of variants using the colorimetric method described in this thesis in order find new thermostable TK variants.

5.3 Conclusions of Mutagenic Libraries

The first conditions for generation of mutagenic libraries with low mutation rates were established using the MEGAWHOP strategy that involves 2 PCR reactions: the first generates the Megaprimers, and the second amplifies the whole plasmid sequence. The TK gene was divided into 5 different parts due to previous observations that the amplification in the second PCR is more effective for shorter Megaprimers (less than 800 bp). While the conditions were successful established in generating mutants, it is necessary in future works to analyse a larger quantity of samples using the new colorimetric method designed in this project with a previous heat-shock step, in order to know whether the libraries contain mutants with higher thermostability or not. Thermostability of the samples were screened at 60 °C considering that the TK variant used as a template H192P-A282P has proved to retain 50% of its activity after 1 hour at 60 °C, which indicates an improved potential to be used in bioconversions at elevated temperatures (Yu H. et al., 2017). In this sense, it is expected to get a variant with higher thermostability than the template. However, for further mutagenic libraries it could be suggested to screen thermostability at lower temperatures to get a thermostable top variant that could be combined with other mutations in order to get more stable TK variants, with improvement in residual activity compared with the template. The next chapter explores the final objective of this project which was to obtain a series of C-terminally truncated variants by directed evolution, enable their purification, and perform a more detailed biophysical analysis of them.

6 Truncated Transketolase Variants

6.1 Introduction of Truncated TK variants

TK structure is composed by three main domains, PP domain, Pyr domain and C-terminal. The first two domains have been reported to be essential for TPP binding and catalysis while the TKC domain remains undefined. In 2007 Costelloe et al., reported interesting data about the relationship between PP and Pyr domains in transketolase. They found that deletion of the TKC-domain does not adversely affect the activity of the enzyme, and so the function of the TKC domain is still unknown. Variants without the TKC-domain improved the activity towards β-hydroxypyruvate and glycolaldehyde, suggesting that the functional unit of TK remains in the PP and Pyr-domains, and in consequence this evolutionary result could be extended to all TPP-dependent enzymes that contain the catalytic PP and Pyr-domains where the TPP cofactor binds.

Different truncated and active variants have been reported, for example the TK 540stop that removed all the TKC-domain showed almost three times more activity using GA and Li-HPA as a substrate than WT. Some other truncated variants (527Z, 492Z, 461Z, 453Z) have been reported introducing stop codons into loop regions of the Pyr-domain, making the TK structure progressively shorter (Costellos S, et al., 2007) using pQR711 as a plasmid to express transketolase (French & Ward 1995). However only the 453Z variant showed better activity than WT, while 527Z, 492Z and 461Z lost their activity.

This chapter examines the effect of removing gradually the C-terminal and the Pyr-domain using stop codons. Continuous truncations of TK from the C-terminal end and part of the Pyr-domain were proposed by insertion of a stop codon in order to generate shorter variants of TK (453Z, 425Z, 403Z, 375Z, 349Z) see Figure 6.1. The stop codons keep the secondary structure of TK intact, and the positions were selected on the loops of the protein structure. Another interesting approach is to have the possibility to improve the desirable characteristics (as activity, stability or thermostability) of the generated truncated versions by applying the methods previously described in Chapters 4 and 5. Stop codons were previously analysed and designed using Pymol, in order to avoid the disruption of the secondary structure of TK. The objective of these experiments is to identify the shortest active TK variant using Li-HPA, GA and PA as substrates.

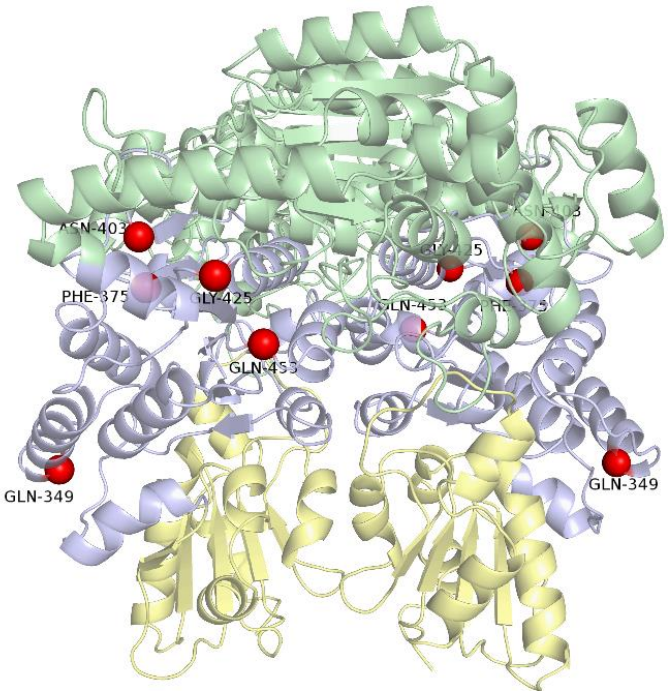


Figure 6.1. Each coloured region corresponds to the different regions of TK, the pyrophosphate (PP)-binding domain (3-322 aa) green, the pyrimidine (Pyr)-binding domain (323-538 aa) blue and the C-terminal domain (539-680 aa) yellow. The C-terminal domain was removed, and the Pyr-domain was also partially removed by insertion of stop codons into the TK gene. Positions of the stop codons are red sphere and amino acids at these positions are also indicated. Image generated using Pymol software.

6.2 Generation of Truncated TK Variants

Truncated variants were generated using site-directed mutagenesis, carried out using the Quickchange kit (Strategene), and the PCR conditions described Section 2.23. After the PCRs, an agarose gel at 1% was made to verify the amplification of the TK gene. However, the DNA gel was not useful for distinguishing between truncated samples because the stop codons on the TK sequence does not alter the size of the TK gene.

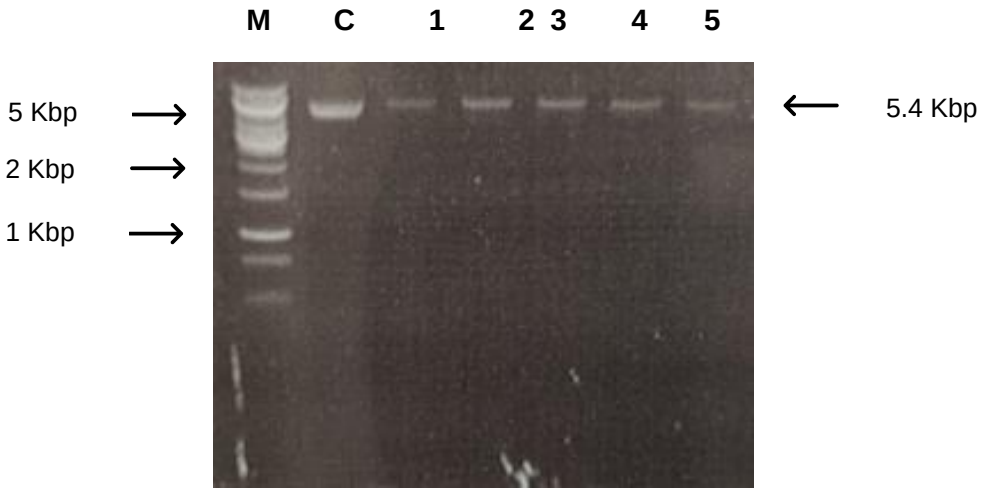


Figure 6.2. 1% Agarose gel of PCRs for construction of TK truncated variants. M (Marker), C (TK - H192P/A282P), 1 (Q453Z), 2 (G425Z), 3 (N403Z), 4 (F375Z), 5 (Q349Z).

As can be seen in Figure 6.2, all samples were amplified correctly with sufficient concentration of each reaction product. These samples were digested with 1.5 µL of Dpn-1 for 2 hours at 37 °C and transformed into *E. coli* XL10 Gold as it is described in the QuickChange XL Site-directed Mutagenesis Kit. 50 µL of each truncated mutant was plated in agar LB + ampicillin and incubated at 37 °C overnight. Table 6.1 indicates the number of colonies grown in each petri dish.

Table 6.1. Number of colonies in each Petri dish after 16 hours of incubation at 37 °C.

TK truncated variants	Number of colonies
Positive Control pUC	300
Negative Control	0
Q453Z	280
G425Z	166
N403Z	245
F375Z	210
Q349Z	130

Six colonies of each truncated mutant were selected randomly and sequenced to verify the presence of the stop codon in the correct position. This would indicate the successful truncation of the TK. All the samples showed the codon correctly. After the verification of the sequence, fermentations to produce each truncated version were performed in baffled flasks of 1L with a working volume of 200 mL of Magic Media for extraction of TK as described in Sections 2.11 and 2.12 (see Table 6.4). Total protein was measured by Bradford method (see Section 2.19) and results are shown in Table 6.2.

Table 6.2. Total protein in lysate of truncated TK variants.

