Novel strategies for automated engineering of enzymatic systems: Structural and functional insights to human pyruvate dehydrogenase complex

Vom Promotionsausschuss der
Technischen Universität Hamburg

zur Erlangung des akademischen Grades

Doktor-Ingenieurin (Dr.-Ing.)

genehmigte Dissertation

von
Sibel Ilhan

aus
Aalen

2020

Erster Gutachter:	Prof. Dr. An-Ping Zeng
Zweiter Gutachter:	PD Dr. habil. Monika Johannsen
Tag der mündlichen Prüfung:	16.11.2020

Acknowledgment

Throughout the writing of this dissertation I have received a great deal of support and assistance.

First I would to like to express my gratitude to my academic supervisor, Professor An-Ping Zeng, for accepting me at his research group and offering me the opportunity to perform my doctoral thesis. Your expertise and feedback was invaluable in formulating the research questions and methodology. I would also like to thank PD Dr. habil. Monika Johannsen for her time and effort as reviewer of this research.

I would like to gratefully acknowledge the financial support provided by the German Ministry for Education and Research (BMBF) in the scope of the funding programme "Biotechnologie 2020+".

My sincere thanks also goes to my fellow colleagues and students from the Institute of Bioprocess and Biosystems Engineering for their wonderful collaboration, the many cheering moments and support in every stage of this study. You made these years memorable to me. I would particularly like to single out my project supervisor Dr. Uwe Jandt, thank you for your valuable guidance, many encouraging words and for providing me the tools along the way on this journey.

In addition, I would like to thank my family and Alexander Mook for their unconditional love, constant encouragement and stimulating discussions. I am forever grateful for your patience and understanding. Finally, I could not have completed this dissertation without the support of my dearest friends, who provided happy distractions to rest my mind outside of my research.

Identifier

DOI: https://doi.org/10.15480/882.3290
Handle: http://hdl.handle.net/11420/8790
Other: urn:nbn:de:gbv:830-882.0125104

Abstract

Subject of investigation in this study is the establishment of theoretical and practical foundations to enable an automated optimization of enzymes and enzymatic complexes in a new and highly efficient type of targeted and feedback guided optimization approach. These include new insights on structural and functional dynamics of a multienzymatic role model, the human pyruvate dehydrogenase complex (hPDC), as well as newly developed methods and a connected process chain to reliably and automatically determine enzymatic fitness. Such a feedback guided approach comprises fully automatable gene generation, protein expression and robust analysis without the need of any purification step and within shortest possible time.
hPDC demonstrates the typical advantages of a coupled multienzymatic system, like higher conversion efficiency through channeling of intermediates and recycling of cofactors. The modular constitution and central role in the metabolic pathway make hPDC an interesting enzyme complex for targeted protein manipulation aimed at enhanced bioprocess efficiency. The catalytic hPDC core is partitioned into N-terminal lipoyl domains for intermediate channeling, the binding domains for non-covalent subcomponent interaction and C-terminal catalytic domains for the self-assembly of the core. A cross-influence of the lipoyl domains of the hPDC core components (E2 and E3BP) to the peripheral subunit (E1 or E3) binding was examined via dynamic protein to protein interaction studies and revealed an impact on binding kinetics by a factor of two to three. This is significant, however much smaller than the influence of the binding domains of the core components itself. E3BP is influencing the E1 to E2 subcomplex interaction, whereas E2 shows no impact on the E3BP to E3 subcomplex. For both subcomplexes, the binding still occurred with high affinity and specificity, whereas the overall hPDC functionality was impaired. These findings on mutual influences within the hPDC will be useful for targeted enzyme manipulation in the future. Beside these discoveries, the mechanism for the flawless functionality of the whole complex depends on an interaction of all components. As reported by previous hPDC reaction modeling, nonlinear dynamics occur under certain conditions. This theory was investigated experimentally in this study by setting

up a multi-parameter screening of variable hPDC subcomponent ratio and substrate concentration in a batchwise *in vitro* hPDC assay. The detailed and fast analysis was enabled by the automatization of the multi-parametric screening approach. A transient oscillation occurring during overall hPDC reaction was found for physiological pyruvate concentrations of about 0.4 mM and lower. In addition, this transient oscillation was elongated if the E1 subcomponent was underrepresented in the hPDC, influencing its dynamics decisively. In order to fully elucidate the underlying mechanism, further investigation approaches are proposed.
Finally, to enable further experimental modulation and validation of hPDC, a fast, flexible and automated targeted enzyme optimization approach was aimed at. A mutant library for the role model green fluorescent protein (GFP) was constructed by rational design. To achieve genetical diversity, a two-step assembly PCR method was established, resulting in linear expression templates (LET). They are produced in short time range and can be manipulated easily to serve as DNA template in cell free protein synthesis (CFPS) without further purification steps or any other manual interaction. Subsequently, a TX-TL *E.coli* lysate was produced to express GFP and its variants in an open environment, directly monitored in the crude lysate by a newly developed, robust and highly sensitive flow cytometry based method. Lastly, the whole process chain was fully automatized at a liquid handling station and can process up to six targeted variants in less than six hours. To this end, further parallelization will lead to higher turnover efficiency.

Zusammenfassung

Gegenstand der Untersuchung in dieser Studie ist die Etablierung theoretischer und praktischer Grundlagen um eine zielgerichtete und automatisierte Optimierung von Enzymen und enzymatischen Komplexen durch einen neuen und hocheffizienten Ansatz zu ermöglichen. Dazu gehören neue Erkenntnisse über die strukturelle und funktionelle Dynamik eines multienzymatischen Vorbildes, des humanen Pyruvatdehydorgenase-Komplexes (hPDC), sowie neu entwickelte Methoden und eine durchgehend automatisierte Prozesskette zur zuverlässigen Bestimmung der enzymatischen Fitness. Ein solcher Feedback gesteuerter Ansatz umfasst die vollautomatische DNA Gengenerierung, Proteinexpression und robuste Analyse ohne jeglichen Aufreinigungsschritt innerhalb kürzester Zeit.
hPDC weist gegenüber einzelenen Enzymreaktionen Vorteile auf, wie z.B. eine höhere Umwandlungseffizienz durch Kanalisierung von Zwischenprodukten und Recycling von Cofaktoren. Der modulare Komplex Aufbau und seine zentrale Rolle im Stoffwechsel machen hPDC zu einem interessanten Enzymkomplex für die gezielte Proteinmanipulation um die Effizienz von Bioprozessen zu steigern. Der katalytische hPDC-Kern ist unterteilt in N-terminale Lipoyldomänen für die Weitergabe des Zwischenprodukts, die Bindungsdomänen für die nicht kovalente Interaktion der Subkomponenten sowie die C-terminalen katalytischen Domänen für die Selbstassemblierung des Kerns. Ein möglicher wechselseitiger Einfluss der Lipoyldomänen beider Kernkomponenten (E2 und E3BP) auf die Bindung der Kern umgebenden Untereinheit (E1 und E3) wurde über dynamische Protein zu Protein Interaktionstudien untersucht und zeigte einen Einfluss auf die Bindungskinetik um den Faktor zwei bis drei. Dies ist signifikant, aber deutlich kleiner als der Einfluss der Bindungsdomänen der Kernkomponenten selbst. Die Kernkomponente E3 beeinflusst den E1 zu E2 Teilkomplex, wohingegen E2 keinen Einfluss auf den E3BP zu E3 Teilkomplex aufweist. In beiden Fällen erfolgte die Protein Interaktion weiterhin mit hoher Affinität und Spezifität, wohingegen die hPDC-Funktionalität beeinträchtigt war. Diese Erkenntnisse über wechselseitige Einflüsse innerhalb des hPDC wird in Zukunft einer gezielten Enzymmanipulation dienlich sein.
Ein reibungsloses Zusammenspiel aller Komponenten im Komplex ist unab-

dingbar für seine Aktivität. Um erstmalig experimentelles Wissen über eine nichtlineare hPDC-Funktion unter verschiedenen Bedingungen zu erhalten, wurde ein Mehrparameter-Screening durchgeführt. Die Substratkonzentration als auch die Verhältnisse der verschieden Enzymuntereinheiten wurden variiert, welches durch die Automatisierung des Assayaufbaus vereinfacht wurde. Wie bereits theoretisch durch die Modellierung der hPDC Reaktion in der Literatur gemutmaßt wurde, konnte in dieser Arbeit eine Nichtlinearität in einer *in vitro* hPDC Reaktion im Batch-Modus festgestellt werden. Hierbei handelt es sich um eine transiente Oszillation um physiologische Zell-Pyruvatkonzentrationen von 0,4 mM und niedriger. Darüber hinaus konnte die Periodenanzahl der Schwingung erhöht werden, wenn die E1-Subkomponente im hPDC unterrepräsentiert war, welcher einen maßgeblichen Einfluss auf die Dynamik des Komplexes hat. Um den zugrunde liegenden Mechanismus vollständig aufzuklären, kann eine weitere Untersuchung erfolgen.
Um eine weitere experimentelle Modulation und Validierung von hPDC zu ermöglichen, wird schließlich ein schneller, flexibler und automatisierter Ansatz zur gezielten Enzymoptimierung angestrebt. Eine nach rationalem Design gestaltete Mutantenbibliothek für das einfache und gut untersuchte GFP wurde erstellt. Um eine genetische Vielfalt zu ermöglichenn, wurde ein zweistufiges Assembly-PCR-Verfahren etabliert, die zu linearen Expressionsvorlagen (LET) führt. Sie werden in kurzer Zeit hergestellt und können ohne weitere Aufreinigungsschritte oder andere manuelle Eingriffe leicht manipuliert werden, um als DNA-Vorlage in der Zellfreien Proteinsynthese (CFPS) zu dienen. Hierzu wurde ein TX-TL *E.coli* Lysat hergestellt, um GFP und seine Varianten in einer offenen Reaktionsumgebung zu exprimieren. Dies ermöglicht eine anschließende direkte Detektion der Zielproteine im Rohlysat durch ein neu entwickeltes, robustes und hochempfindliches Durchflusszytometrie-Verfahren. Die gesamte Prozesskette wurde letztendlich an einem Liquid-Handling-System vollautomatisiert und kann bis zu sechs gezielte Varianten in weniger als sechs Stunden verarbeiten. Weitere Parallelisierung der Prozesskette kann in Zukunft zu höheren Umsätzen führen.

Contents

Abbreviations

aa	Amino acid
Amp	Ampicillin
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BLI	Bio-layer interferometry
bp	Basepair
BD	Binding domain
BSA	Bovine serum albumin
CF	Cell free
CFP	Cyan fluorescent protein
CFPS	Cell free protein synthesis
CL	Cell lysate
CoA	Coenzyme A
C-terminal	Carboxyl-terminus
cv	Column volume
Da	Dalton
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E.coli	*Escherichia coli*
et al.	And others
E1α-E1β	Pyruvate dehydrogenase (EC 1.2.4.1)
E2	Dihydrolipoamide acetyltransferase (EC 2.3.1.12)
E3	Dihydrolipoamide dehydrogenase (EC 1.8.1.4)
E3BP	Dihydrolipoamide dehydrogenase (E3) binding protein
EC	Enzyme Commission number
eCFP	Enhanced cyan fluorescent protein
eCFPv	Variant of enhanced cyan fluorescent protein
em.	Emission
EtOH	Ethanol

ex.	Excitation
eYFP	Enhanced Yellow fluorescent protein
eYFPv	Variant of enhanced yellow fluorescent protein
FAD	Flavin adenine dinucleotide
FITC	Fluorescein isothiocyanate
FPLC	Fast protein liquid chromatography
FRET	Förster resonance energy transfer
FSC	Forward-scatter
g	Gram
GFP	Green fluorescent protein
GFP-PC	pETDuet-1-OR2-OR1-Pr-UTR1-eGFP
h	Hour
HCP	Host Cell Protein
hPDC	Human Pyruvate Dehydrogenase Complex
HPLC	High-performance liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IMAC	Immobilized metal ion affinity chromatography
Kan	Kanamycin
k_a	Association rate constant [M^{-1} s^{-1}]
k_d	Dissociation rate constant [M^{-1}]
K_D	Equilibrium binding constant [M]
k_{obs}	Observed binding rate constant [s^{-1}]
k	Kilo
KO	Krome Organe
L	Liter
LB	Lysogenic broth
LD	Lipoyl domain
LET	Linear expression template
M	Molar
MCS	Multiple cloning site
min	Minute
MilliQ water	Ultrapure water filtered through a 0.22 µm membrane filter
MS	Mass spectrometry
MW	Molecular weight
NaCl	Soldiumchloride
NC	Negative control
NAD	Nicotineamide adenine dinucleotide
N-terminal	Amino-terminus
NTP	Nucleoside triphosphate

OD_{600}	Optical density at 600 nm
oligomer	Oligunucleotide
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PC	Positive control
PCR	Polymerase chain reactions
PDB	Protein data bank
PE	R-phycoerythrin
PI	Protease inhibitor
ppm	Parts per million
R_g	radius of gyration
R_h	hydrodynamic radius
RNA	Ribonucleic acid
RNAse	Ribonuclease
rpm	Rounds per minute
S	Svedberg unit [10^{13} s]
SD	Standard deviation
SDS	Sodium dodecyl sulfate
s	Second
SOC	Super optimal broth (with catabolite repression)
SPR	Surface Plasmon Resonance
SSC	Side-scatter
TEMED	Tetramethylethylenediamine
TPP	Thiamine triphosphate
Tris	Tris(hydroxymethyl)aminomethane
TX-TL	Transcription-translation
w/v	Weight/volume
WT	Wild type
U	Unit [µmol/min]
Δ	Gene deletion

Chapter 1

Introduction

The astonishing properties of microorganisms performing specific chemical reactions have been exploited for centuries. The most prominent representative of this area is probably the historical use of yeast for the bioprocess of beer fermentation (Zuidam, 2010). Synthetic biology and metabolic engineering approaches are migrating into the field of chemical industry in recent years, creating new possibilities for conversions and processes. This trend is driven by an increase of environmental awareness in combination with fossil resources coming to an end (Jullesson et al., 2015). A prognosis of Inkwood Research (2017) estimated growth of market share and value for bio-based chemicals worldwide from $6474 million in 2016 to $23976 million by 2025. These numbers highlight the rapid growth in the biotech industry. To realize this growth, increasing the efficiency of bioprocesses is essential, which can be promoted on the process level as well as on the biological level.