TK truncated variants	Total protein [mg/mL]
Q453Z	10.79
G425Z	4.88
N403Z	8.41
F375Z	9.29
Q349Z	10.79

To determine whether the five truncated versions were active or not, bioconversions were carried out over 96 hours, taking samples at 0, 24, 72 and 96 hours, and performing two different bioconversions, the first one using Li-HPA and GA, and the second one using Li-HPA and PA as substrates. Samples were quenched with TFA (0.1%) and stored at 4 °C until analysis. As the colorimetric method was not suitable for bioconversions using glycolaldehyde as substrate (see Section 4.3), measurements of both bioconversions were performed by HPLC, using the conditions described in Section 2.21. In this case bioconversions were carried out over 96 hours, allowing sufficient time to consume 95% of Li-HPA. Conditions for these bioconversions are given in Table 6.3.

Tables 6.3. Volumes and concentrations for bioconversions of truncated TK versions.

Reactive or substrate	[Initial]	Volume [µL]	[Final]
TK lysate (total protein)	1 mg/mL	400	0.3 mg/mL
Li HPA	150 mM	400	45 mM
Gly	150 mM	400	45 mM
PA	150 mM	400	45 mM
Cofactor solution	12 x	140	1x

In addition, two different variants previously generated, G425Z and R520Z, were produced and analysed to determine their TK expression level, using sonication as the extraction method. For both variants, the total protein concentration was below 1 mg/mL (see Figure 6.3). These results showed that transketolase was not well expressed. Even the lysate did not show a band that could correspond to TK (~72 KDa). In this sense, for future experiments, both variants must be sequenced to confirm if they have a stop codon in the corresponding position. Concentration of both variants were quantified in their lysate and pure enzyme forms and results that are shown in Table 6.4, confirming the very low total protein concentration obtained from these mutants (Figure 6.3). Based on these results it is very difficult to express TK in these two variants, so both were discarded from further experiments and bioconversions.

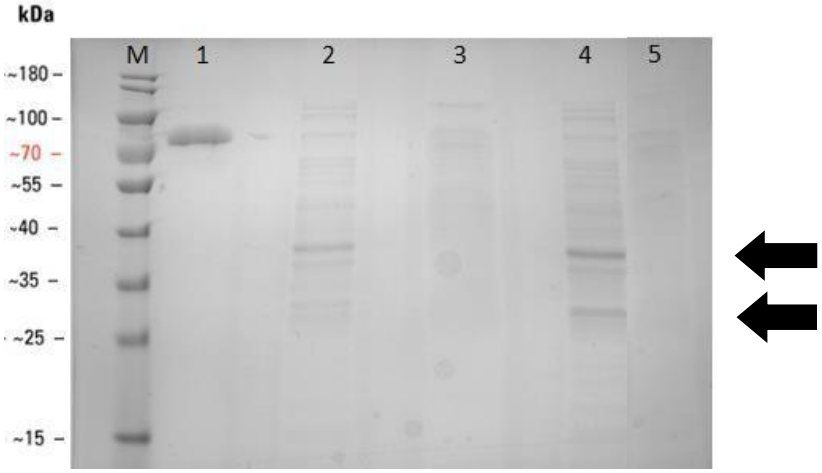


Figure 6.3. SDS PAGE gel of transketolase (WT-TK) lysate and pure protein using different methods of TK extraction. Arrows show bands of proteins (between 25 and 40 kDa) that does not correspond to TK (Lane 1: WT purified standard, Lane 2: G425-stop lysate, Lane 3 G-425 stop purified, Lane 4: R520-stop lysate, Lane 5: R520 stop purified.

Table 6.4. Concentration of total protein in variants G425-stop and R520-stop after 8 hours of fermentation at standard conditions. Last column shows the protein loaded on SDS gel.

Sample	Name	Concentration (mg/mL)	Protein loaded on gel (mg)
M	Marker		
1	WT Pure	1.47	0.01
2	G425-stop lysate	0.91	0.006
3	G425-stop pure	0.14	0.001
4	R520-stop lysate	1	0.007
5	R520-stop pure	0.14	0.001

6.3 Bioconversion of Truncated Variants

Finally, activity of truncated variants was tested using GA and PA as substrate in standard conditions for 24, 72 and 96 hours. A negative control was used as a reference to quantify the change of Li-HPA concentration through the time. According to Figure 6.4, after 24 hours the total concentration of Li-HPA decreased to less than 1 mM in bioconcersions with all the truncated TK versions, showing a fast consumption of Li-HPA.

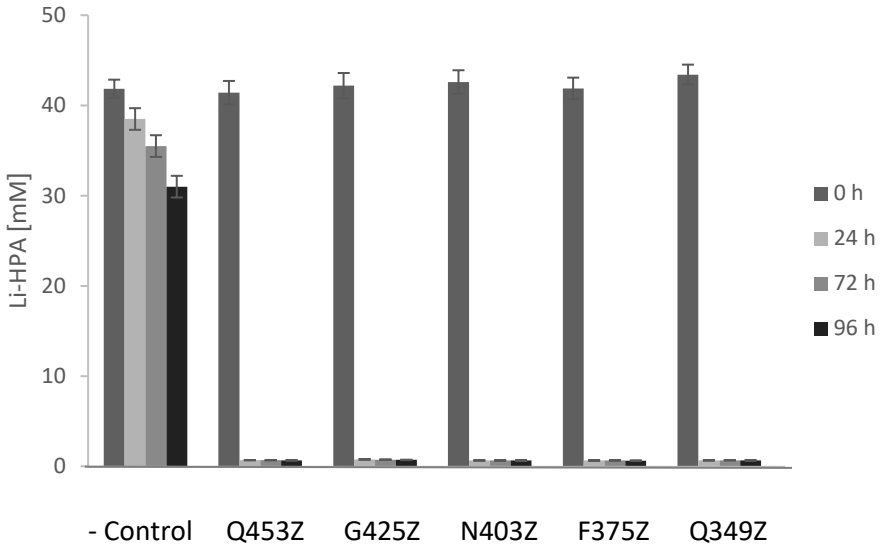


Figure 6.4. Remaining Li-HPA in the bioconversion of truncated TK versions with GA at different times. Error bars represent one standard deviation of the mean (n=3).

On the other hand, results shown in Figure 6.5 also confirm that the yield of the bioconversion of GA to erythrulose was around 75% after 24 hours, which remained relatively constant even after 96 hours for all five different truncated TK variants. However, results using PA as a substrate were surprisingly different. Bioconversions were much slower compared with reactions using GA. Using PA with the same hours of bioconversion, only 45% of Li-HPA were consumed in the reaction, compared with the 95% Li-HPA consumption after 24 hours when using GA (see Figure 6.6).

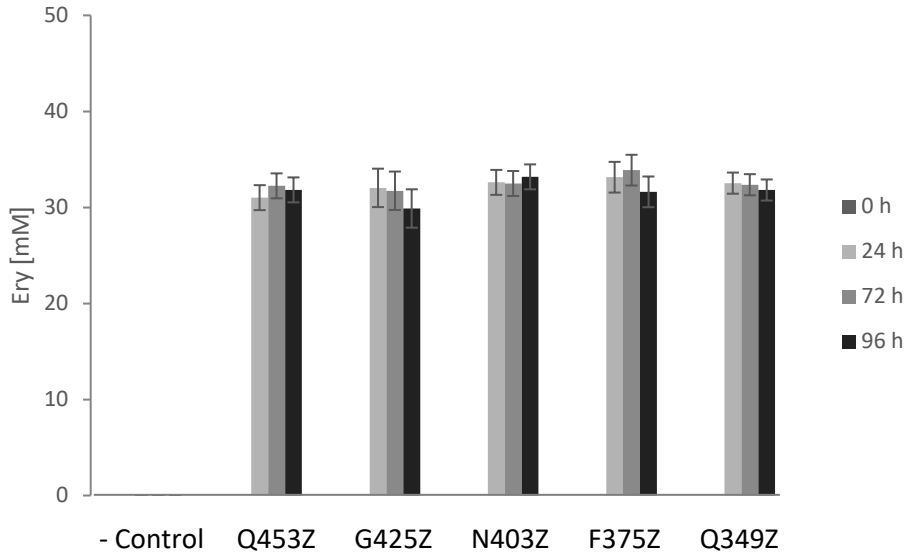


Figure 6.5. Production of erythrulose in bioconversion of truncated TK versions with GA at different times. Error bars represent one standard deviation of the mean (n=3).

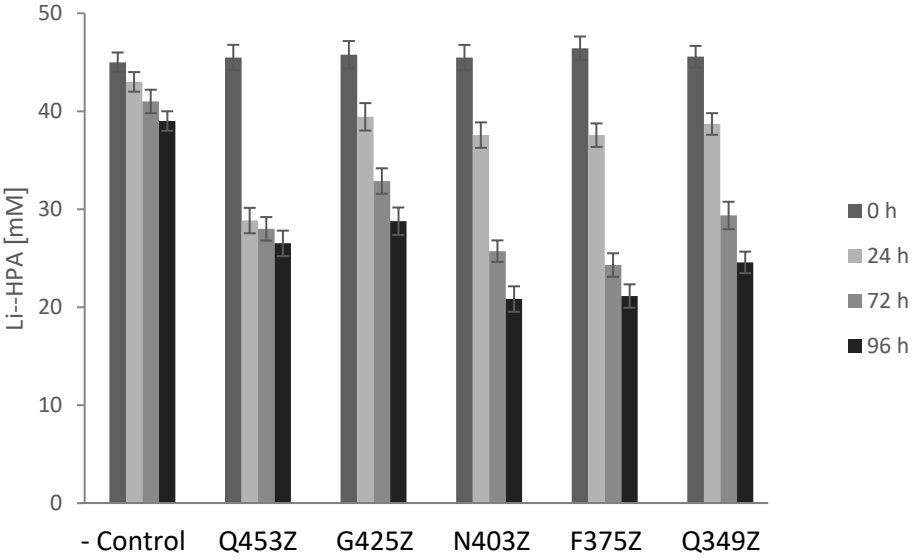


Figure 6.6. Remaining Li-HPA in bioconversion of truncated TK versions using PA as a substrate at different times. Error bars represent one standard deviation of the mean (n=3).

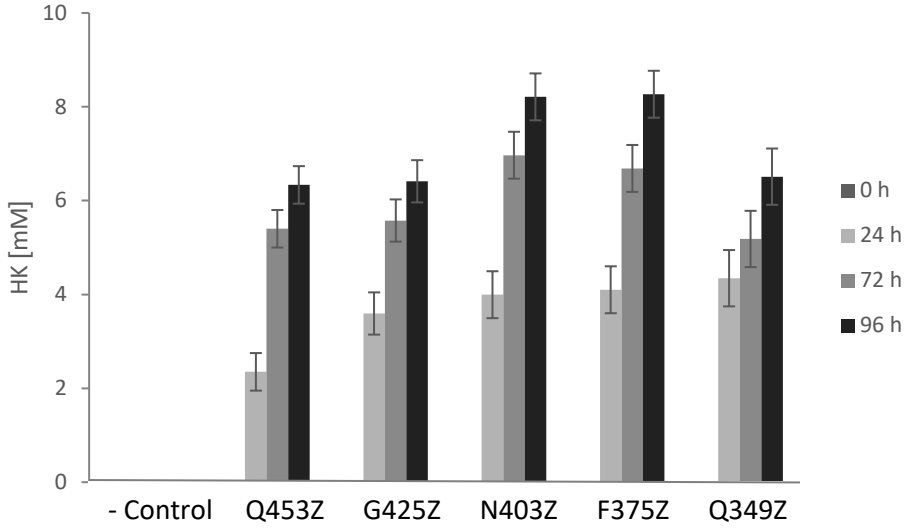


Figure 6.7. Production of 1,3-dihydroxypentan-2-one in bioconversion of truncated TK versions with PA at different times. Error bars represent one standard deviation of the mean (n=3).

Due to the slow bioconversion of Li-HPA and PA, the yield of product was quite low. Variants N403Z and F375Z gave the highest and similar yields, of around 16% after 96 hours of reaction (see Figure 6.6). Another interesting point was that even in the control sample, the concentration of Li-HPA decreased by 10% every 24 hours, presumably due to degradation, and this must be considered for future analysis and experiments.

6.4 Conclusions of Truncated Variants

Six different truncated and active variants were generated keeping the PP-domain and N-terminal 6-His tag tail, using primers that contained a stop codon 1 (Q453Z), 2 (G425Z), 3 (N403Z), 4 (F375Z), 5 (Q349Z). All variants expressed activity after 96 hours of bioconversion using Li-HPA + GA and Li-HPA + PA as substrates. However, all truncated variants showed higher activity using GA as a substrate, giving 66% of bioconversion yield contrary to only 16% of yield with PA. Hence, it is suggested to define a strategy to mutate the generated variants to improve the activity of them.

7 Final Conclusions

In this project, several fermentation conditions for TK production and extraction were defined. The influence of different growth media for *E.* coli and TK production (Sections 3.9, 3.10 and 3.11) was shown. Moreover, advantages and disadvantages of different extraction methods were analysed (Section 3.7). The best growth media for biomass and TK production was the commercial Magic Media that allows to reach six times higher biomass and total protein production compared to the most common media such as TB and LB. On the other hand, it was also shown that addition of 7.5% of glycerol to LB and TB did not resulted in a significant difference in terms of biomass production and in consequence it is not necessary to add glycerol as a source of carbon for further experiments. Additionally, it was shown that even when chemical reagents were effective for protein extraction, they are not convenient to use due to the high viscosity that is produced in the protein solution, making samples not suitable to measure and manipulate for further steps such as purification using Ni-NTA columns or bioconversions. On the other hand, sonication does not change the viscosity of the protein solution, and the difference of protein extraction is only 14% less compared to the chemical methods, hence this method is very convenient and effective to use for protein cell disruption. Finally, the freeze-thawing technique could be considered as an effective alternative for protein extraction when a high number of samples must be analysed. The main disadvantage of this methodology is its relatively low yield in protein extraction (around 25%) compared with the sonication method. The aim of chapter three of this thesis was achieved, providing fundamental knowledge for setting up basic conditions for successful transketolase production and protein extraction that can be used as a guide for future experiments related to generation and screening of TK variants.

The main achievement on this project was the development of a colorimetric high-throughput screening method that allows to identify new TK variants from mutant TK libraries with specific desired characteristics such as thermostability. The methodology to screen active TK variants and measure the product of bioconversion was established using Li-HPA and PA as substrates. Compared with previous colorimetric methods as the one reported by Smith et al., (2006), the method proposed in this project was more reliable as the coloured compound does not precipitate making it suitable for screening large number of TK variants and quantify the bioconversion product. The standard curve obtained from the colourimetric method was validated with the experiment described in section 4.10 in which WT-TK and H192P-A282P-

I365L-G506A TK were used for the bioconversion an equimolar concentration of Li-HPA and PA, after 72 h the reaction and used for colorimetric analysis. Also the same reaction mix was prepared for HK concentration analysis using HPLC. The concentration results using colorimetric method and HPLC were the same with a SD of +/- 0.005 mM, which indicates that using the colorimetric method avoids the need of HPLC analysis for this reaction scheme. Furthermore, the stablished conditions of the developed method are suitable to screen and quantify from TK bioconversions, making this method comparable to analytic methods such as HPLC. However, the method has some limitations and the product concentration of the bioconversion must be between 20 and 50 mM that corresponds to a reasonable HK concentration in TK bioconversions. Efficiency of the method was tested using different TK variants such as the D469T for qualitative measurements (Section 4.7) and the H192P-A282P-I365L-G506A for quantitative experiments (Section 4.10). In addition, it was shown that the use of MP resin to absorb the remaining Li-HPA would not be necessary if the conditions of the bioconversion allow to reduce the concentration of Li-HPA below 5 mM (Section 4.9).

In chapter five, conditions for the generation of mutagenic libraries using the MEGAWHOP strategy were stablished using the H192P-A282P TK mutant as a template. Libraries generated were sequenced to acquire the number of mutations per gene, showing an average mutation between two and five changes of aminoacids per gene, that have been reported to be suitable to get variants with desirable characteristics such as higher thermostability. Generation of mutations around active site (10 Å), could improve the enantioselectivity and stability of TK. These findings provide the key parameters for the generation of mutants with potential higher thermostability that can be screened and measured using the HTS method described also in this thesis. Finally, the hypothesis of generating truncated active TK variants was proven, using five designed stop codons at different positions, deleting most of the Pyr-domain that has been associated with TK activity. All TK variants generated (453Z, 425Z, 403Z, 375Z, 349Z) showed activity using Li-HPA and GA or PA as substrates. However, the activity of all truncated versions was significantly higher using GA than using PA under the same conditions. These variants are a starting point to explore the importance of the Pyr-domain in the TK structure. The generation of new TK mutants with these characteristics would be of great interest for a future application in biocatalysis. Overall, this thesis established the basis for further research in directed evolution of transketolase.

8 Future Work

This thesis defined several parameters and methodologies to generate and screen TK variants with significant improved characteristics such as thermal stability. Nevertheless, future work is needed to apply the methodologies here described to get TK variants with the desirable characteristics. A list of proposed future works is described below.

- In order to get higher diversity of TK variants, mutagenic libraries using the conditions described in chapter four could be generated. Mutagenic libraries must explore a range of different mutation rates through optimisation of the protocol parameters, that then generate a greater diversity. Achieving a larger library through greater diversity would proportionally increase the probability of obtaining variants with improved thermostability. The most sensitive parameters to optimize include the manganese concentration, the template DNA purity, and the primer concentrations.

- Future mutagenic libraries generated could be screened using the HTP colorimetric method established in this project. This would validate the new screening approach directly on a library that contains a wide range of activities. By combining the assay with different heat challenges prior to the assay itself, the screening approach could also be used to identify variants with improved thermostability.

- Considering that the HTS methodology reported in this thesis is still performed manually, it is suggested to optimize the methodology using an automatic system to have the possibility to analyze higher number of samples per day. The methodology could be adapted to an automated system such as TECAN that is able to test and analyse thousand of samples more efficiently. This approach would not only increase the throughput achievable, but it would also improve the assay accuracy and so minimize further the likelihood of false positive or negative variants.