One intriguing representative enzyme complex for bioconversion can be found within the metabolic pathway of all organisms, namely the pyruvate dehydrogenase complex (PDC). This complex composed of different enzymes is involved in the energy metabolism of the cell, since it catalyzes the oxidative decarboxylation of pyruvate to acetyl-CoA (Yu et al., 2008). It fulfills its role by using several advantageous mechanisms. Main reasons for increased efficiency are the channeling of intermediates from active center to active center of the involved enzymes, a structured setup of modular subunits, and the ability to regenerate its own cofactor, a feature which is promising and in high demand (Patel et al., 2014). Furthermore, its product acetyl-CoA is a key intermediate for a variety of demanded products like biofules (1-butanol, alkanes/alkenes), pharmaceuticals and fragrances (isoprenoids), nutraceutical ingredients and antioxidants (polyketides and polyphenols) and chemical building blocks (fatty alcohols) (Nielsen, 2014). PDC has already been metabolically engineered to

lower precursor cost of acetyl-CoA in *S. cerevisiae* and to establish it as cell factory for sustainable production of chemicals and fuels. Kozak et al. (2014) expressed functional PDC in the cytosol of *S. cerevisiae* to efficiently produce acetyl-CoA by circumventing the by-product ethanol. This example illustrates the eligibility of PDC for industrial application, however, its features are also of high relevance for basic research, and in the context of synthetic biology, can be utilized to improve bioprocesses in a bottom-up approach, but in future probably more interesting in a top-down approach.

A decisive approach to the optimization of entire bioprocesses is represented by process control. Various strategies are being pursued, which can be miniaturization, high throughput processing, automation, and lately, digitalization and bioprocess intensification. These concepts lead either to reduction of time, enable higher reproducibility, increase yields, save/combine processing steps and finally, save costs (Ball et al., 2018; Bhambure et al., 2011; Knepper et al., 2014; Woodley, 2017). Breakthroughs in high throughput technologies were pivotal for industrial application of biotech based processes. One example is the automation of DNA synthesis, which have lead to a drop of production from \$50 per bp around 20 years ago to less than \$0.2 per bp today (Jullesson et al., 2015). Process control is crucial for highly efficient bioprocesses, because of the complexity of biological systems which should be monitored to track significant process variability (Denard et al., 2015). In this regard, the automation of analytics is already advanced, mainly through fluorescent readying or liquid handling, but automation of whole bioprocesses including their control is still ongoing (Knepper et al., 2014). High-risk bioprocessing steps should therefore be prioritized, since here the advances of automation take effect, and manual errors can be avoided, reproducibility increased and number of open operations to avoid microbial contamination reduced (Ball et al., 2018; Leong et al., 2018). One point to consider when it comes to process automation is the required hardware, which might be limiting (Ball et al., 2018).

Besides the process design itself and its periphery, the central component, the biocatalyst, offers great potential for optimization. The efficiency of bioprocesses can be enhanced throughout improving specificity of the reaction, and quality and yield of the conversion product. Cell factories can use renewable feedstocks and convert them to valuable compounds ranging over a variety of areas (Jullesson et al., 2015). Research is ongoing to enhance biocatlyst properties focusing on enzymatic reactions in order to increase the activity and selectivity and to stabilize the enzymes (Denard et al., 2015). The advances in the fields of synthetic biology and metabolic engineering offer new possibilities for strain engineering. They deliver tools to develop an improved organism host or to engineer biosynthetic pathways and enzymes (Panke and Wubbolts, 2002; Porter

et al., 2016). However, biocatalysts can be dependent on cofactors, which need to be considered from a process engineering perspective. These can be pyridine dinucleotides or phosphorylated trinucleotides. Usually these substrates are costly, necessitating recycling strategies for economical use. For the industrial application this is mostly achieved by whole cell process with cofactor recycling in the metabolic pathway (Schmid et al., 2001). Biocatalyst pursuing their own regeneration strategy can therefore ease the design of a bioprocess.

The engineering of proteins (biocatalysts) can be performed with evolutionary or rational methods. The latter is gaining more importance because it offers the potential to save time and labour, but on the other hand requires *a priori* knowledge of sequence and structure relationship (Arnold, 2018; Porter et al., 2016). Nowadays, the focus of industry is more on altered enzyme properties than on enzyme mechanism. Consequently, a rational or semi-rational library design can provide a mutant library which is diminutive and highly promising and thus affordable to screen (Porter et al., 2016). Currently, a third strategy is accelerating the evolution of enzymes, which is *in silico* protein design (Wang et al., 2019). Here, new proteins can be designed, but also a redesign of existing protein structure is pursued. Accordingly, rational as well as *in slico* methods can benefit from gaining knowledge about enzyme function and dynamics, which can be related to structural properties and help to sort and model enzymatic reactions.

The technologies applied onto a process impact parameters like time and robustness. The cell free protein synthesis offers a platform technology which has already served as a tool for fundamental research, but experienced a "technical renaissance" over the recent years (Gregorio et al., 2019; Silverman et al., 2020). It offers flexibility because of the open reaction environment, the potential for cost savings since it accelerates protein production and obviates any kind of energy investment to maintain the cell, and is easily compatible with high throughput applications because of direct monitoring and fast sampling (Gregorio et al., 2019; Liu et al., 2019). Therefore, it was described as a next-generation biomanufacturing tool, but also demonstrates its potential in fundamental areas such as genetic networks and metabolic engineering, on-demand biosensing, gene expression in non-model organism and synthetic biology education tool (Liu et al., 2019; Silverman et al., 2020). This protein expression toolbox is still developing, yet is already divers enough to cover a variety of systems enabling the production of proteins with different complexity and needs (Liu.2019). The developments in synthetic biology can receive a boost when combining these cell free protein synthesis properties with rational genetic diversity and high throughput screening methods.

1.1 Motivation of this work

The main goal of this work is to build a foundation for efficient and automated protein optimization. Thereby, multimeric human pyruvate dehydrogenase complex (hPDC) serves a role model performing specific reaction cascades enabled through a high degree of organization and modularity of the enzyme assembly. hPDC was investigated at different aspects - the modular structured setup of the core and the enzyme dynamics under different conditions - to enable the incorporation of the gained knowledge into an process chain of fast and targeted DNA manipulation, cell free protein expression and new and robust functionality screening method. Hence, the specific tasks were as follows:

- Establishing a new and automated process chain as a platform for fast enzyme engineering and screening. Towards this end, a flexible method for fast DNA generation, expression of proteins in an open environment via cell free protein synthesis and their property testing via a new and reliable screening method was set up. These steps were fully automated at a liquid handling workstation. A proof of concept was provided with the simple reporter molecule GFP. This new type of rational process opens up the possibility to focus on new biotechnological relevant applications, by utilizing hPDC own mechanism (Chapter 4).

- Acquiring knowledge about hPDC performance at different conditions, namely varying substrate and complex subcomponent ratio, gained through hPDC overall activity assay. In this way critical process related parameters could be identified and a better understanding for the enzyme complex dynamics was obtained (Chapter 5).

- The kinetic subcomplex assembly in hPDC and a possible cross-influence of structural different hPDC core components were investigated. N-terminal lipoyl and hinge domain deficient core variants were constructed and their interaction to peripheral hPDC subunits was examined via the method of bio-layer interferometry. This enabled insight to possible cooperativity of different core domains on the overall complex assembly (Chapter 6).

Chapter 2

Theoretical and Technological Background

This chapter serves to convey information about the state of art of methods and techniques. It provides theoretical background to study the topics discussed in this work.

2.1 Pyruvate dehydrogenase complex

PDC is a multimeric enzyme assembly which occurs in all living organisms and exhibits sizes of several mageadaltons. It belongs to the superfamily of 2-oxo acid dehydrogenase complexes. Different species show divergent architecture and function of subunits (see Table 2.1), but all of them comprise pyruvate dehydrogenase (E1) and dihydrolipoyl acetyltransferase (E2), which work together to decarboxylate pyruvate and generate acetyl-CoA, while dihydrolipoyl dehydrogenase (E3) performs redox recycling of E2 (Patel et al., 2014). The interaction between these subunits result in the astonishing properties of PDC: It can self assemble out of the modular blocks, channel reaction intermediates from one catalytic active center to the subsequent, regenerate its cofactors and regulate itself via manifold mechanisms (Bates et al., 1977; Hezaveh et al., 2018; Patel et al., 2014).
In eukaryotic species, PDC is present in the mitochondrial inner membrane-matrix space (Wieland, 1982). It links the glycolysis to the TCA cycle, where it plays a key role in energy and metabolic homeostasis (see Figure 2.2).
Due to its central role and localization, PDC is also of interest for clinical studies. Its altered regulation can lead to neurodegenerative diseases, cancer and type 2 diabetes mellitus because of mitochondrial dysfunction (Park et al., 2018).

Table 2.1: Overview of PDC subunits in the human (h) species as representative for eukaryotic PDC, in contrast to *E. coli* (ec). Data was extracted from Patel et al. (2014). *These values represent one substitution model.

Enzyme	Cofactor	Subunits	Comment
E1 (EC 1.2.4.1) pyruvate dehyrogenase	TPP	h: 20-30 heterotetr. ec: 12 dimers	
E2 (EC 2.3.1.12) dihydrolipoyl acetyltransferase	lipoic acid CoA	h: 48* (icosahedral) ec: 24 (cubic)	h: 2 LDs ec: 3 LDs
E3BP E3 binding protein	lipoic acid	h: 12* ec: -	not present in ec, E1 and E3 bind to E2
E3 (EC 1.8.1.4) dihydrolipoyl dehydrogenase	FAD	h: 6-12 dimers ec: 6 dimers	

2.1.1 Pyruvate metabolism

The human multi-enzymatic pyruvate dehydrogenase complex (hPDC) is an enzyme cluster which performs coupled enzymatic reaction cycles in a modular fashion, thus it was already referred to as molecular machine (Perham, 2000). Its main actor is the 60-meric core consisting of two subunits, namely E2 and E3 binding protein (E3BP). They contain N-terminal flexible lipoic arms to channel intermediates, and further contain binding domains to interact with surrounding subunits as shown in Figure 2.3. These are thiamine diphosphate dependent E1 and FAD dependent E3 (see Table 2.1). This complex belongs to the branched-chain α-ketoacid dehydrogenase family (Harris et al., 1986) and mediates the oxidative decarboxylation of pyruvate (α-keto acid) to actyl-CoA (Reed, 2001), as sketched in the following Figure 2.1:

$$\text{Pyruvate} + \text{CoA} + \text{NAD}^+ \rightarrow \text{Acetyl-CoA} + \text{NADH} + \text{CO}_2$$

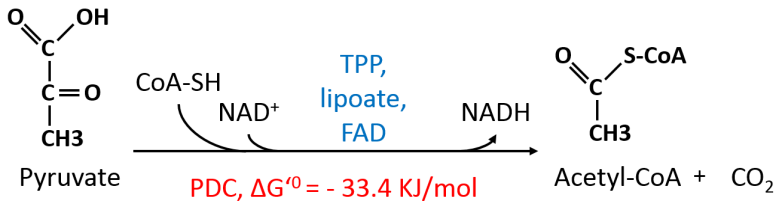


Figure 2.1: Reaction of PDC in dependence of its cofactors (in blue). Energetically, the conversion series is a strong exergonic, thus spontaneous reaction. Scheme adopted according to Nelson et al., 2009.