- Truncated variants could be combined with specific mutations to improve their activity or even the substrate specificity. Furthermore, these combinations could keep improving additional specific characteristics of TK variants such as thermostability or stability to organic solvents.

- Generate different shorter truncated TK variants using stop codons at different

positions to investigate and determine the shortest and active TK variant. It is of significant interest to investigate the role of the C-terminal domain of TK, through step-wise truncation. This would pick apart the role of specific contacts at the C-terminal interface in the dimer, including their impact on thermostability and activity. Additionally, it would lead to a smaller enzyme variant, potentially enabling greater expression levels to be achieved through a reduced metabolic burden per copy of the protein.

9 References

Abedinia M., R. Layfield, S. M. Jones, P. F. Nixon, J. S. Mattick, *Biochem. Biophys.* 1992**.** *Res. Commun*, *183*, 1159-1166.

Acker, M G., et al., 2013. Considerations for the design and reporting of enzyme assay in high-throughput screening applications. Perspective in Science 1, 56-73.

Advances in small-molecule screening and selection. Annu. Rev. Biochem. 79, 563-590.

Aguiar et al. Parasites & Vectors. 2017. 10:300 DOI 10.1186/s13071-017-2240-3

Aharoni, A.; Thieme, K.; Chiu, C. P.; Buchini, S.; Lairson, L. L.; Chen, H.; Strynadka, N. C.; Wakarchuk, W. W.; Withers, S. G. 2006. Highthroughput screening methodology for the directed evolution of glycosyltransferases. Nat. Methods 2006, 3 (8), 609−614.

Allen A. Cheng and Timothy K. Lu. 2012. Synthetic Biology: An Emerging engineering Discipline, Annual Review of Biomedical Engineering, 14: 55-78.

Arnold F H, Chen K. 1993. Tuning the activity of an enzyme for unusual environments: Sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide, Biochemistry, Vol. 90, pp. 5618-5622.

Arnold F H. 1993. Protein engineering for unusual environments, Current Opinion in Biotechnology, 4:450-455.

Association of Manufacturers and Formulators of Enzymes Products. 2020. https://amfep.org/}

Barahona, E., Jiménez-Vicente, E., & Rubio, L. M. 2016. Hydrogen overproducing nitrogenases obtained by random mutagenesis and high-throughput screening. *Nature Scientific Reports*, *6* (November), 1–10. https://doi.org/10.1038/srep38291.

Baud, D. Ladkau, N. Moody, T S. Ward J M. Hailes, H.C. 2015. A rapid, sensitive colorimetric assay for the high-throughput screening of transaminases in liquid or solid phase, ChemComm, Royal Society of Chemistry, 51, 17225-17228.

Becker, S.; Schmoldt, H. U.; Adams, T. M.; Wilhelm, S.; Kolmar, H. 2004. Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. Curr. Opin. Biotechnol. 15 (4), 323−329.

Behrendorff, J. B. Y. H.; Vickers, C. E.; Chrysanthopoulos, P.; Nielsen, L. K. 2013. 2,2-Diphenyl-1-picrylhydrazyl as a screening tool for recombinant monoterpene biosynthesis. Microb. Cell Fact. 12.

Blass J. P., S. Piacentini, E. Boldizsar, A. Baker, *J.* *1997. Neurochem.* 1982, *39*, 729-733. Boder, E. T.; Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. 1997, 15 (6), 553−557.

B.L. Horecker. 2002. The pentose phosphate pathway, J. Biol. Chem. 277, 47965–47971.

Bolívar, F., 2004. Fundaments and Successful Cases of the Modern Biotechnology.

Bradford M. 1976. A Rapid Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding, Analytical Biochemistry 72, 248- 254.

Branden C, Tooze John. 1999. Introduction to Protein Structure, Second Edition, Taylor & Francis Group, Garland Publishing.

Breuer, M. et al., 2002. High-throughput assay of (R)-phenylacetylcarbinol synthesized by pyruvate decarboxylase. Anal. Bional Chem. 374: 1069-1073.

Buchner, J. & Kiefhaber, T. 2005. Protein Folding Handbook (Wiley-VCH, Weinheim).

Bud, Robert. 1993. The uses of life A history of biotechnology, Cambridge University Press, 1993.

Bylina EJ, Coleman WJ, Tanner MA, Young MM, Youvan D.C. 2000. Solid phase enzyme screening. ASM News, 66:211-217.

Cázares, A. et al., 2010. Non-alpha-hydroxylated aldehydes with evolved transketolase enzymes. *Organic & biomolecular chemistry*, 8(6), pp.1301–1309.

Chen B, Cai Z, Wu W, Huang Y, Pleiss J, Lin Z. 2009. Morphing activity between structurally similar enzymes: from heme-free bromoperoxidase to lipase. Biochemistry, 48:11496- 11504.

Chen, I.; Dorr, B. M.; Liu, D. R. 2011. A general strategy for the evolution of bond-forming enzymes using yeast display. Proc. Natl. Acad. Sci. U. S. A. 108 (28), 11399−11404.

Chen, K. C.; Wu, C. H.; Chang, C. Y.; Lu, W. C.; Tseng, Q. Z.; Prijovich, Z. M.; Schechinger, W.; Liaw, Y. C.; Leu, Y. L.; Roffler, S. R. 2008. Directed evolution of a lysosomal enzyme with enhanced activity at neutral pH by mammalian cell-surface display. Chem. Biol. 15.

Chen, K. & Arnold, F.H. 1993. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences of the United States of America*, 90(12), pp.5618–5622.

Chen, M. M. Y., Snow, C. D., Vizcarra, L., Mayo, S. L., & Arnold, F. H. 2012. *Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes*. *25*(4), 171–178. https://doi.org/10.1093/protein/gzs004.

Chica R A, Doucet N, Pelletier J. 2005. Semi-rational approaches to engineering enzymes activity: combining the benefits of directed evolution and rational design, Current Opinion in Biotechnoloogy, 16:378-384.

Clark M. A. et al. 2018. OpenStax, Biology, https://openstax.org/books/biology-2e/pages/3-4-proteins

Copley, S. D. 2015. An evolutionary biochemist's perspective on promiscuity. *Trends*

in Biochemical Sciences, *40*(2), 72–78.

Copp, J. N., Editors, D. F. A. & Walker, J. M. 2014. *Directed Evolution Library Creation IN Series Editor*.

Costelloe, Seán. 2006. Natural evolution and engineering of transketolase. University College London. PhD Thesis.

Costelloe, S.J., Ward, J.M. & Dalby, P. A. 2008. Evolutionary analysis of the TPP-dependent enzyme family. *Journal of Molecular Evolution*, 66(1), pp.36–49.

Dalby, P. A. 2003. Optimising enzyme function by directed evolution. *Current Opinion in Structural Biology*, 13(4), pp.500–505.

Dalby, P. A. 2011. Strategy and success for the directed evolution of enzymes. *Current Opinion in Structural Biology*, 21(4), pp.473–480. Available at: http://dx.doi.org/10.1016/j.sbi.2011.05.003.

Daugherty, P. S., Chen, G., Iverson, B. L. and Georgiou, G. (2000) Proc Natl Acad Sci U S A 97(5):2029-34.

De Carvalho CC. 2011.Enzymatic and whole cell catalysis: finding new strategies for old processes. Biotechnol. Adv. 29:75–83. [PubMed: 20837129].

De Groeve MRM, de Baere M, Hoflack L, Desmet T, Vandamme EJ, Soetaert W. 2009. Creating lactose phosphorylase enzymes by directed evolution of cellobiose phosphorylase. Protein Eng Des Sel, 22:393-399.

De Jesus, A. J., Yin, H. 2017. Supramolecular Membrane Chemistry, Comprehensive Supramolecular Chemistry II, Elsevier, pahes 311-328, ISBN 9780128031995, https://doi.org/10.1016/B978-0-12-409547-2.12572-7.

De los Santos, E.L., Meyerowitz, J.T., Mayo, S.L., Murray, R.M. 2016. Engineering transcriptional regulator effector specificity using computational design and in vitro rapid prototyping: developing a vanillin sensor. ACS Synth. Biol. 5 (4), 287-295.

Demuynck C., J. Bolte, L. Hecquet, V. Dalmas. 1991. Tetrahedron Lett., 32, 5085-

5088.

Dietrich JA, McKee AE, Keasling JD. 2010. High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu Rev Biochem*. 79:563-590. doi:10.1146/annurev-biochem-062608-095938.

Doig, A. J., & Williams, D. H. 1991. Is the hydrophobic effect stabilizing or destabilizing in proteins?. The contribution of disulphide bonds to protein stability. *Journal of Molecular Biology*, *217*(2), 389–398. https://doi.org/10.1016/0022-2836(91)90551-G.

Doi N, Kumadaki S, Oishi Y, Matsumura N, Yanagawa H. 2004. In vitro selection of restriction endonucleases by in vitro compartmentalization. Nucleic Acids Res. 32(12): e95. [PubMed: 15247328].