The cofactor lipoic acid is anchored covalently to C-terminal lysine residue of E2 and E3BP and acts as "swinging arm", which visits all three active centers of the complex (Vijayakrishnan et al., 2010). The reaction begins with a decarboxylation of pyruvate at E1, whereas the lipoic moiety undergoes a reductive acetylation with E1 bound cofactor TPP (see Figure 2.2). Then, the intermediate is channeled to E2 and a subsequent acyl transfer to Coenzyme A (CoA) follows. This forms acetyl-CoA which will be released for further downstream reactions. The remaining dihydrolipoyl moiety of E2 is now to be regenerated. It is reoxidized at FAD dependent subunit E3 by cofactor NAD^+ accepting electrons and being reduced to NADH. This finishes one enzyme catalyzed cycle (Ciszak et al., 2003; Reed, 2001). This reaction sequence is restricted by the irreversible and rate limiting step of subcomponent E1 (Patel et al., 2014; Sanderson et al., 1996).

The regulation of the complex is performed mainly on the E1α subunit by two regulatory enzymes. Since E1 is the first enzyme in the cascade (and as just mentioned, the rate limiting step and irreversible) this is favorable for the complex (Hucho, 1975). It can fulfill its gatekeeper role in the central carbohydrate metabolism through sophisticated mechanism by sensing changes in substrates and products, but also the intramitochondrial redox state of NAD^+/NADH (Patel and Korotchkina, 2006). Mammalian PDC can be regulated by protein to protein interaction. It is inactivated by phosphorylation through four different isoenzymes of pyruvate dehydrogenase kinase (PDK), which can act on three phosphorylation sites of E1 (Patel and Korotchkina, 2001). The four PDK have different affinity and rates towards the phosphorylation sites and act either on the interaction with the lipoyl domain of E2 (sites 1 and 2) or impair TPP binding of E1 (site 3). This can be counteract by two different isoforms of E1 phosphatases which dephosphorylate the E1 residues. PDC is feedback inhibited by its product acetyl-CoA, high ATP levels but also NADH which acts on E3. Thereby, for mammalian PDC, the ratios of NADH/NAD^+ and acetyl-CoA/CoA are regulating (Patel and Korotchkina, 2006).

2.1.2 Organization and stoichiometry of hPDC subcomponents

Eukaryotic PDC exhibits a more complex and more ordered structure compared to most prokaryotic PDC (Izard et al., 1999; Vijayakrishnan et al., 2010). This is, amongst reduced copy number of subunits, due to the core of prokaryotic PDC which lacks the E3BP protein (as shown in Table 2.1, Patel et al., 2014). This subcomponent has structural similarity with E2, but does not catalyze

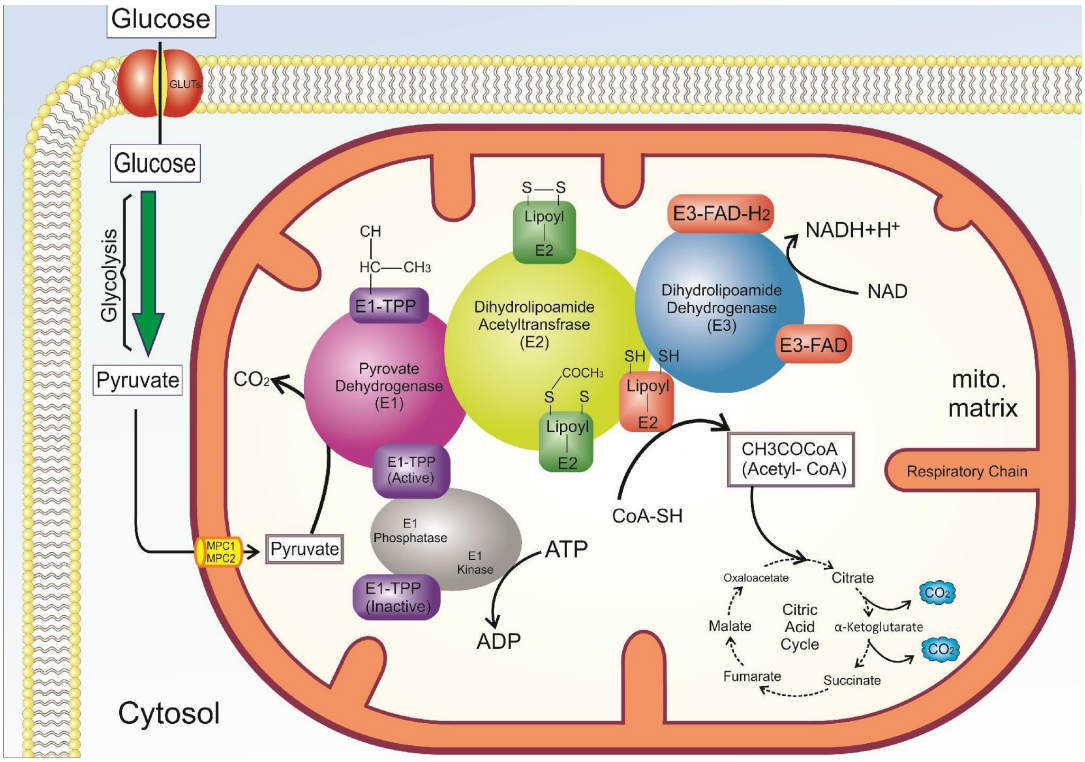


Figure 2.2: The general function of the PDC in the context of the metabolic pathway from glycolysis to the citric acid cycle. Pyruvate is the substrate of the first reaction step in PDC activity which is performed by E1 (purple). This enzyme is controlled strictly and performs the irreversible decarboxylation of pyruvate by binding the resulting acetyl-group at E1-TPP. This acetyl-group is subsequently transferred to the lipoyl-group of E2 (yellow), to enable a transesterification reaction with coenzyme A. acetyl-CoA is formed and released to enter the citric acid cycle. The regeneration of the reduced E2 lipoyl-group is performed by cofactor FAD at E3 (blue). The reduced $FADH_2$ is oxidized by cofactor NAD^+, forming and releasing NADH + H^+ as a last step. This Figure was taken from Nasiri et al. (2020). Copyright 2020, Abolfazl Nasiri, Masoud Sadeghi, Asad Vaisi-Raygani, Sara Kiani, Zahra Aghelan, Reza Khodarahmi. License by creativecommons.org/licenses/by/4.0/.

the acetylation of CoA since a catalytic histidine residue is missing (Harris et al., 1997; Park and Patel, 2010; Prajapati et al., 2019). Despite no catalytic activity, it anchors E3 to an exclusive site on the core and thus allows full functionality of the complex by enabling direct transfer of reducing equivalents between E2 and E3 (Marsac et al., 1993; Yang et al., 1997). It is suggested that E3BP arose from its cognate E2 through gene duplication, since both are located on chromosome 11 (Brautigam et al., 2006; Harris et al., 1997). However, the binding mode of E3 has diverged for more complex architecture of eukaryotic PDC to adopt to special organizational and functional requirements.

The stoichiometry of the modular setup of the eukaryotic PDC is not resolved yet. The core consists of 60 subunits of E2 and E3BP. How to distribute them is discussed in two substitution models, which suggest that some E2 units are exchanged by E3BP, either $E2_{48}$ + $E3BP_{12}$ or $E2_{40}$ + $E3BP_{20}$ (Brautigam et al., 2009; Hiromasa et al., 2004; Vijayakrishnan et al., 2010). Hezaveh et al. (2017) showed, on base of coarse-grained simulations, a higher stability for the $E2_{48}$ + $E3BP_{12}$ model. Whereas Prajapati et al. (2019) verified $E2_{40}$ + $E3BP_{20}$ because of incapability of E3BP to bind CoA. However, the architecture of the eukaryotic PDC core is known to be arranged in a pentagonal dodecahedron shape, self-assembled via interaction of E2 and E3BP subunits at their catalytically C-terminus (Yang et al., 1997).

2.1.2.1 hPDC core structure and domains

The catalytic domains of E2 and E3BP form the core structure of hPDC (see Figure 2.3). On each of the edges of the 60-mer pentagonal dodecahedron core the smallest catalytic unit of 20 trimers is located. Flexible linkers protrude from the catalytic domains of each trimer (Yu et al., 2008). It is assumed that the trimers are not a homogeneous species, but is built of different combinations of E2 and E3BP. The trimer described as $E2_2$ and $E3BP_1$ or a homotrimer consisting of $E2_3$. According to Prajapati et al., 2019 even more constructs are possible, although they would result in units without function if not at least $E2_2$ units are present. Because two active chains per trimer are needed to bind CoA and donate to a catalytic histidine at a neighboring active site (Prajapati et al., 2019).

Having a closer look at the monomeric core structure, different domains constitute E2, and in a slightly modified version E3BP (see Figure 2.3). Both contain large inner catalytic domain at the C-terminus (around 250 aa, Patel et al., 2014). In case of E2, this is connected to an E1 subunit binding domain (E1-BD) via a hinge region. This hinge region contains 20 to 30 residues and is rich in charged amino acids but also contain multiple alanine and proline which makes

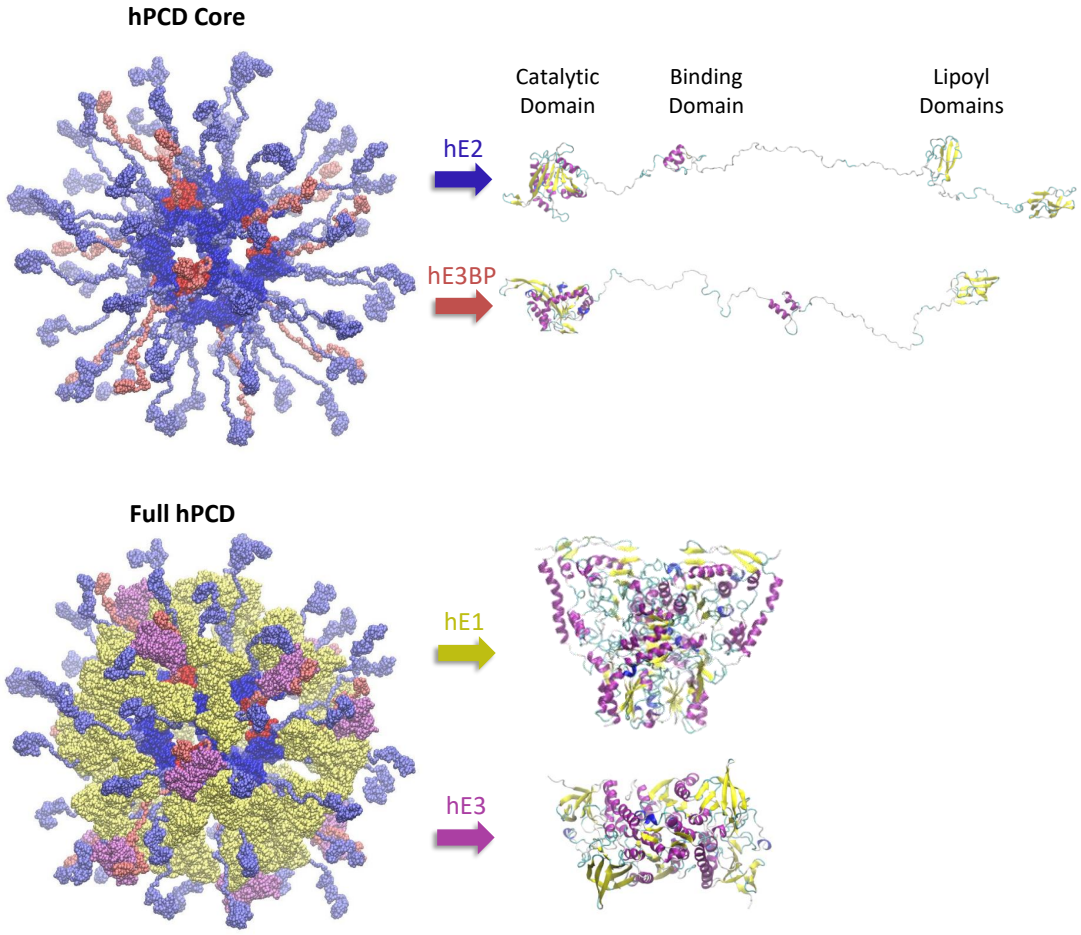


Figure 2.3: Visualization of hPDC at different levels. The upper graph shows the framework, i.e. the hPDC core. This is build by the subcomponents E2 and E3BP, which interact at their catalytic domains. The whole complex is then assembled with the peripheral subcomponents, as shown in the lower graph. Here, the subcomponents E1 and E3 are attached to their respective binding partners. Figure adapted with permission from Depta et al. (2019). Copyright 2019 American Chemical Society.

them flexible (Perham, 2000; Yu et al., 2008). After E1-BD, two further and longer hinge regions follow connecting two globular lipoyl domains of about 80 residues (Deng et al., 2008; Hiromasa et al., 2004; Izard et al., 1999). E3BP shows a similar setup, it has around 50 % sequence identity of the C-terminal catalytic core (Yu et al., 2008). A hinge region connects it to the E3 binding domain (about 35 residues, Mande et al., 1996), this is followed, in contrary to E2, by only one lipoyl domain and two further and longer hinge regions (Hiromasa et al., 2004).