Dragosits, M., Mattanovich, D. 2013. Adaptative laboratory evolution – principles and applications for abiotechnology. Microb. Cell Factories 12, 64.

Duetz, W. A. 2007. Microtiter plates as mini-bioreactors: miniaturization of fermentation methods. Trends Biotechnol. 15 (10), 469−475.

Du J, Shao Z, Zhao HM. 2011. Engineering microbial factories for synthesis of value-added products. J. Ind. Microbiol. Biotechnol. 38:873–890. [PubMed: 21526386].

Esakova, O., Meshalkina L. E. et al., 2004. Donor substrate regultation of transketolase. *Eur. J. Biochem.* 271, 4189-4194.

Falke, J. A. B. and J. J. 2010. Purification of Proteins Using Polyhistidine Affinity Tags. *Methods Enzymol.; 326: 245–254.*

Fengxue Xin, Weiliang Dong, D. Z., Jiang Yujia, W. Y., & Ziyao Lv, Yan Fang, M. J. 2019. *Reactors in Process Engineering* (Issue Current Developments in Biotechnology and Bioengineering. https://doi.org/10.1016/B978-0-444-64085-7.00009-5, pp. 207–232).

Ferla M. P. 2016. Mutanalyst, an online tool for assessing the mutational spectrum of epPCR libraries with poor sampling. *BMC Bioinformatics*. 17-152. DOI

10.1186/s12859-016-0996-7.

Fisher, A. K., Freedman, B. G., Bevan, D. R., & Senger, R. S. 2014. A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories. *CSBJ*, *11*(18), 91–99. https://doi.org/10.1016/j.csbj.2014.08.010.

Fu, Yun-Xin & Huai Haying. 2003. Estimating Mutation Rate; How to Count Mutations? Genetics Society of America. Genetics 164: 797-805.

Gary L, Foutch, Johannes Arland. 2017. Reactors in Process Engineering, Research Gate.

Gauri Singhal, Vartika Verma, S. S. B. and N. S. 2018. Fermentation Technology Prospecting on Bioreactors/Fermenters: Design and Types. In *Principles and Applications of Fermentation Technology* (pp. 65–83).

Ghadessy FJ, Ong JL, Holliger P. 2001. Directed evolution of polymerase function by compartmentalized self-replication. Proc Natl Acad Sci U S A. 98(8):4552–7. [PubMed: 11274352].

Gianella, P., Snapp, E., Levy, M. 2016. *of Bond-Forming Enzymes From Large Libraries*. *113*(8), 1647–1657.

Giepmans, B. N.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y. 2006. The fluorescent toolbox for assessing protein location and function. Science, 312 (5771), 217−224.

Gocke, Dörte. 2007. New and optimised thiamine diphosphate (ThDP)-dependent enzymes for carboligation- Creation of a toolbox for chiral 2-hydroxy ketones, PhD Thesis, Köln.

Griffiths AD, Tawfik DS. 2003. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. Embo J. 22(1):24–35. [PubMed: 12505981].

Grunwald, P. 2018.Biocatalysis Biochemical Fundamentalals and Applications, Second Edition, World Scientific.

G. Schneider, Y. Lindqvist. 1998. Crystallography and mutagenesis of transketolase: mechanistic implications for enzymatic thiamin catalysis, Biochim. Biophys. Acta 1385, 387–398.

Guo, W., Sheng, J., & Feng, X. 2017. Mini-review : In vitro Metabolic Engineering for Biomanufacturing of High-value Products. *Computational and Structural Biotechnology Journal*, *15*, 161–167. https://doi.org/10.1016/j.csbj.2017.01.006.

Guruprasad, L. 2019. Protein Structure. Resonance, 24(3), 327–338. https://doi.org/10.1007/s12045-019-0783-7.

Hailes, H.; A. Dalby, P.; J. Lye, G.; Baganz, F.; Micheletti, M.; Szita, N.; M. Ward, J. 2010. α,α'-dihydroxy ketones and 2-amino-1,3-diols: Synthetic and Process Strategies Using Biocatalysts. Current Organic Chemistry, Volume 14, Number 17, pp. 1883-1893.

Hanson-Manful P, Patrick WM. 2013. Construction and analysis of randomized protein-encoding libraries using error-prone PCR. Methods Mol Biol. 996:251–67.

Hecquet, L., Fessner, W. D., Hélaine, V., & Charmantray, F. 2014. New Applications of Transketolase: Cascade Reactions for Assay Development. *Cascade Biocatalysis: Integrating Stereoselective and Environmentally Friendly Reactions*, *9783527335220*, 315–338. https://doi.org/10.1002/9783527682492.ch15

Hecht, M. H., Sturtevant, J. M. & Sauer, R. T. 1986. Proteins: Struct. Funct. Genet. 1, 43-46.

He, Y.-C.; Ma, C.-L.; Xu, J.-H.; Zhou. 2011. L. A high-throughput screening strategy for nitrile-hydrolyzing enzymes based on ferric hydroxamate spectrophotometry. Appl. Environ. Microbiol. 89 (3), 817−823.

Hibbert, E.G. et al., 2007. Directed evolution of transketolase activity on non-phosphorylated substrates. *Journal of Biotechnology*, 131(4), pp.425–432.

Hibbert, T. Senussi, M. E. B. Smith, S. J. Costelloe, J. M. Ward, H. C. Hailes, P. A. Dalby, 2008. *J. Biotechnol. 134*, 240-245.

Hobbs G. R., M. D. Lilly, N. J. Turner, J. M. Ward, A. J. Willets, J. M. Woodley, ***J.*** *1993. Chem. Soc., Perkin Trans. 1*, 165-166.

Hokanson, C. a. et al., 2011. Engineering highly thermostable xylanase variants using an enhanced combinatorial library method. *Protein Engineering, Design and Selection*, 24(8), pp.597–605.

Hoyos, P. et al., 2010. Biocatalytic strategies for the asymmetric synthesis of alpha-hydroxy ketones. *Accounts of chemical research*, 43(2), pp.288–299.

Hudlicky T, Reed JW. 2009. Applications of biotransformations and biocatalysis to complexity generation in organic synthesis. Chem. Soc. Rev. 38:3117–3132. [PubMed: 19847346].

Hughes, J. Gregory., Lewis, J. C. 2018. Introduction: Biocatalysis in Industry. *Chemical Reviews*, *118*, 1–3. https://doi.org/10.1021/acs.chemrev.7b00741.

Hulse, J.H. 2004. Biotechnologies: Past, history, present state and future prospects. *Trends in Food Science & Technology*, 15(1), pp.3–18.

Hult, K., & Berglund, P. 2007. *Enzyme promiscuity: mechanism and applications*. *25*(5). https://doi.org/10.1016/j.tibtech.2007.03.002

Illanes, A. et al., 2012. Recent trends in biocatalysis engineering. *Bioresource Technology*, 115(July), pp.48–57.

Ishiyama, M. et al.1993. Chem. Pharm. Bull. 41(6), 1118-1122.

Jaenicke, R. & Bohm, G. 1998. The stability of proteins in extreme environments. Curr. Opin. Struct. Biol. 8, 738–748.

Jahromi, R.R.F. et al., 2011. Structural stability of *E. coli* transketolase to temperature and pH denaturation. *Journal of Biotechnology*, 155(2), pp.209–216. Available at: http://dx.doi.org/10.1016/j.jbiotec.2011.06.023.

Johannes, T.W., Simurdiak, M.R. & Zhao, H. 2006. Biocatalysis. *Encyclopedia of Chemical Processing*, pp.101–110.

Jo, H. J., Lee, J.-W., Noh, J.-S., Kong, K.-H. 2012. Site-directed mutagenesis of cysteine residues in Phi-class glutathione S-transferase F3 from Oryza sativa. Bull. Korean Chem. Soc. 33, 4169–4172.

Johnson B.H., Hecht M.H. 1994. Cells by repeated cycles of freezing and hawing. Biotechnology. 1994;12:1357. doi: 10.1038/nbt1294-1357.

Joo, E. J., Cress, B. F., & Koffas, M. A. G. 2014. Using Recombinant Microorganisms for the Synthesis and Modification of Flavonoids and Stilbenes. In *Polyphenols in Human Health and Disease*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-398456-2.00036-0.

Joo H, Arisawa A, Lin Z, Arnold F. H. 1999. A high-throughput digital • imaging screen for the discovery and directed evolution of oxygenases. Chem Biol, 6:699-706.

Kazlauskas, R J. 2005. Enhancing catalytic promiscuity for biocatalysis. Current Opinion in Chemical Biology 9. 195-201.

Kazlauskas, R J., Bornscheuer, U. T. 2012. Enzyme Catalytic Promiscuity: Expanding the Catalytic Action of Enzymes to New Reactions. In *Comprensive Chilarity Vol 7.* (Vol. 7, pp. 465–480). chttp://dx.doi.org/10.1016/B978-0-08-095167-6.00729-1.