Lipoic acid in free form is naturally occurring only in low levels. Instead, for its biological function, lipoic acid is bound covalently to the amino group of lysine residues of several proteins (Patel and Hong, 1998). This is also the case for the core subunits of PDC, where the lipoyl moieties fulfill their function as "swinging arms". They actively channel intermediates between the active centers of the PDC subunits. Whereas human E2 contains two (outer and inner) lipoyl moieties and E3BP only one with no catalytic, but structural role. They transfer substrates, products and reduction equivalents (Patel and Korotchkina, 2006). Thereby, the lipoyl moieties undergo a reductive acetylation at E1, pass further to E2 and transfer latter to CoA, and finally get reoxidized at E3 by reduction of NAD^+ to NADH (Patel et al., 2009).

2.1.2.2 Protein to protein interaction

Subunits E1 and E3 arrange around the inner core unit by interacting with their respective binding partners as shown in Figure 2.3. For mammalian PDC there are distinct binding domains connecting E1 to E2 and, due to the presence of E3BP, allowing for monospecificity of E3BP to E3. In case of prokaryotic PDC only one binding domain is present for both, E1 and E3 (Patel et al., 2009). The subunit E1 is actually a heterotetramer and consists of two E1α and two E1β enzymes. The E1 binding domain of E2 interacts highly specifically with the C-terminals of both E1β subunits. This interaction is mediated through electrostatic and hydrophobic forces between E1 and E2 (Patel et al., 2009). The E3BP to E3 binding system is organized in a similar fashion. E3 forms a homodimer, this interface interacts highly specific with the E3BP core unit, also via electrostatic bonds and hydrophobic interaction mediated by specific residues (Ciszak et al., 2006; Park and Patel, 2010; Patel et al., 2009). It was shown that a residual affinity of E2 to E3 still remains since for E3BP deficient cases of hPDC a remaining activity of 4 – 20 % was measured (Marsac et al., 1993; Patel et al., 2009; Perham, 2000; Yang et al., 1997). This indicates a loose interaction of E2 to E3, although there is no E3 binding site for E2.
One measure to describe kinetic interaction between proteins are the association

(k_a) and dissociation rate constants (k_d). These two rates finally lead to the equilibrium binding constant K_D (M), also referred to as equilibrium dissociation constant, a measurement for protein to protein affinity. A more detailed explanation and equations will be presented later in section 2.8.1.

2.1.2.3 Intermediate channeling

Enzyme efficiency is also a question of the availability of its substrate. In case of a pathway route this can be an intermediate. They are likely to diffuse away, and so cannot be processed by their downstream enzyme. Therefore, evolutionarily different intermediate channeling mechanisms have developed, like enzyme tunnels or very close proximity of active centers. Another way to accelerate intermediate processing is through cluster mediated channeling (Castellana et al., 2014). Here, agglomerates, which consist of co-clustering multiple enzymes, enhances the probability of an intermediate being processed on an enzyme concentration based effect. In this way metabolic efficiency is increased without the need of an specific arrangement of the enzyme molecules. This mechanism is also known in multienzyme complexes like the α-keto acid dehydrogenase or fatty acid synthase which consist of multiple copies of different enzymes (Hucho, 1975). The direct channeling from enzyme to enzyme prevents the intermediate from diffusing away leading to a kinetic advantage, while branching of intermediates is prevented as well (Hucho, 1975). Also, it is known that protein to protein interaction is used to allosterically regulate enzymes.
PDC does not organize in a loose form of agglomerates, but is more ordered. It exhibits a specific structure as already described in section 2.1.2.1. Nevertheless it utilizes the agglomerate strategy with multiple copies of the same enzyme for the efficient intermediate channeling.

2.2 Enzyme structure to function relationship

The development of a mechanistic understanding of molecular and cellular biological processes depend upon protein sequence, structure and function (Siltberg-Liberles et al., 2011).
The native tertiary protein structure is crucial for biological protein function, which can be described in three layers. In the first layer protein function is associated as a synonym to catalytic activity (of the single enzyme). In the next layer function is defined as a biological process, summing up molecular functions to a metabolic pathway, including protein to protein interactions. Finally, biological function can be used to describe cellular components, the compartments where the functions take place (Lee et al., 2007). All of this functions can be predicted by different computational methods.

Structural information can be gained among others from the primary sequence and the associated tertiary structure. The primary amino acid sequences of many proteins are known or can be identified with low effort, to the contrary not many tertiary structures are resolved. Nevertheless, immense effort has been undertaken within the last decades to identify protein structure and collect them in databases like the Protein Data Bank (PDB). In 2018 the number of available structural entities in the PDB database for 3D structures of proteins and nucleic acid were already 147557, and a growth of nearly 10 % annually is stated (PDB website, 2019). This database is collected through methods like X-ray crystallography, cryo-electron microscopy, or to a diminishing proportion by NMR spectroscopy.
In contrary, protein sequence entities in the database UniProtKB/TrEMBL count up to over 140 million in February 2019. This is due to various genome sequencing projects coupled with high-throughput methods. Characterizing all of them experimentally would be too costly and time intense.
Different approaches are followed to understand the relation between structure and function. Data for sequence and structure is available, but often with no functional annotation. Computational tools which could predict protein function reliably based on structure would increase value of this genomic data and would also benefit the experimental/biochemical verification of the function (Hvidsten et al., 2009; Mills et al., 2015).

The known protein structures only represent a part of the whole protein variety, for instance membrane associated proteins or structurally disordered proteins are underrepresented in data bases. Also proteins are modified to enable the complex and expensive crystallization process for structure elucidation and thus are biased (Siltberg-Liberles et al., 2011). Structurally disordered proteins

are functional proteins which are in need of interaction partners to adopt a well-defined tertiary structure, otherwise these proteins are too dynamic and missed in crystallization studies (Eliezer, 2013; Siltberg-Liberles et al., 2011).

Sequence homology and motifs in general are a good indicator for biochemical function similarity since the same ancestor can be assumed (Lee et al., 2007). Although this holds true for most proteins, there are exceptions. Different functions for homologous sequences as well as common function for different sequences are described (Sleator and Walsh, 2010). Hereby, either sequence similarities or structure similarities are used to derive the corresponding function. This described mismatch can be illustrated by pairwise sequence alignment. Generally when utilizing pairwise sequence alignment and more than 50% identity exists, usually less than 30% of the pairs exhibit same function; or even structural superfamilies that have a variety of functions (Mills et al., 2015). Concretely, this is demonstrated by the TIM barrel and the Rossmann fold. Together, they constitute more than 50 different biochemical functions (Hvidsten et al., 2009; Mills et al., 2015). From an evolutionary point of view, it is believed that structure is more conserved than sequence (Hvidsten et al., 2009; Siltberg-Liberles et al., 2011). A structure to function relationship is more promising than the standard homologous sequence to function relationship. Nevertheless, for both approaches the prediction reliability increases if not the global (fold), but the local (motif) sequence of the active site is compared (Mills et al., 2015; Sleator and Walsh, 2010).

PDC exists in different species performing the same role, the oxidative decarboxylation of pyruvate. The enzymes of the different species exhibit large differences in structure. However, when comparing the overall PDC structures, still similarities can be found (Patel et al., 2014). The E3 subunit binding to E2 via its binding domain reveals one example. Although human and *E.coli* E3 perform both the same function and have similar structures, their binding modes within the PDC are different.

2.3 Oscillation in biological systems

Understanding the dynamic regulation of an enzyme in a metabolic pathway is essential for the ability to control the enzyme. Generally, it is assumed that molecular processes reach a steady state at which fluxes of metabolites are constant. An environmental stimulus can disturb the steady state and lead to oscillatory behavior (Goodwin, 1965).

Oscillations in biological system are well known at different levels, like population or community waves, periodic change in body temperature, heart contractions or rhythms of breathing (Rubin and Riznichenko, 2014). Nonlinearities also occur on a cellular level, because of enzyme or genetic oscillations. This periodic behavior may fulfill a function (like the circadian clock) or just display self-organization of a network (Kruse and Jülicher, 2005). In nonlinear dynamic systems spontaneous oscillation can occur which illustrate a periodically oscillating component in the overall system (Kruse and Jülicher, 2005).

Higgins (1967) predicted a mathematical model for oscillation in the glycolytic pathway before actual fluctuations in metabolite concentrations in yeast and muscle extracts were observed. Here, fluctuations in the metabolite concentrations occur, provoked by the enzymes involved in glycolysis. Another case is the oscillation of intracellular calcium concentration induced by the signaling of permeable calcium channels (Dupont et al., 2011; Rubin and Riznichenko, 2014). A further enzyme assumed to show undamped activity oscillations is the PDC (Selivanov et al., 1994). Kinetic models of *in vitro* and *in vivo* hPDC systems show nonlinear dynamic behavior such as oscillation and multiplicity. Since PDC controls glucose homeostasis via its function as linker from the glycolysis to the TCA cycle, it needs to be strictly regulated via manifold mechanisms. Pyruvate dehydrogenase is deactivated/activated by site specific phosphorylation/dephosphorylation. Furthermore, PDC senses a change in substrate levels as well as the intramitochondrial redox state (NAD^+/NADH) and Acetyl-CoA/CoA ratio (Patel and Korotchkina, 2006).

Enzymes in central metabolic positions may keep steady state conditions despite of fluctuating stimuli like the environment (Higgins, 1967). This might be due to low frequencies, which also explains the quite low number of oscillating bio processes detected. High resolution experimental techniques are essential to detect oscillations, and in case of populations, they need to be synchronized to exhibit the same phase. However, obtained kinetics can be characterized according to their oscillatory or non-oscillatory responses, as Higgins (1967) is describing them:
Firstly, a simple monotonic function describing a rising or decaying curve can be established (see equation 2.1). This can be modified to an overshoot-undershoot function, which exhibits an oscillatory response, but no periodicity (see equation 2.2). This means the time between two maxima is not constant. Furthermore, only a finite number of over- and undershoot are present (n-1). Finally, there can be oscillatory responses with kinetics exhibiting an infinite number of oscillation. Equation 2.3 represents a sustained oscillatory behavior, where $p(t,\omega)$ is the

periodic function. Equation 2.4 also contains this periodic function, but in addition, contains the monotonic decay function $d(t)$ to describe damped oscillations until a new steady state is reached. In case the decay function becomes small ($t \to \infty$), it can be difficult to distinguish damped oscillations from an overshoot-undershoot function.

$$y(t) = c \pm d(t) \tag{2.1}$$

$$y(t) = c + \sum_{i=1}^{n} (\pm)_t d_i(t) \tag{2.2}$$

$$y(t) = c + ap(t, \omega) \tag{2.3}$$

$$y(t) = c + d(t)p(t, \omega) \tag{2.4}$$

c	constant
$d(t)$	monotonically decaying function
a	constant
$p(t, \omega)$	periodic function
ω	frequency

2.3.1 Kinetic model simulation

Biochemical reaction networks contain intermediates which cannot always be captured by wet lab experiments. Therefore, mathematical modeling of networks can be deployed. They allow for a mechanistic understanding of functional role of different molecules, their regulation and reaction kinetics (Crampin et al., 2004). In this work, a system of ordinary differential equations (ODE) based modeling technique is used. Depending on the problem to be solved, other modeling frameworks are available, like a stochastic description used for birth and death processes, Boolean models used for metabolic networks or Petri net models for translation in complex systems (Klipp et al., 2016). Usually, a set of species like proteins can be described by a set of variables describing their amount and state. The variable values in biological systems are often concentrations and a parameter is often time. ODEs are able to describe the dynamic behavior of a variable (thus species) over time. Depending on the biochemical reaction of each species, either a producing or a consuming reaction, their rate laws can be set up. Among others there are the constant flux, mass action, Michaelis-Menten, inhibition or activation kinetics. Once these ODEs are specified with initial conditions, they can be solved by numerical integration. One possibility to solve ODEs is via the Livermore Solver for Ordinary Differential Equations (LSODA),

which delivers a numerical solution of the initial value problem (Radhakrishnan and Hindmarsh, 1993).

2.4 Protein engineering methods

The synthesis of some high value molecules can be accomplished via the chemical route, but can be feasible more effectively as well as environmental friendly and enantioselective via biological catalysts (Steiner and Schwab, 2012). The industrial biotechnology is in need of a broad range of specific catalysts which usually don't occur in nature since enzymes evolve for their own specific purpose. Nevertheless, to make use of the advantages of enzymatic catalysts, tools for synthetic biology, namely protein engineering comes to fruition. This optimization occurs at amino acid (aa) level and can address different protein characteristic like important conversion steps as well as the substrate spectrum or substrate/product inhibition. But also general conditions like protein solubility, temperature, medium components, pH or storage stability are taken into account to result in specific and highly active biocatalysts. One further aspect is the engineering of proteins as biosensors to allow for high through-put screening (Eriksen et al., 2014).

Protein engineering is approached via two different routes as can be seen in Figure 2.4. One is called directed evolution. Iterative rounds of selection and screening of altered protein characteristics are performed, which mimics Darwinian evolution (Porcar, 2010). Thereby, genetic diversity can be generated via random recombining of a set of related sequences by the method of gene shuffling (Steiner and Schwab, 2012). Another alternative to create genetic diversity can be accomplished by random mutagenesis of the protein sequence by different methods like error prone PCR, usage of mutator strain or chemical mutagenesis (Eriksen et al., 2014; Steiner and Schwab, 2012). These methods of random mutagenesis allow for a broadening of the landscape diversity. Information about the structure is required to follow this route, which also aims at mutations that aren't located in the active site. Moreover, a huge mutant library is generated which needs to be screened and evaluated over several rounds, whereas each round only yields in minor improvement. A reliable high-throughput assay as well as time consuming and labor intense work is a perquisite for directed evolution.
These drawbacks don't account for the second protein engineering route, namely rational design. Because of knowledge of biochemical data, protein structure and molecular modeling, a targeted enhancement of protein property is possible

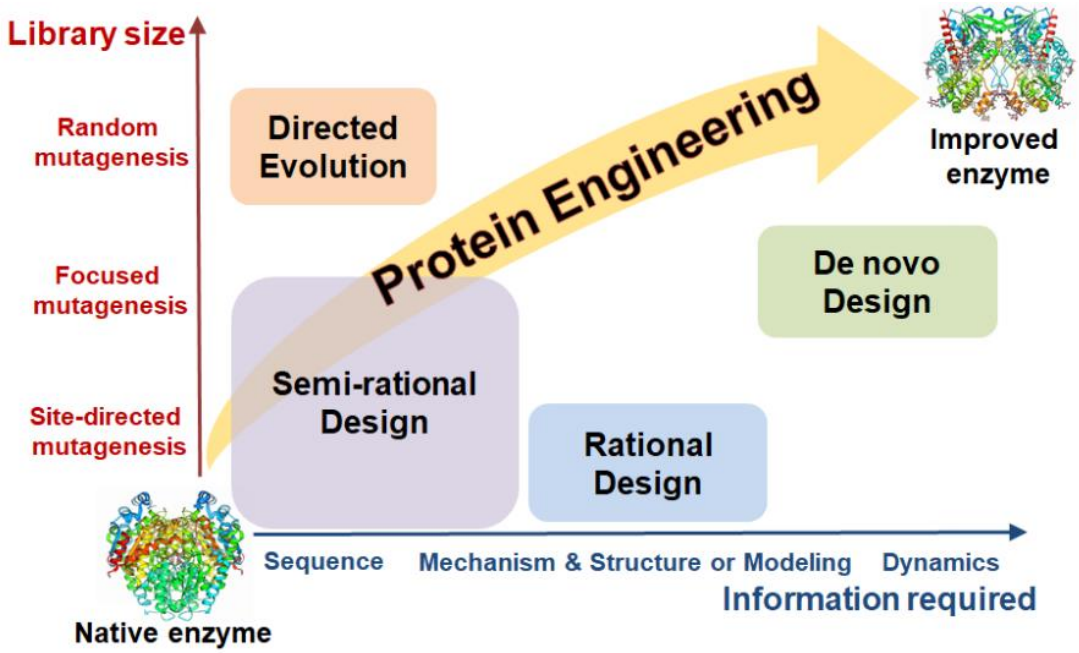


Figure 2.4: Scheme of different protein engineering methods compared according to their library size and *a priori* protein information required. The latter is sorted after the complexity of the information (sequence, structure and dynamics). Reprinted with permission from Liu et al., 2019. Copyright 2019 Elsevier.

(Steiner and Schwab, 2012). Amino acid (aa) specific mutation predictions can be made which likely have a strong influence on the protein characteristic. Either sequence based comparison of homologous proteins is performed or a structure–function relationship study of residues within the catalytic active enzyme pocket can be examined (Eriksen et al., 2014). This route can not only modify proteins, but can also be used to design *de novo* proteins. All together these actions result in a small library to be screened with larger influence of the mutation points.

Nowadays, combing both concepts of protein engineering to benefit of both strategies is established as a "third" alternative. Semi-rational protein design exploits the mechanisms of directed evolution with a more rational choice of key residues to be screened, thus reducing library size. Also, less screening rounds are required to obtain an enhanced protein function (Eriksen et al., 2014). Another option is the random mutagenesis at different target positions which can reveal synergistic effects of sites, but increases library size again (Steiner and Schwab, 2012).

2.4.1 Fluorescent proteins

Research, especially in the field of cell biology, has already benefited from the green fluorescent protein (GFP). This is illustrated by the fact that this small molecule has earned its discoverers and explorers the Nobel Prize in Chemistry in 2008. Namely these scientists were Osamu Shimomura, Martin Chalfie and Roger Y. Tsien. GFP was first discovered in the jellyfish *Aequorea victoria* by Shimomura in 1961 (Tsien, 1998) and ever since developed further.
GFP is one of many different fluorescent molecules with the ability to absorb light in a specific wavelength range. This means its electrons are transferred to a higher energy level (excited state). When these electrons return back to their ground state via the radiative pathway they emit light, which is called fluorescence (Andrews, 2009; Ormerod, 2008). Here, the electrons reach their ground state by emission of photons with retention of spin multiplicity. The excitation will take place at lower wavelengths (higher energy) than the emission (fluorescence) because of energy loss in the excited state by vibrational relaxation (Andrews, 2009; Ormerod, 2008). The gap between the absorption peak maxima and the fluorescent peak maxima is called stokes shift (Andrews, 2009). This fluorescence characteristic cannot only be used for visualization of samples, but also be useful as a sensor, indicator or tracer to provide information about physical and chemical parameters (Andrews, 2009).

All these many applications of GFP and its variants make them a versatile tool for cell biology in a variety of cells and organisms. Most common, GFP is employed as a reporter of gene expression, the localization of proteins in the cell or to measure protein to protein interaction (Cubitt et al., 1995; Dahndayuthapani et al., 1995; Zhang et al., 1996). Another unusual reported application is the usage of a GFP variant (EGFP) been employed as a molecular thermometer (Wong et al., 2007). The application of fluorescent proteins is eased because of their intrinsic capacity to form the chromophore by simple excitation with appropriate light, without the need of exogenously added substrates or cofactors (Dahndayuthapani et al., 1995; Zhang et al., 1996). Also, once expressed, GFP is a robust protein which can be detected in living cells by using noninvasive techniques like fluorescence microscopy or flow cytometry (Zhang et al., 1996).
The structure of GFP consists of an eleven-stranded β- barrel containing a central helix from which the chromophore is formed in the core. This structure protects the chromophore from quenching through water dipoles, paramagnetic oxygen or cis-trans isomerization (Arpino et al., 2012; Watcher et al., 1998). The chromophore is autocatalytically formed by cyclization of residues Ser65, Tyr66 and Gly67 in the case of wild type GFP (Palm et al., 1997). This leads to a covalent rearrangement of the amino acids leading to the formation of a

heterocyclic imidazolinone ring, followed by αβ oxidation of Tyr66 in the presence of molecular oxygen (Arpino et al., 2012; Watcher et al., 1998), so oxygen is essential for the fluorescence signal. In addition, this mechanism can only work for correctly folded protein, thus the function require proper protein folding (Watcher et al., 1998).

Over years GFP has been studied extensively and developed further. For its application in living cells mutations that lead to temperature stability at 37 °C were required. Furthermore, GFP has two excitation maximum. For the application in living cells mild conditions are preferred, thus lower energy input (higher wavelength). To meet this requirements eGFP has been engineered. It is more temperature stable, has only one distinct excitation wavelength (ex.: 489 nm, em.: 508 nm) and its fluorescence is more intense with longer lifetimes and faster maturation which means higher sensitivity. Two main mutations here are F64L which is considered to improve protein folding at 37 °C whilst S65T is responsible for the suppression of the second excitation peak of GFP at lower wavelengths (395 nm) (Arpino et al., 2012).

Other engineering approaches affected the spectral properties. So mutants were generated which could be exited and emit at lower wavelengths compared to GFP, like cyan fluorescent protein (CFP) with maximum excitation at 434 nm and emission at 472 nM (Spiess et al., 2005). One significant change in the protein sequence here is the mutation of Y66W. Others were engineered to the contrary, like yellow fluorescent protein (YFP) with maximum excitation at 514 nm and emission at 527 nM (Spiess et al., 2005). One of the most important mutation point is T203Y (Watcher et al., 1998). The distinguishable emission spectrum of CFP and YFP allows for protein interaction studies on the basis of Förster resonance energy transfer (FRET) measurements. Here two different fluorescent molecules are chosen in such a way that the emission wavelength of the molecule with higher energy sate matches the excitation wavelength of the molecule with lower energy state (Watcher et al., 1998). If the molecules are close together (less than approximately 80–100 Å, Day and Davidson, 2009) an energy transfer takes place, thus the proteins interact. A pair of eYFP and eCFP is one of the most common FRET partners.

2.5 Gene transcription and translation

Before proteins can be employed in various applications, their production must be preceded. The generation of proteins can be summarized by the central dogma of molecular biology. It "deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid" (Crick, 1970). Thus, all the information required for the construction of proteins is encoded in deoxyribonucleic acid (DNA).

The information needed for protein synthesis is encoded by some parts of the DNA sequence, these are referred to as genes. A gene is flanked by an upstream part consisting of a promotor and a 5' untranslated region (UTR), also called leader sequence, followed by the open reading frame (ORF). A start codon (usually AUG coding for Methionine) marks this ORF. Its downstream regulatory parts contain a 3' UTR and a terminator. Assisted by transcription factors a RNA polymerase binds to the promotor region and transcribes the DNA sequence spanning from the 3'UTR, ORF and 5'UTR to a RNA in 5' to 3' direction (Clark et al., 2009). Here nucleoside triphosphates are added to the nascent strand to elongate the template DNA strand. After reaching the terminator region the transcription will be terminated. In eukaryotic cells additionally a subsequent modification consisting of splicing, capping and tailing happens which result in the final messenger RNA (mRNA). The bases of the mRNA are subsequently translated to amino acids (aa) and tied to a polypeptide chain. First, an initiation phases takes place. Here, a ribosome holoenzyme is assembled around the mRNA assisted by initiation factors. The initiator tRNA loaded with methionine binds with the help of aminoacyl-tRNA synthetases to the start codon. This is followed by an elongation phase in which a codon consisting of three bases will be translated to one aa (Clark et al., 2009). Except for methionine and tryptophan all other aa are coded by different codons. The existing four bases are coding for all aa by base-triplets in the end. In total 61 codons exist that encode for 20 aa, this shows the existing redundancy in the genetic code (Spencer and Barral, 2012). Usually an organism prefers one or two codons for an aa which differ between different species, this should be taken into account when expressing genes in a non-natural host. For each codon a transfer RNA (tRNA) must be present which bears the anticodons and via hydrogen bonds pairs the matching bases (Spencer and Barral, 2012). One aa after the other is located at the active site of the ribosome, and catalyzed by elongation factors and hydrolysis of GTP, peptide bonds are tied to generate an aa-sequence. Finally, after reaching a stop codon the translation is terminated with the help of termination factors. The polypeptide chain

is not active upon folding to its native 3D structure and post translational modification (in eukaryotic systems), if required (Clark et al., 2009; Spencer and Barral, 2012). In general, the proteins have the ability to perform self-folding (Anfinsens dogma), but chaperone systems help and facilitate this process.

Protein synthesis in eukaryotic systems is more tedious when compared to prokaryotic systems. This is due to the more complex structure of the eukaryotic system. Gene transcription and translation happen simultaneously in the cytoplasm for prokaryotic cells. Since eukaryotic cells contain more compartments, transcription takes place in the nucleus and translation in the endoplasmic reticulum. Furthermore, eukaryotic cells contain different RNA polymerases depending on its purpose and thus being in need of different transcription factors. This requires different regulation mechanisms of gene control for the different systems. Also the ribosome composition differs. Eukaryotes contain 80S ribosomes assembled out of 60 S and 40 S subunits. The prokaryotic cells have 70 S ribosomes which consists of a small subunit of 30 S and a large subunit of 50 S (Clark et al., 2009). Thereby S is the sedimentation coefficient in Svedberg unit.

2.6 Cell free protein synthesis

For a long time protein expression was believed to require a living cell, in which the processes of transcription and translation occur. Since the early 1950 this theory has been overturned. Various laboratories found evidence that cell homogenates continue to produce proteins (Spirin and Swartz, 2008). Starting with this discovery, intense research was conducted in the field of cell free protein synthesis (CFPS) to provide an efficient new tool for molecular biology.

CFPS exhibits various advantages over conventional protein synthesis. One is due to its open environment which eases direct modification of the reaction condition. This might be interesting for additives like chaperones, mild detergents and lipids for proper protein folding, especially of membrane associated proteins (Katzen et al., 2005). In addition, no cell growth or viability needs to be maintained. Thus optimization of CFPS can be directed exclusively towards the desired protein (Carlson et al., 2012). This allows for toxic proteins to be expressed or a simplified incorporation of non-natural or synthetic aa. Furthermore, conventional cell cloning and transformation problems from cell based workflows are circumvented in CFPS. Finally, and maybe most important for future development and protein engineering, CFPS offers reduction of process time, which in turn enables the

great potential for high-throughput production of protein libraries to investigate protein evolution in terms of functional and structural analysis (Carlson et al., 2012; Katzen et al., 2005).