Kelly, R.M., Leemhuis, H., Rozeboom, H.J., van Oosterwijk, N., Dijkstra, B.W., Dijkhuizen, L. 2008. Elimination of competing hydrolysis and coupling side reactions of a cyclodextrin glucanotransferase by directed evolution. Biochem. J. 413, 517–525.

Khersonsky, O. & Tawfik, D.S. 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual review of biochemistry*, 79, pp.471–505.

Kiely M. E., E. L. Tan, T. Wood. 1969. *Can. J. Biochem. 47*, 455-460.

Kim, Y. S.; Jung, H. C.; Pan, J. G. 2000. Bacterial cell surface display of an enzyme library for selective screening of improved cellulase variants. Appl. Environ. Microbiol. 66 (2), 788−793.

Kirk, O., Borchert, T.V. & Fuglsang, C.C. 2002. Industrial enzyme applications. *Current Opinion in Biotechnology*, 13(4), pp.345–351.

Kirschner, A., and Bornscheuer, U. T. 2008. Directed evolution of a Baeyer-Villiger monooxygenase to enhance enantioselectivity. *Appl. Microbiol. Biotechnol.* 81, 465-472.

Kochetov, G. A., & Solovjeva, O. N. 2014. Structure and functioning mechanism of transketolase. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1844*(9), 1608–1618. https://doi.org/10.1016/j.bbapap.2014.06.003

Kone, F. M. T.; Le Bechec, M.; Sine, J.-P.; Dion, M.; Tellier, C. 2009. Digital screening methodology for the directed evolution of transglycosidases. Protein Eng., Des. Sel. 22 (1), 37−44.

Korendovych, I. V. 2018. HHS Public Access. *Methods Mol Biol.*, *1658*, 15–23. https://doi.org/10.1007/978-1-4939-7366-8

Kronberg, B. 1991. The hydrophobic effect. *Current Opinion in Colloid and Interface Science*, *22*, 14–22. https://doi.org/10.1016/j.cocis.2016.02.001

Kronberg, B. (2016). Current Opinion in Colloid & Interface Science The hydrophobic effect. Current Opinion in Colloid & Interface Science, 22, 14–22. https://doi.org/10.1016/j.cocis.2016.02.001.

Kühne, W. 1876. Über das Verhalten verschiedener organisirter und sog. ungeformter Fermente.- Über das Trypsin (Enzym des Pankreas). FEBS Lett., 62, 3-7.

L. Mitschke, Ch. Parthier, K. Schröder-Tittmann, J. Coy, S. Lüdtke, K. Tittmann. 2010. The crystal structure of human transketolase and new insights into its mode of action, J. Biol. Chem. 285, 31559–31570.

Leemhuis, H.; Kelly, R. M.; Dijkhuizen, L. 2009. Directed evolution of enzymes: Library screening strategies. IUBMB Life, 61 (3), 222− 228.

Lessard, J. C. 2013. Growth media for *E. coli.* In *Methods in Enzymology* (1st ed., Vol. 533). Elsevier Inc. https://doi.org/10.1016/B978-0-12-420067-8.00011-8.

Lee, S. Y.; Choi, J. H.; Xu, Z. 2003. Microbial cell-surface display. Trends Biotechnol. 21 (1), 45−52.

Leung, D. W., Chen, E. & Goeddel, D. V. 1989. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Technique 1, 11–15.

Lin, J., Wagner, J. M., & Alper, H. S. 2017. Enabling tools for high-throughput detection of metabolites: Metabolic engineering and directed evolution applications. *Biotechnology Advances*, *35*(8), 950–970. https://doi.org/10.1016/j.biotechadv.2017.07.005

Lindqvist Y, Schneider G, Ermler U, Sundstrom M. 1992. 3-Dimensional structure of transketolase, a thiamine diphosphate dependent enzyme, at 25 angstrom resolution. EMBO J 11:2373–2379.

Lindskog, E. K. 2018. The Upstream Process: Principal Modes of Operation. In *Biopharmaceutical Processing*. Elsevier Ltd. https://doi.org/10.1016/B978-0-08-100623-8.00031-1.

Li, S., Yang, X., Yang, S., Zhu, M., & Wang, X. 2012. Technology Prospecting on Enzymes: Application, Marketing and Engineering, Computational and Structural Biotechnology Journal, Volume No:2, Issue:3, September 2012.

López-Llano, J. et al., 2006. α-Helix Stabilization by Alanine Relative to Glycine: Roles of Polar and Apolar Solvent Exposures and of Backbone Entropy. PROTEINS: Structure, and Bioinformatics 64:769-788.

Luli, G. W., & Strohl, W. R. 1990. Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations. *Applied and Environmental Microbiology*, *56*(4), 1004–1011. https://doi.org/10.1128/aem.56.4.1004-1011.1990.

Lutz, S. 2010. Beyond directed evolution-semi-rational protein engineering and design. *Current Opinion in Biotechnology*, 21(6), pp.734–743. Available at: http://dx.doi.org/10.1016/j.copbio.2010.08.011.

Ma, S.K., Gruber, J., Davis, C., Newman, L., Gray, D., Wang, A., Grate, J., Huisman, G. W., Sheldon, R.A. 2010. A green-by-design biocatalytic process for

atorvastatin intermediate. Green Chem. 12, 81–86.

Mack, M.; Burger, M.; Pietschmann, P.; Hock, B. 2008. A highthroughput microtiter plate-based screening method for the detection of full-length recombinant proteins. Protein Expression Purif. 61 (1), 92−98.

Madhavan, A., Sindhu, R., Binod, P., Sukumaran, R. K., & Pandey, A. 2017. Strategies for design of improved biocatalysts for industrial applications. *Bioresource Technology*, *245*, 1304–1313. https://doi.org/10.1016/j.biortech.2017.05.031.

Magnusson, A.O. et al., 2005. An S-selective lipase was created by rational redesign and the enantioselectivity increased with temperature. Angew. Chem. Int. Ed. 44, 4582–4585.

Mahr, R., Frunzke, J., 2016. Transcription factor-based biosensor in biotechnology: current state and Future prospects. Appl. Microbiol. 100, 79-90.

Masri S. W., M. Ali, C. J. Gubler. 1989. *Comp. Biochem. Phys. B* 1988, *90*, 167-172. Mocali A., F. Paoletti, *Eur. J. Biochem. 180*, 213-219.

Mastrobattista E, Taly V, Chanudet E, Treacy P, Kelly BT, Griffiths AD. 2005. High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. Chem Biol. 12(12):1291–300. [PubMed: 16356846].

Matthews, B. W. et al., 1987. Enhanced protein thermostability from site-directed mutations that decreased the entropy unfolding. Proc. Natl. Acad. Sci. USA. Vol. 84. Pp. 6663-6667. Biochemistry.

Meshalkina L., Nilsson U., et al., 1997. Examination of the thiamin diphosphate binding site in yeast transketolase by site-directed mutagenesis. *Eur. J. Biochem.* 244, 646-652.

Meng Wang, T. S. and H. Z. 2012. Biocatalyst Development by Direct Evolution. *Bioresour Technol. National Institutes of Health*, *23*(1), 1–17. https://doi.org/10.1038/jid.2014.371.

Miyasaki, K., and Takenouchi, M. 2002. Megawhop cloning: A method for creating

random mutagenesis libraries by Megaprimer PCR of Whole Plasmids. *Biotechniques* 33 (1033-1034), 1036-1038.

Miyazaki K, T. M. 2002. METHODS Short Technical Reports. *Biotechniques*, *1038*(November), 33(5): 1033-4, 1036–1038.

Miyasaki Kentaro. 2011. MEGAWHOP cloning: A method of creating random mutagenesis libraries via Megaprimer PCR of Whole Plasmids. Methods in Enzymology, Volume 498.

Miyazono, Y., Shiga, M., Sasamoto, K., Examiner, P., & Richter, J. 2000. *United States Patent (19)*. *19*.

Mohamed, M. F., & Hollfelder, F. 2013. *Efficient, crosswise catalytic promiscuity among enzymes that catalyze phosphoryl transfer*. *1834*, 417–424.

Molitor, R, et al. 2019. Agar plate-based screening methods for the identification of polyester hydrolysis by *Pseudomonas* species. Microbial Biotechnolog, **https://doi.org/10.1111/1751-7915.13418**

Morley KL, Kazlauskas R. J. 2005. Improving enzyme properties: when are closer mutations better? Trends Biotechnol, 23:231-237.

Müller, M., Gocke, D. & Pohl, M., 2009. Thiamin diphosphate in biological chemistry: Exploitation of diverse thiamin diphosphate-dependent enzymes for asymmetric chemoenzymatic synthesis. *FEBS Journal*, 276(11), pp.2894–2904.

Muller M, Sprenger Ga, Pohl M. *2013. CC bond formation using ThDP- dependent lyases. Curr Opin Chem Biol; 17:261–270.*

Nair N, Zhao H. 2008. Evolution in reverse: engineering a D-xylosespecific xylose reductase. ChemBioChem, 9:1213-1215.