2.6.1 History and landmarks of CFPS

CFPS is still a fast developing technique, but already having reached landmark advances during last decades. This technique helped to gain the understanding of molecular biology that we have today. The resolution of the genetic code, three bases encoding for an amino acid as well as the contribution to the establishment of the central dogma of molecular biology by Crick are two of its acquisitions (Chong, 2014). To establish the technique as an easy to use and widely spread protein production platform, some improvements were already made. The energy supply was costly and depleted fast, furthermore nucleotides are expensive, thus research focused amongst others on these parts.

One significant improvement on CFPS was made in the 1960s. One drawback was the low amount of expression of exogenous proteins, since endogenous proteins were still produced and required energy and resources. To tackle this problem in case of *E.coli*, the cells were disrupted and subsequently centrifuged at 30,000 g to obtain all the soluble cell components for protein production, but remove macromolecules such as cell-surface membranes and genomic DNA. This active supernatant was known as S30 extract (Shimizu et al., 2005). Subsequently, it was improved by Nirenberg and Matthaei (1961). They performed a run-off reaction of this lysate in presence of all the 20 canonic amino acids and ATP regeneration components (Chong, 2014; Nirenberg and Matthaei, 1961). This procedure is thought to release endogenous mRNA from the ribosomes and in addition allow endogenous ribonucleases to degrade the mRNA, which enhances endogenous protein expression (Liu et al., 2005).

For the purpose of studying gene regulation *in vitro*, Zubay and his colleagues added DNA instead of mRNA to the S30 extract in 1973. By doing so, they implemented the coupled transcription-translation system. Generally, as long as the lysate contains the machinery for translation it can produce proteins. Nevertheless, essential factors like amino acids (aa), energy supply, nucleoside triphosphates (NTPs) and a DNA template need to be added (Katzen et al., 2005). This is the case for the coupled system. Here transcription and translation are performed simultaneously. This presupposes next to DNA and NTP also RNA polymerase addition. For an uncoupled system the aforementioned factors are not added to the CFPS reaction, but instead purified mRNA. Achieved protein yields are usually higher for the coupled system (Katzen et al., 2005). This

might me due to synchronized rates of freshly synthesized mRNA, immediately translated to protein and thus not exposed to degradation or inhibitory tertiary mRNA structures (Chong, 2014; Spirin and Swartz, 2008).

The introduction of non-natural aa is feasible via CFPS, especially with the reconstituted systems. Since ribosomes, their factors and aminoacyl-tRNA synthetases aren't too strict with unnatural aa incorporation, the genetic code can be enlarged (Chong, 2014). This helps to study proteins or to change their properties. Furthermore, Exonuclease V or RNAse I deficient strains can be used to stabilize the DNA template and the resulting mRNA. As a further step the procedure to obtain the cell extract is systematically optimized by considering costs and time expenses. Also, research efforts were directed to shorten and simplify the extract production protocol, improving batch to batch extract consistency (Liu et al., 2005). Isotope labeling is easy to perform with the open environment of CFPS and already helped to determine protein structure via NMR or X-ray crystallography (Carlson et al., 2012).

2.6.2 Lysate systems

Basically, every organism can be used to produce a crude lysate for protein expression. Today some systems are already well studied and commercially available: *E.coli*, wheat germ, rabbit reticulocytes and insect cells. Depending on the protein of interest, some lysates exhibit advantages over others. Eukaryotic based systems are better choice for protein functionality which requires post-translationally modifications, multi domain proteins, membrane associated or proteins requiring disulfide bonds (Chong, 2014; Katzen et al., 2005). This is challenging for *E.coli* lysate, but it exhibits one of the highest protein yields by also showing low costs for fermentation and reaction (Carlson et al., 2012). By using strong phage promotors in combination with efficient phage RNA polymerases in the coupled transcription-translation mode protein yields were increased (Chong, 2014). This was due to a higher efficiency of the phage polymerase over bacterial polymerase, and because of high promotor specificity enabling the expression of the gene of interest only (Spirin and Swartz, 2008). The choice of the system should comprise thoughts about desired protein yield, the origin and size of the protein, its folding and posttranslational modification as well as costs and applications (Chong, 2014).

2.6.3 Cell lysate and reconstituted system

In the crude cell lysate all components are present for transcription and translation as well as protein folding and energy metabolism. Ribosomes are the molecular machine for translation which interact with aminoacyl-tRNA synthetases, factors for translation initiation, elongation and for termination (release factors); proteins for the recycling of nucleotides, and chaperones and foldases for the proper folding of proteins (Carlson et al., 2012). In addition, substrates need to be fed to perform a CFPS which includes aa, nucleotides, DNA template, energy substrates, cofactors and salt.

Next to the cell lysate system another approach arose which utilizes a reconstituted cell free protein production system. Therefore, crucial components for the protein expression are reconstituted separately to a functional suspension (Ohashi et al., 2010). This option, called PURE (Protein synthesis Using Recombinant Elements) system was established by Shimizu in 2001 and consists of a minimal set of individually purified components of translation apparatus of cells (Ohashi et al., 2010; Shimizu et al., 2001). Among others, aminoacyl-tRNA synthetase, translation factors, energy source and regeneration as well as aa and polymerase are reconstituted to a functional PURE system (Ohashi et al., 2010; Shimizu et al., 2005). A big advantage of the PURE system is the fact it contains only minimal nucleases or protease activities that are present in a cell lysate system and lower DNA/mRNA and protein half time (Matsuura et al., 2007; Ohashi et al., 2010). The PURE system enables another kind of flexibility and controllability. It is missing all the factors in cells that were not added intentionally, making it easy to modify. Molecular chaperones and protein disulfide isomerase can be added to help protein folding or change of a reductive milieu (by dithiothreitol) to oxidative milieu (oxidized glutathione) if the protein requires such conditions (e.g. secretory proteins) by exchanging the buffer (Shimizu et al., 2005). The yield of the PURE system is lower compared to the lysate system, but there is still potential to improve this method.

2.6.4 Reaction mode

The accumulation up to inhibiting concentrations of side products like inorganic phosphates and monophosphates is a drawback that cannot be overcome in batch, but in continuous exchange mode. Hereby, consumable substrates are supplied which comprise aa, nucleoside triphosphates (NTPs) and energy regeneration compounds (Spirin and Swartz, 2008). In this fashion reaction duration can be extended from around one hour up to a day which increases protein yields to the range of mg/mL (Schwarz et al., 2007). Beside the mode that is used, the

principle behind this idea is the usage of a porous membrane that separates the cell free protein solution (with large molecular weight components like ribosomes, mRNA and enzymes) from a feeding solution containing small molecular weight substrates. The different existing modes display different advantages. The so called continuous-flow cell-free (CFCF) systems established by Spirin and coworkers continuously pumps a feeding solution into the reaction chambers simultaneously removing products (Spirin et al., 1988). This allows for extended reaction times but is quite costly due to continuous aa and energy supply and exhibits less convenient handling compared to dialysis mode reactions. Here, a passive dialysis performs the feeding supply and the removal of byproducts. This technology called continuous exchange cell free (CECF) is suitable for automation as well (Aoki et al., 2009; Kim et al., 1996; Spirin and Swartz, 2008). Although a batch reaction is simpler to perform and well suited for automation, the dialysis reaction enables more constant reaction conditions.

2.6.5 Energy recycling and costs

The energy needed for the protein synthesis is generated from energy rich phosphate bonds found in compounds like dNTPS. In living cells this energy need to be divided for the maintenance and growth of cell as well as protein expression. In the cell lysate, the energy resource can be solely directed to the latter. However, the energy supply is one of the limiting factors for CFPS. If it gets depleted the reaction stops and no further protein expression can take place. Not only the depletion of of the energy supply impairs further protein production, but also the accumulation of free phosphates, which form complexes with magnesium (Katzen et al., 2005). Therefore, next to energy supply optimization and regeneration, the supplementation of additives has to be regarded to enable higher efficiency.
Free magnesium availability should be optimized carefully for each extract. It is needed for complex formation with phosphate rich compounds which in turn is needed for the kinase activity to generate an energy transfer reaction (Spirin and Swartz, 2008).

NTPs are not only used as building blocks for mRNA, but especially ATP and GTP are also used as energy donor for the translation machinery (Spirin and Swartz, 2008). To spare NTPs, energy rich compounds like phosphoenol pyruvate, creatine phosphate or acetyl phosphate are employed. The accumulation of free phosphates after their hydrolysis leads to problems as described above, so internal recycling strategies are more efficient. Thereby, the central metabolism of the organism is used to regenerate ATP (Carlson et al., 2012). For instance the phosphorylation of ADP by acetyl phosphate which is generated via the

PANOx system from pyruvate. Or the highly effective cytomin system which regenerates ATP through oxidative phosphorylation (Carlson et al., 2012; Spirin and Swartz, 2008). Jewett et al. (2008) was able to increase protein yields up to 1.2 mg/mL in 2 hours utilizing latter energy recycling system and activating central catabolism through growth of lysed cells on excess glucose.
More additives can be added to stabilize the lysate, some will be listed in the following: Dithiothreitol (DTT) is added to avoid formation of disulfide bonds to stabilize the polymerase. The addition of CoA and NAD serve the purpose of ATP recycling via energy metabolism. Further, folinic acid is added as precursor for formyl methionine to push translation initiation. To stabilize nucleic acids spermidin is included as well (Rolf et al., 2019; Spirin and Swartz, 2008).

A great reduction of costs was made by investigating the energy supply not in the CFPS but during extract preparation. Here, the runoff incubation made up 70% of overall reagent costs accordingly to Liu et al. (2005). The main reason was the use of the expensive phosphoenol pyruvate (PEP) for this step. They performed an "empty" runoff reaction in which no additional energy source was added, but claim that the endogenous residual amount of components for the energy metabolism in the lysate is enough for the ribosomes to be released. (Liu et al., 2005).

Compared to commercially available kits, the costs for CFPS can be drastically decreased when utilizing self-made lysates. According to Sun et al. (2013), a self-made *E.coli* based lysate costs \$0.03 per µL of reaction volume and produces 0.75 mg/mL of deGFP at 29 °C and 8 hours of reaction time (this yield was increased with the same lysate over time). Krinsky et al. (2016) shortened and optimized *E.coli* and T7 promotor based lysate production even further and claim a quick lysate production which costs \$0.022 per µL, with 0.15 mg/mL of sfGFP yield within two hours.

2.6.6 DNA templates for CFPS

A notable part of the CFPS reaction is occupied by the exogenous DNA template. Here, quality and concentration are crucial parameters. A contamination of DNA with RNAse lowers the final protein yield. Also DNA concentrations should be sufficient to make full use of the provided substrates in the cell free extract to increase efficiency (Schwarz et al., 2007). Without the need for a transformation step in CFPS, there is increased flexibility in the form of supplied DNA template. The DNA template for CFPS can be a circular plasmid, like it is required for cells, or a simple linear sequence named Linear Expression Template (LET). Nevertheless, both should contain all regulatory parts which are also needed for

a conventional protein expression, like a promotor and untranslated region with strong ribosome binding site. A variety of strong phage promotors as well as endogenous sigma promotors are available for *E.coli* lysate (Sun et al., 2014).

2.6.6.1 LET based CFPS

LETs can be produced fast and convenient via PCR and employed directly in the CF lysate. Thus, LETs bypass tedious cloning and transformation steps which represents a significant saving of time (from days to hours)(Yabuki et al., 2007). Another advantage is the compatibility with automation providing great future development potential. However, in comparison to a plasmid based CFPS, LETs display only a small percentage of their protein yields. The vulnerability towards nucleases due to free accessible ends shortens their lifespan and lead to a fast cessation of the protein production (Schinn et al., 2016).

The RecBCD (ExoV) helicase-nucleases is an instrument in bacterium to repair DNA double strand breaks and serves in genetic recombination. It utilizes NTPs to unwind and hydrolyze DNA on a rate depending on ATP and magnesium levels, what makes it to a compound adversely for CFPS (Amundsen et al., 2012). Therefore, research has been undertaken to inhibit RecBCD activity. Murphy (2007) investigated the bacteriophage encoded lamda Gam protein which is able to bind to RecBCD as a dimer and inhibit in turn the binding of the exonuclease to the double strand DNA. Despite an overload of Gam protein, nevertheless a total inhibition of RecBCD is not attained. In the case of Sun et al. (2014), around ⅓ of plasmid performance could be reached for LETs through addition of the Gam protein to the CFPS. Further improvement for LETs were achieved via steric hindrance. Therefore, protective (random) sequences were added up- and downstream of the LET. A base pair length of around 250 is a good trade-off between bp length and yield which was increased six fold compared to no steric protection (Sun et al., 2014). The effect of small molecules as RecBCD inhibitor is studied by Amundsen et al., 2012. He identified some small compounds that display an inhibitory effect on RecBCD and can be useful for LET based CF reaction. Finally, LET concentration is an easily adjustable variable with big impact on the final protein yield.

2.6.6.2 LET generation

A possible technique for fast and reliable LET production was described by Stemmer et al., 1995. He derived an assembly PCR method based on DNA shuffling by using short DNA sequences, 40mer oligonucleotides (oligos), and assembled them into a full gene of 2.7 kb length via polymerase reaction. Each

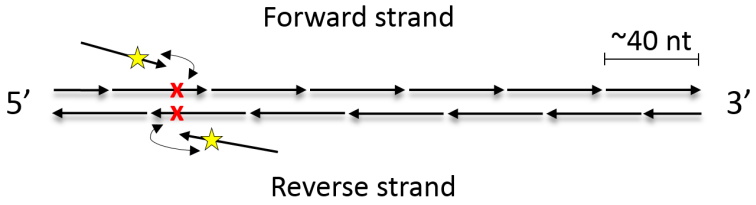


Figure 2.5: Schematic representation of set of adjacent and overlapping oligonucleotides for the assembly of LETs. Possible point mutations (red X) can be introduced by exchanging one forward and one reverse oligonucleotide with the respective change in the sequence.

neighboring oligomer overlapped with each other by 20 complementary base pairs. Likewise, in a one pot reaction gene length fragments were generated (Kosuri and Church, 2014). In comparison to approaches like high stringency ligation based gene synthesis, the polymerase cycling assembly (PCA) exhibit higher error rate, but also higher flexibility which is desired to introduce targeted diversity (Kosuri and Church, 2014). No gap between consecutive oligos is acceptable for ligase chain reaction, but if desired works for PCA (Rouillard et al., 2004). However, both approaches suffer from GC rich sequences or inhibitory secondary structures. They lead to miss-annealing or inhibit assembly and thus are only feasible for constructs in a length of maximal few kbp (Kosuri and Church, 2014).

Due to efficient high fidelity polymerases, the error rates per base of PCA could be decreased over time (Yabuki et al., 2007). Kosuri and Church (2014) report an error rate for a two-step PCR of 0.3 per kbp, Sequeira et al. (2016) stated an error rate of 1.13 per kbp. Further decrease of error rate in gene synthesis can be achieved for a smart oligo design which considers thermodynamically constraints (Sequeira et al., 2016). Melting temperature, GC content as well as specificity are crucial parameters to pay attention to when fragmenting a DNA sequence into neighboring oligo set with overlap to represent the double strand. Rouillard et al. (2004) introduced a degree of freedom for the bp length of each oligomer in one set. In this way uniform hybridization melting temperature as well as adequate specificity can be ensured. Also the quality and concentration of inner and outer oligos and the number of cycles in the PCA are critical parameters influencing the final nucleic acid yield (Sequeira et al., 2016).

Finally, in a one pot reaction via a two-step PCR, LETs can be generated from a pool of overlapping oligos. In the first step the full set of all oligos are assembled together with each oligo in a low concentration. After, this mixture is utilized as template DNA, and together with the outermost forward and reverse oligos

amplified to a full length gene with concentrations sufficient for CFPS (Yabuki et al., 2007). A purification of the PCR product is not strictly needed between or after gene synthesis. The generated template can be directly used for CFPS, allowing simple automation. This modular setup is well suited for targeted mutation. By exchanging a set of complementary oligos owning a mutation site and replacing them with newly designed oligos with exchanged bases, variants are produced with low effort (see Figure 2.5).

2.7 Automation in bioprocess development

The achievements and progress in production and detection in bioprocess development over the last decades have accelerated the field of biotechnology. This can be seen for the exponential growth of genomic and metagenomic databases (Sequeira et al., 2016). Technical progress has already led to the establishment of high-throughput methods, which resulted in a reduction of time and costs (resources as well as personnel). Thereby, high throughput is characterized by miniaturization, parallelization and automation (Bhambure et al., 2011). Miniaturization results in lower volumes, reagent requirements and faster experimental processing. Parallelization allows for implementation of process analytical technology, which helps to screen and evaluate a large experimental space in parallel allowing for faster optimal condition screening and gaining of a deeper process understanding. A further aspect, the automation in bioprocess development, is still limited to single step applications (Knepper et al., 2014). Although automation enhances the information content of each process step in terms of reliability and amount of process parameters monitored. The first fully automated robotic systems came up in the 1990s for bioassay development using microtiter plates. In this way plate to plate variations and manual handling errors could be avoided (Bhambure et al., 2011). Nowadays, the market is focusing on fluid handling platforms and analytic detection for robotic instruments, which enhance speed, accuracy and robustness of the processes (Bhambure et al., 2011). Fully automated process chains and evaluation will increase the efficiency of industrial bioprocess development and decrease the "time to market" in the biotech industry (Knepper et al., 2014).
Process development thereby relies on one or several techniques. The heuristic or knowledge-based process development relies on experience and knowledge of the personnel. An algorithm- or optimization-based method relying on mathematical models for the process optimization is desirable, but it requires mechanistic understanding of the process. Moreover, further high throughput experimental

optimization using robotic equipment, sensitive detectors and data progressing and controlling software for rapid process development is needed (Bhambure et al., 2011).

2.8 Measuring principles

2.8.1 Bio-layer interferometry

Bio-layer interferometry (BLI) offers an optical analytical platform to study dynamic protein to protein interaction over time without the need of labeling samples. An advantage of this technology over other well established label free methods like Surface Plasmon Resonance (SPR) is the missing microfluidics part which is prone to clogging (Abdiche et al., 2008; Rich and Myszka, 2007). When using label-free method, additional processing steps as well as the concern about an influence of the label to the sample is omitted. Furthermore, BLI is suitable with high throughput analysis because of parallelization and the ability to deliver multiple samples simultaneously via well plates (Rich and Myszka, 2007). Moreover, because of capture based immobilization method, samples in crude background can be employed. The sensitivity of detectable molecules ranges from less than 1 kDa up to few thousand kDa (Pall ForteBio LLC, 2017).

The BLI technique utilizes disposable fiber optic biosensors which bind the ligand to be investigated in an open shaking microplate. This fact allows to reuse the sample since it is not depleted, but nevertheless depends on the stability of itself (Abdiche et al., 2008).
Incident white light is guided through the biosensor and is reflected from its functionalized tip (see Figure 2.6). When a ligand is loaded onto the tip by dipping into a solution of protein, the bio-layer thickness increases which causes a wavelength shift of the reflected beam. After a baseline step, the biosensor is dipped into an analyte solution. The interaction of the loaded ligand to the analyte causes a second wavelength shift which is monitored online. Finally a dissociation step follows in which the breakage of the interaction can be studied. By considering analyte concentration and nature, kinetic parameters and binding mechanism can be derived. It is possible to adjust parameters like temperature of reaction, run time of each step, pH and analyte concentration (Abdiche et al., 2008). If an interaction reaction is allowed to run long enough, a reaction equilibrium can be reached and studied (Rich and Myszka, 2007). Furthermore, it is possible to obtain specific kinetics from a target protein in a crude sample. This however is dependant on the non-interaction of crude molecules with neither the biosensor nor the ligand, as well as no influence on the refractive index (Pall ForteBio LLC, 2017).

The nature of binding is essential for data evaluation. In the most simple case a 1:1 binding model is assumed. Interaction partner ligand (L) and interaction partner analyte (A) form a complex of (LA), whereas the reaction is reversible and

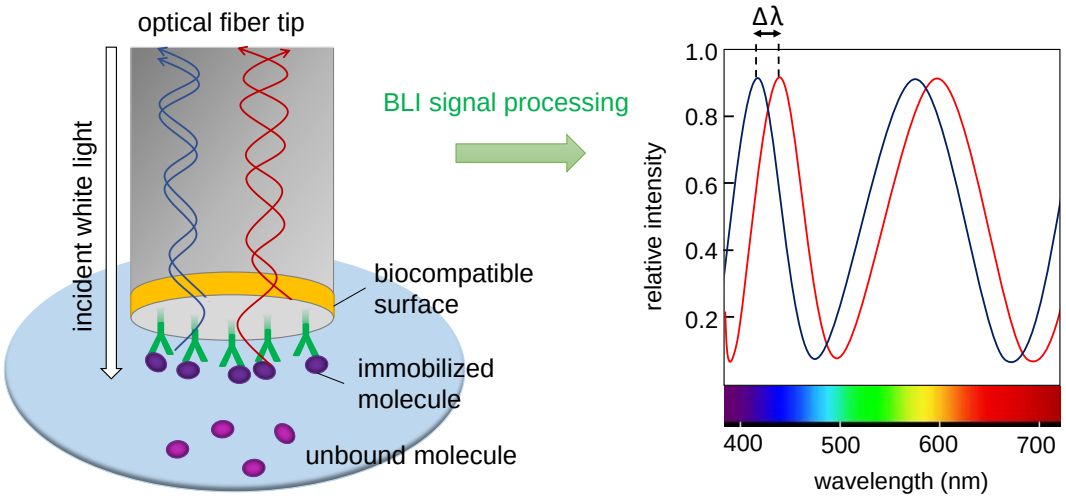


Figure 2.6: Optical analytic principle of bio-layer interferometry. Incident white light is guided to the surface of an optical fiber (left). Any interaction of molecules at this surface leads to an increased bio-layer thickness, which causes a shift in the reflected beam pattern. The recorded spectral shift (right) is the base for the bio-layer interferometry analysis. This Figure was adopted according to Wallner et al. (2013).

can be described by an association rate constant k_a ($M^{-1}s^{-1}$) and a dissociation rate constant k_d (M^{-1}) (see equation 2.5 which is the law of mass action). When regarding the association and dissociation rate constants individually, which are dependent on the concentration of ligand [L] and the concentration of analyte [A], they can be described by a single exponential function (see equation 2.6 and 2.7).

$$L + A \underset{k_d}{\overset{k_a}{\rightleftharpoons}} LA \tag{2.5}$$

$$\text{association rate} = k_a \cdot [L] \cdot [A] = \frac{d[L][A]}{dt} \tag{2.6}$$

$$\text{dissociation rate} = -k_d \cdot [LA] = \frac{d[LA]}{dt} \tag{2.7}$$

[L]	concentration of ligand	(M)
[A]	concentration of analyte	(M)
k_a	association rate constant	($M^{-1}s^{-1}$)
k_d	dissociation rate constant	(s^{-1})

During the association phase of an experiment, both, association and dissociation occur, therefore the interaction of the molecules in the association phase is described by the kinetic equation 2.8. From the association step, a rate constant is derived which is dependent on the analyte concentration, namely the observed binding rate k_{obs} (s^{-1}) (see equation 2.9). k_{obs} thereby is a fitted observed rate constant indicating how fast the equilibrium loading is achieved. When plotting k_{obs} over [A], it is possible to derive k_a and k_d, however mostly it is calculated (as shown in equation 2.9)(Marquart, 30.05.2020). To fit the association phase, a differential equation is integrated in which the rate of association is presented as a function of the decrease of unbound ligand molecules due to analyte binding (see equation 2.10). The dissociation phase occurs in an environment free of unbound analyte. Therefore only a decay function is needed to describe the kinetics dependent on k_d and [LA]. Equation 2.11 describes the decay in the concentration of bound [LA] over time (Tobias and Kumaraswamy, 2013). By solving the equations 2.10 and 2.11, k_a can be calculated (by using equation 2.9). Finally, the affinity constant, or equilibrium dissociation constant K_D (M) can be derived which is the quotient of k_d to k_a (see equation 2.12). Thereby, the interaction is tighter, thus the ligand to analyte affinity greater, the smaller the affinity constant is.

$$\text{assoc. phase kinetic} = \frac{d[LA]}{dt} = k_a \cdot [L] \cdot [A] - k_d \cdot [LA] \tag{2.8}$$

$$k_{obs} = k_a \cdot [A] + k_d \tag{2.9}$$

$$Y = Y_0 + A(1 - e^{-k_{obs} \cdot t}) \tag{2.10}$$

$$Y = Y_0 + A \cdot e^{-k_d * t} \tag{2.11}$$

$$K_D = \frac{k_d}{k_a} = \frac{[L][A]}{[LA]} \tag{2.12}$$

A	asymptote	(s^{-1})
K_D	equilibrium binding constant	(M)
k_{obs}	observed binding rate constant	(s^{-1})
Y	level of binding	(nm)
Y_0	binding at start of assoc.	(nm)

K_D can be utilized to describe the thermodynamic potential of the Gibbs free binding energy $\Delta G°$(J mol^{-1}) as shown in 2.13.

$$\Delta G = RTlnK_D \tag{2.13}$$

ΔG°	Gibbs free energy	(J mol^{-1})
R	gas constant	8.314 (J K^{-1} mol^{-1})
T	temperature	(K)

In a real experiment using a BLI biosensor, only a limited number of ligand can be immobilized at the biosensor surface, resulting in a maximum number of possible binding sites for the analyte binding at equilibrium R_{max}. Therefore, the amount of unbound ligand [L] and bound ligand in the complex [LA] are dependent on each other, as shown in equation 2.14. When considering this relationship for the complex concentration after time, and introducing it to equation 2.8, a function results as presented in the equation 2.15. In this equation, [A] can be regarded as a constant, because the amount of free analyte is much larger than the amount of interaction sites to the ligand, so although the complex [LA] is forming, the analyte concentration nearly remains constant over time. When solving this differential equation, equation 2.16 is obtained. Because there are two unknowns in this equation, k_a and k_d, firstly k_d should be determined independently. In the dissociation phase there is no analyte present in the solution, thus the term for [A] is eliminated in equation 2.17 and k_d can be calculated. Thereby, $[LA]_0$ refers to the amount of analyte bound to the start of dissociation phase.

$$[\mathrm{L}] = [\mathrm{L}]_{\max} - [\mathrm{LA}] \tag{2.14}$$

$$\frac{d[LA]}{dt} = k_a \cdot ([\mathrm{L}]_{\max} - [\mathrm{LA}]) \cdot [\mathrm{A}] - k_d \cdot [\mathrm{LA}] \tag{2.15}$$

$$[\mathrm{LA}](\mathrm{t},[\mathrm{A}]) = [\mathrm{L}]_{\max} \cdot \frac{1}{1 + \frac{k_d}{k_a \cdot [\mathrm{A}]}} \cdot (1 - \mathrm{e}^{-(k_a \cdot [\mathrm{A}] + k_d) \cdot t}) \tag{2.16}$$

$$[\mathrm{LA}](\mathrm{t}) = [\mathrm{LA}]_0 \cdot \mathrm{e}^{-k_d \cdot t} \tag{2.17}$$

$[\mathrm{L}]_{\max}$	maximum ligand binding level	-

The simple 1:1 model does not account for cases where mass transport limitations or heterogeneity of the ligand or analyte occur. In these cases the equations need to be modified to take parameters such as additional binding sites into account (Tobias and Kumaraswamy, 2013).