Neuberg, C. & Karczag, L. 1911. Über zuckerfreie Hefegärungen. IV. Carboxylase, ein neues Enzym der Hefe. Biochem. Zeitschr., 36, 68.

Nikkola M., Lindqvist, Y. & Scheneider, G. 1994. Refined structure of transketolase from Saccharomyces cerevisiae at 2.0 Å resolution. J. Mol. Biol. 238, 387-404.

O'Brien, T.E., Bertolani, S.J., Zhang, Y., Siegel, J.B., Tantillo, D.J. 2018. Predicting productive binding modes for substrates and carbocation intermediates in terpene synthases—bornyl diphosphate synthase as a representative case. ACS Catal. 8, 3322–3330. https://doi.org/10.1021/acscatal.8b00342.

Ole Kirk, Torben Vedel Borchert and Claus Crone Fuglsang. 2002. Industrial enzyme applications, Protein technologies and commercial enzymes, Current Opinion in Biotechnology 2002, 13:345-351.

Olsen, M., Iverson, B., & Georgiou, G. 2000. *High-throughput screening of enzyme libraries*. 331–337.

Packer M S, Liu R.D. 2015. Methods for direct evolution of proteins, Nature Reviews Genetics, Volume 16.

Packer, M. S., & Liu, D. R. 2015. *Nature Publishing Group*, *16*(July). https://doi.org/10.1038/nrg3927.

Paoletti F, D. Aldinucci. 1986. *Arch. Biochem. Biophys. 245*, 212-219.

Park S, Morley KL, Horsman GP, Holmquist M, Hult K, Kazlauskas R. J. 2005. Focusing mutations into the P. fluorescens esterase binding site increases enantioselectivity more effectively than distant mutations. Chem Biol, 12:45-54.

P. Asztalos, C. Parthier, R. Golbik, M. Kleinschmidt, G. Hübner, M.S. Weiss, R. Friedemann, G. Wille, K. Tittmann. 2007. Strain and near attack conformers in enzymic thiamin catalysis: X-ray crystallographic snapshots of bacterial transketolase in covalent complexwith donor ketoses xylulose 5-phosphate and fructose 6-phosphate, and in noncovalent complexwith acceptor aldose ribose 5-phosphate, Biochemistry 46 12037–12052.

Patel R. N. 2011. Biocatalysis: synthesis of key intermediates for development of pharmaceuticals. ACS Cata. 1:1056–1074.

Pavia, Donald L., Gary M. Lampman, George S. Kritz, Randall G. Engel. 2006. *Introduction to Organic Laboratory Techniques (4th Ed.)*. Thomson Brooks/Cole.

Pham, P. V. 2018. Medical biotechnology: Techniques and applications. *Omics Technologies and Bio-Engineering: Towards Improving Quality of Life*, *1*, 449–469. https://doi.org/10.1016/B978-0-12-804659-3.00019-1.

Pino, M. S., Rodríguez-Jasso, R. M., Michelin, M., Flores-Gallegos, A. C., Morales-Rodriguez, R., Teixeira, J. A., & Ruiz, H. A. 2018. Bioreactor design for enzymatic hydrolysis of biomass under the biorefinery concept. *Chemical Engineering Journal*, *347*(April), 119–136. https://doi.org/10.1016/j.cej.2018.04.057.

Priscilla E.M. Purnick & Ron Weiss. 2009. The second wave of synthetic biology: from modules to systems, Nature Reviews Molecular Cell Biology, 10, 410-422.

Philippov P, I. K. Shestakova, N. K. Tikhomirova, G. A. Kochetov. 1980. *BBA-Enzymol. 613*, 359-369.

Ran, N. et al., 2008. Recent applications of biocatalysis in developing green chemistry for chemical synthesis at the industrial scale. *Green Chemistry*, 10(4), p.361.

Ranoux A. et al., 2012. Enhancement of the Substrate Scope of Transketolase, ChemBiochem, 13, 1921-1931. DOI:10.1002/cbic.201200240.

Rayment I. 2003. Encyclopedia of Physical Science and Technology, Reference Module in Chemistry, Molecular Sciences and Chemical Engineering, Academic Press, 191-218. https://doi.org/10.1016/B0-12-227410-5/00616-5.

Reetz, M.T.; Carballeira, J.D. 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nat. Protoc. 2, 891–903.

Reetz M. T. 2009. Directed evolution of enantioselective enzymes: an unconventional approach to asymmetric catalysis in organic chemistry. J. Org. Chem. 2009; 74:5767–5778. [PubMed: 20560561].

Reetz, M. T. et al., 2010. Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. *Chemical communications (Cambridge, England)*, 46(45), pp.8657–8658.

Reetz, M.T., Wang, L.-W. & Bocola, M. 2006. Directed evolution of enantioselective

enzymes: Iterative cycles of Casting for probing protein-sequence space. Angew. Chem. 118, 1258–1263 Erratum 2556.

Rice, A. J., Truong, L., Johnson, M. E., & Lee, H. 2013. A colorimetric assay optimization for high-throughput screening of dihydroorotase by detecting ureido groups. *Analytical Biochemistry*, *441*(1), 87–94. https://doi.org/10.1016/j.ab.2013.05.035.

Riet Klass, Tramper Johannes. 1991. Basic Bioreactor Design, Wageningen Agricultural University, Marcek Dekker.

Robinson, P. K. 2015. *Enzymes: principles and biotechnological applications*. 1–41. https://doi.org/10.1042/BSE0590001.

Roche. 2007. Cell Proliferation Reagent WST-1. *Roche Diagnostics Cat.*, *1*(11), 1–4. https://doi.org/10.1016/j.bbrc.2008.07.050.

Rubin-Pitel SB, Zhao HM. 2006. Recent advances in biocatalysis by directed enzyme evolution. Comb. Chem. High Throughput Screen. 9:247–257. [PubMed: 16724916].

Sakkos, J. K., Wackett, L. P., & Aksan, A. 2019. Enhancement of biocatalyst activity and protection against stressors using a microbial exoskeleton. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-40113-8.

Saambrook, J., and Rusell, D. W. 2001. Molecular Cloning: A Laboratory Manual. 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanchez, S. & Demain, A. L. 2011. Enzymes and bioconversions of industrial, pharmaceutical, and biotechnological significance. *Organic Process Research and Development*, 15(1), pp.224–230.

Schoemaker HE, Mink D, Wubbolts MG. 2003. Dispelling the myths--biocatalysis in industrial synthesis. Science. 299:1694–1697. [PubMed: 12637735].

Scott. K. A. et al., 2006. Conformational entropy of alanine versus glycine in protein denaturated states. PNAS. Vol. 104. No. 8. 2661-2666. www.pnas.orgcgidoi10.1073pnas.0611182104

Scudiero, D. A. et al., 1988. Evaluation of a Soluble Tetrazolium / Formazan Assay for Cell Growth and Drug Sensitivity in Culture Using Human and Other Tumor Cell Lines Evaluation of a Soluble Tetrazolium / Formazan Assay for Cell Growth and Drug Sensitivity in Culture Using Human. pp.4827–4833.

Seebeck, F.P. and Hilvert, D. 2003. Conversion of a PLP-dependent racemase into an aldolase by a single active site mutation. J. Am. Chem. Soc. 125, 10158–10159.

Sheh-Yi Sheu, Dah-Yen Yang, H. L. Selzle, E. W. Schlag, 2003. Energetics of hydrogen bonds in peptides, Proceedings of the National Academy of Sciences, 2003, 100 (22) 12683-12687; DOI: 10.1073/pnas.2133366100

Sheryl B. Rubin-Pitel, CatherineM-H. et al., 2007. Chapter 3 Direct Evolution Tools in Bioproduct and Bioprocess Development, Bioprocessing for Value-Added Products from Renewable Resources, pages 49-72.

Shim, J.H., Kim, Y.W., Kim, T.J., Chae, H.Y., Park, J.H., Cha, H., Kim, J.W., Kim, Y.R., Schaefer, T., Spendler, T., Moon, T.W., Park, K.H. 2004. Improvement of cyclodextrin glucanotransferase as an antistaling enzyme by error prone PCR. Protein Eng. Des. Sel. 17, 205–211.

Shukla, V. B. & Kulkarni, P. R. 2000. L-Phenylacetylcarbinol (L-PAC): biosynthesis and industrial applications. World J. Microbiol. Biotechnol., 16, 499-506.

Sillitoe I, et al., 2015. The history of the CATH structural classification of protein domains. Biochimie 119, 209-217.

Siloto Rodrigo M.P. Randal J. Weselake. 2012. Site saturation mutagenesis:Methods and applications in protein engineering, Biocatalysis and Agricultural Biotechnology 1, 181 - 189.

Singh, J., Kaushik, N., & Biswas, S. 2014. *Bioreactors – Technology & Design Analysis*. *01*(06).

S.J. Costelloe, J.M. Ward, P.A. Dalby. 2008. Evolutionary analysis of the TPP-dependent enzyme family, J. Mol. Evol. 66, 36–49.

Smith, M.E.B., Kaulmann, U., et al., 2006. A colorimetric assay for screening transketolase activity. *Bioorganic and Medicinal Chemistry*, 14(20), pp.7062–7065.

Smith, M.E.B. et al., 2008. Enhancing and reversing the stereoselectivity of Escherichia coli transketolase via single-point mutations. *Advanced Synthesis and Catalysis*, 350(16), pp.2631–2638.

Smith, M.E.B., Smithies, K., et al., 2006. The first mimetic of the transketolase reaction. *European Journal of Organic Chemistry*, (5), pp.1121–1123.

Speltz, E. B. & Regan Lynne. 2013. White and green screening with circular polymerase extension cloning for easy and reliable cloning. The Protein Society. DOI: 10.1002/pro.2268.

Sprenger, G. a. & Pohl, M. 1999. Synthetic potential of thiamine diphosphate-dependent enzymes. *Journal of Molecular Catalysis - B Enzymatic*, 6(3), pp.145–159.

Steele, H.L. et al., 2008. Advances in recovery of novel biocatalysts from metagenomes. *Journal of Molecular Microbiology and Biotechnology*, 16 (1-2), pp.25–37.

Stemmer P. 1994. DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution, Genetics, Vol. 91, pp. 10747-10751.

Stemmer P. 1994. Rapid evolution of a protein in vitro by DNA shuffling, Nature 370, 389- 391.

Strafford, J. et al., 2012. Directed evolution to re-adapt a co-evolved network within an enzyme. *Journal of Biotechnology*, 157(1), pp.237–245. Available at: http://dx.doi.org/10.1016/j.jbiotec.

Strausberg SL, Ruan B, Fisher KE, Alexander PA, Bryan PN. 2005. Directed coevolution of stability and catalytic activity in calcium-free subtilisin. Biochemistry, 44:3272- 3279.

Takahashi Masato and Hideki Aoyagi. 2020. Analysis and effect of conventional flasks in shakin culture of *Escherichia coli.* AMB Expr. 10:77.

Torres, S. et al., 2009. An organic-solvent-tolerant esterase from thermophilic Bacillus licheniformis S-86. *Bioresource Technology*, 100(2), pp.896–902. Available at:

http://dx.doi.org/10.1016/j.biortech.2008.07.009.

Torres, S., Pandey, A. & Castro, G.R. 2011. Organic solvent adaptation of Gram positive bacteria: Applications and biotechnological potentials. *Biotechnology Advances*, 29(4), pp.442–452. Available at: http://dx.doi.org/10.1016/j.biotechadv.2011.04.002.

Ukkonnen K. et al., 2011. High-yield production of biologically active recombinant protein in shake flask culture by combination of enzyme-based glucose delivery and increased oxygen transfer. Microbial Cell Factories. 10:107.

Valdevit, A. 2018. Bioreactors: System design and application for regenerative engineering. In *Encyclopedia of Biomedical Engineering* (Vols. 1–3). Elsevier. https://doi.org/10.1016/B978-0-12-801238-3.64124-7.

Valetti, F., & Gilardi, G. 2013. Improvement of biocatalysts for industrial and environmental purposes by saturation mutagenesis. *Biomolecules*, *3*(4), 778–811. https://doi.org/10.3390/biom3040778.

Valle, A., Cantero, D., & Bolívar, J. 2019. Metabolic engineering for the optimization of hydrogen production in Escherichia coli : A review. *Biotechnology Advances*, *37*(5), 616–633. https://doi.org/10.1016/j.biotechadv.2019.03.006.

Vartanian, J. P., Henry, M. and Wain-Hobson, S. 1996. Nucleic Acids Res 24(14):2627-31.

Voet, Donald; Judith Voet; Charlotte Pratt. 2008. *Fundamentals of Biochemistry*. John Wiley & Sons Inc.

Vogel, C., & Pleiss, J. 2014. The modular structure of ThDP-dependent enzymes. *Proteins: Structure, Function and Bioinformatics*, *82*(10), 2523–2537. https://doi.org/10.1002/prot.24615.

Voulgaris, I., Chatel, A., Hoare, M., Finka, G., & Uden, M. 2016. Evaluation of options for harvest of a recombinant *E. Coli* fermentation producing a domain antibody using ultra scale-down techniques and pilot-scale verification. *Biotechnology Progress*, *32*(2), 382–392. https://doi.org/10.1002/btpr.2220.

Wahler, D., & Reymond, J. 2001. *High-throughput screening for biocatalysts*. 535–

544.

Wang, Y.H., Xie, H., Zhou, J., Zhang, F., Wu, R. 2017. Substrate folding modes in trichodiene synthase: a determinant of chemo- and stereoselectivity. ACS Catal. 7, 5841–5846. https://doi.org/10.1021/acscatal.7b01462.

Weis, R.; Luiten, R.; Skranc, W.; Schwab, H.; Wubbolts, M.; Glieder, A. 2004. Reliable high-throughput screening with Pichia pastoris by limiting yeast cell death phenomena. FEMS Yeast Res. 5 (2), 179−189.

Wen F, Nair NU, Zhao HM. 2009. Protein engineering in designing tailored enzymes and microorganisms for biofuels production. Curr. Opin. Biotechnol. 20:412–419. [PubMed: 19660930].

Wiggelsworth Mark. 2020. *High-thoughput screening (HTS),* https://www.bmglabtech.com/high-throughput-screening/.

Woodley, J.M. 2008. New opportunities for biocatalysis: making pharmaceutical processes greener. *Trends in Biotechnology*, 26(6), pp.321–327.

Xiao, H., Bao, Z., & Zhao, H. 2015. High throughput screening and selection methods for directed enzyme evolution. *Industrial and Engineering Chemistry Research*, *54*(16), 4011–4020. https://doi.org/10.1021/ie503060a.

Yadav, D., Tanveer, A., Malviya, N., & Yadav, S. 2016. Overview and Principles of Bioengineering : The Drivers of Omics Technologies. In *Omics Technologies and Bio-engineering Volume 1*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-804659-3.00001-4.

Yang, C., Xia, Y., Qu, H., Li, A., Liu, R., Wang, Y., Zhang, T. 2016. Discovery of new cellulases from the metagenome by a metagenomics-guided strategy. Biotechnol. Biofuels 9, 138.

Yang, G.; Withers, S. G. 2009. Ultrahigh-throughput FACS-based screening for directed enzyme evolution. ChemBioChem, 10 (17), 2704−2715.

Yang, H., Liu, L., Shin, H.D., Chen, R.R., Li, J., Du, G. 2013. Structure- based engineering of histidine residues in the catalytic domain of alpha-amylase From Bacillus subtilis for improved protein stability and catalytic efficiency under

acidic conditions. J. Biotechnol. 164, 59–66.

Yaoi, K., Kondo, H., et al., 2007. The structural basis for the exo-mode of action in GH74 oligoxyglucan reducing end-specific cellobionhydeolase. *J. Mol. Biol.* 370, 53-62.

Yi, L.; Gebhard, M. C.; Li, Q.; Taft, J. M.; Georgiou, G.; Iverson, B. L. 2013. Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. Proc. Natl. Acad. Sci. U. S. A. 110 (18), 7229−7234.

Yin, L. M., Wei, Y., Wang, Y., Xu, Y. D., & Yang, Y. Q. 2012. Long term and standard incubations of WST-1 reagent reflect the same inhibitory trend of cell viability in rat airway smooth muscle cells. *International Journal of Medical Sciences*, *10*(1), 68–72. https://doi.org/10.7150/ijms.5256.

Y. Lindqvist, G. Schneider, V. Erwler, M. Sundström. 1992. Three-dimensional structure of apotransketolase. Flexible loops at the active site enable cofactor binding, EMBO J. 11, 2373–2379.

You C. et al., 2012. Easy preparation of a large-size random gene mutagenesis library in *Escherichia coli*. Analytical Biochemistry. YABIO 10976. http://dx.doi.org/10.1016/j.ab.2012.05.022.

Yu Haoran, Dalby P. A. et al., 2017. Two strategies to engineer flexible loops for improved enzyme thermostability, Scientific Reports, 7:41212.

Yu, H., & Dalby, P. A. 2018. *Coupled molecular dynamics mediate long- and short-range epistasis between mutations that affect stability and aggregation kinetics*. *115*(47).

Yu, H., Hernández López R. I. et al., 2019. Engineering transketolase to accept both unnatural donor and acceptor substrates and produce a-hydroxyketones. The FEBS Journal 287, 1758–1776.

Zhao HM. 2011. Highlights of biocatalysis and biomimetic catalysis. ACS Catal. 1:1119–1120.

Zhao J, Kardashliev T, Joëlle Ruff A, Bocola M, Schwaneberg U. 2014. Lessons from diversity of directed evolution experiments by an analysis of 3,000 mutations. Biotechnol Bioeng. 111:2380–9

Zhou, N. E., Kay, C. M., & Hodges, R. S. 1993. Disulfide Bond Contribution to Protein Stability: Positional Effects of Substitution in the Hydrophobic Core of the Two-Stranded α-Helical Coiled-Coil. *Biochemistry*, *32*(12), 3178–3187. https://doi.org/10.1021/bi00063a033.