2.8.2 Flow cytometry

A further technique which relies on optical laws is flow cytometry. It is used to examine particles/cells suspended in a fluid for their optical characteristics. An advantage of this technique in contrast to other optical methods such as

microscopy is that each particle suspended in liquid is analyzed once (Givan, 2001). The fluidic part enables a high particle throughput (multiple thousands) in a short time span, and because of high number of investigated particles also the calculation of robust statistics (Givan, 2001).

Usually a flow cytometer is triggered by cells exhibiting intrinsic fluorescence. In this work fluorescence signal of proteins produced in a cell free manner were detected using flow cytometer. To enable the trigger event for a measurement, a currently established protocol by Wurm et al. (2019) was used in which cells were mimicked by 0.909 nm polystyrene particles. This beads were non-functionalized and triggered measurement events to allow for the detection of fluorescence in the bulk solution surrounding the beads.

The sample is transported via a fluidic system through the flow cytometer. It passes a focused light beam in a sheath fluid stream which is adjusted to a flow rate that allows only one particle to pass at a time. For illumination, lasers are used because of their ability to focus light very precise (Givan, 2001). The sheath fluid is used to adjust a flow rate and center the particle in the light beam. The principle of lasers producing light relies on energized electrons which are transferred to high energy orbitals with high voltage electricity (Wilkerson, 2012). Photons of light are created when the electron falls back on its lower energy orbital. When the light strikes the particles it deflects to the forward-scattered light (FSC) and side-scattered light (SSC, 90 degree to laser beam), and the fluorescent signal. The collection optics consisting of filters and mirrors guide the light to an electronic network (photodetectors) which turns the analog signal into a light intense proportional digital signal (Adan et al., 2017; Wilkerson, 2012). The scattered light signal is converted into voltage pulse, if there is no particle in the beam the pulse reaches baseline. The light scattering allows to draw conclusion on the size in case of FSC because the light refracts around the particle. The granularity of the particle is detected at 90 degree angle scatter (SSC), thus different populations in the sample can be distinguished by investigating these both scatter signals (Wilkerson, 2012).
The fluorescence emission is detected at the same angle as SSC and shows a direct amount of the fluorescent sample (Adan et al., 2017). A band pass filter enables the specific detector configuration whereas it filters a narrow wavelength range in which the emission peak of the fluorophore has its maximum. Most flow cytometers have different lasers built in with different excitation wavelengths, the most common here is the argon laser with an excitation wavelength of 488 nm (Wilkerson, 2012).

Once a digital signal is generated, data evaluation is a crucial point. Usually a plot of FSC over SSC is generated to gate populations that will be analyzed. Populations which represent for example particle/cell debris can be filtered out and won't be analyzed. After, the fluorescence signal of these populations can be investigated in histograms, 2-D plots (dot plots, contour and density plots) and also 3-D plots (a tomogram plot) (Adan et al., 2017). In a histogram the signal of a population in a channel is presented over the number of events in this channel. This results in a peak at a distinct channel intensity, thus gives information about the amount of fluorescence.

Chapter 3

Materials and Methods

In the following, materials and methods used for this work will be presented. A part of the used materials can be found in the Appendix as indicated in the text.

3.1 Strains and plasmids

Strains and plasmids used for this study are listed in the following Tables 3.1 and 3.2. For general cloning purposes *E.coli* NEB 10-beta (New England Biolabs, Frankfurt) was used. For routine T7 protein expression purposes *E. coli* BL21(DE3) derivatives were used. For coexpression of two target genes the pETDuet-1 plasmid (Novagen, Massachusetts, USA) was used. This plasmid is designed with two multiple cloning sites (MCS) each of which is preceded by a T7 promoter, lac operator and ribosome binding sites.

3.2 Molecular cloning techniques

In the following subsections different methods are going to be presented that are used to perform molecular cloning work.

3.2.1 Plasmid DNA isolation (Miniprep)

To extract plasmid from bacterial cells the NucleoSpin® Plasmid kit from Machery-Nagel was used after manufacturer's recommendation. Routinely a five mL over night (ON) culture was processed after manufacturer's recommendation and the plasmid was eluted in 40 µL to result in a plasmid concentration around 100 ng/µL. DNA concentration and purity were determined via NanoDrop One (VWR, Darmstadt). For pure DNA the ratio of 260:280 nm should be between 1.8 - 2.

Table 3.1: Strains used for this study.

Strain	Genotype / description	Source
E. coli NEB10-beta	*(ara-leu) 98;9 araD35; fhuA lacX96 galK38 galE37 e36- :2dlacZ M37 recA3 relA3 endA3 nupG rpsL (Str^R) rph spoT3 (mrr-hsdRMS-mcrBC)*	NEB
E. coli BL21(DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] hsdS λ DE3 = λ sBamHIo EcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 nin5*	NEB
E. coli T7 Express LysY	*MiniF lysY (Cam^R)/fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10–TetS)2 [dcm] R(zgb-210::Tn10–TetS) endA1 Δ(mcrC-mrr)114::IS10*	NEB
E.coli BL21(DE3)pLysS	*F- ompT hsdSB(rB- mB-) gal dcm (DE3) pLysS (Cam^R)*	Novagene
E. coli Lemo21(DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] hsdS/ pLemo(Cam^R) λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 pLemo = pACYC184-PrhaBAD-lysY*	NEB
Rosetta(DE3)	*F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE (Cam^R)*	Novagen

Table 3.2: Plasmids used for this study.

Plasmid	Genotype / description	Source
pETDuet-1	Expression vector with IPTG inducible T7 promotor	NEB
pACYCDuet-1	Expression vector with IPTG inducible T7 promotor	NEB
pET-28a(+)	Expression vector with IPTG inducible T7 promotor	Novagen
pBAD	Expression vector	Addgene

3.2.2 Polymerase chain reaction

Polymerase chain reactions (PCR) were carried out to either insert or delete a sequence into a plasmid, to verify a plasmid (colony PCR, see section 3.2.8) or for LET production (see sction 3.6.2). Depending on the specific purpose, different polymerases were used. The cycling protocols are dependent on the employed polymerase, but general steps are common among all:
PCR reaction mixture were prepared in PCR tubes containing primers, polymerase, buffer and DNA template. The PCR reaction was conducted using Thermocycler Doppio from VWR. The cycling reaction starts with an initiation phase, followed by multiple cycles of denaturation → annealing → elongation and finishes with an final extension step.
For routine gene insertion/deletion as well as *de novo* gene synthesis DNA polymerase Phusion™High-Fidelity from NEB was employed. For colony PCR Maxima™Hot Start Green 2x PCR Master from Thermo Fisher was employed according manufacturer's instructions.

3.2.3 DpnI digestion

PCR products were digested with FastDigest DpnI (Thermo Fisher Scientific, Waltham, Massachusetts, USA) to get rid of the template DNA. DpnI cleaves only when its recognition site is methylated and so is sensitive to plasmid that was obtained from cells. The reaction mixture was prepared by following the manufacturer's guidelines and 37 °C incubation in a heat block for 30 min.

3.2.4 Agarose gel electrophoresis

To analyze PCR products or any kind of DNA template agarose gel electrophoresis was performed. Firstly, a 1 % (w/v) solution of agarose dissolved in 1x Tris-acetate-EDTA (TAE) was prepared. The recipe for 50x TAE buffer is indicated in section 3.9. This solution was boiled using a microwave and poured into a agarose gel chamber assembled with a comb to produce wells. After the gel has solidified, it was placed in a electrophoresis chamber filled with 1x TAE. One well was loaded with 5 µL of GeneRuler DNA Ladder Mix (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The PCR samples were mixed with commercial 6x loading dye (Thermo Fisher Scientific, Waltham Massachusetts, USA) and loaded into the remaining gel pockets. The chamber was connected to the power supply (Bio-Rad Laboratories Inc., Hercules California, USA) and run for 45 min at constant voltage of 120 V. Afterwards, DNA was visualized by staining the gel with SYBR Gold (Thermo Fisher Scientific, Waltham Massachusetts, USA) diffusively over the surface for around 30 minutes and captured with Gel

Doc 2000 (Bio-Rad, Hercules, California, USA).
In case the DNA was needed for further application the band of interest was excised from the agarose gel and cleaned up with NucleoSpinGel and PCR Clean-up (Macherey-Nagel, Düren, Germany) according to manufacturer's specifications.

3.2.5 Circularization of linear DNA

Linear DNA can be circularized by first phoshorylating the linear end and and then ligating both ends together. Therefore, cleaned PCR product is 5'-phosphorylated by using T4 Polynucleotide Kinase (PNK) (Thermo Fisher Scientific, Waltham Massachusetts, USA). The protocol was slightly modified (see 3.3) to make it better compatible with the subsequent ligation step.

Table 3.3: Protocol for phosphorylation of DNA.

Linear DNA	20 pmol
T4 ligase buffer	2 µL
ATP, 10 mM	2 µL
T4 Polynucleotid kinase	1 µL (10U)
Water	to 20 µL

The suspension was mixed thoroughly and incubated at 37 °C for 30 minutes. PNK was then inactivated by heating the mixture at 75 °C for 10 minutes.
The 5‘ phosphorylated PCR products were then circularized by using T4 DNA Ligase (Thermo Fisher Scientific, Waltham Massachusetts, USA). Since the protocol of the phosphorylation was modified, also this subsequent protocol had to be altered. To the heat inactivated phosphorylated mixture 1 µL T4 DNA Ligase (5 U µL-1) and 0.5 µL T4 Ligation buffer were added and incubated at 22 °C for at least 1 hour. The modified plasmid is now ready to be transformed into a host.

3.2.6 Competent cell production

A colony of a bacterial host from a LB plate was inoculated into two mL LB liquid medium and incubated at 37 °C, 200 rpm ON. This preculture was used to inoculate 100 mL LB medium in a 1:100 ratio and shake at 37 °C until OD_{600} reaches 0.25 to 0.3 (usually within 1.5 to 2 h). From here all further work was performed strictly on ice if not stated otherwise. For harvesting, the culture was separated into two falcon tubes of 50 mL and chilled on ice for 15 min. Also the sterile 0.1 M $CaCl_2$ solution and 0.1 M $CaCl_2$ plus 15 % glycerol were rest on ice. Subsequently, the cells were centrifuged for 10 min at 3300 g at 4 °C. The

supernatant was discarded and the cell pellet was resuspended in 15 ml of cold 0.1 M $CaCl_2$. A further chill step on ice for 30 min followed. Again, the cells were centrifuged for 10 min at 3300 g at 4 °C. Afterwards, the supernatant was discarded and cells were resuspend in two ml of cold 0.1 M $CaCl_2$ + 15% glycerol solution. Finally, under sterile conditions 0.2 mL aliquots of the competent cells were prepared and frozen at - 80 °C. Once thawn they cannot be frozen again, the frozen competent cells are stable up to 1 year.

3.2.7 Transformation

To transform a bacterial host with plasmid, following protocol was used: competent cells were withdrawn from -80 °C freezer and thaw on ice for 10 min. Around 1 µg of insert, but not more than 5 µL of DNA template was added to 50 µL competent cells and chilled on ice for 30 min. Subsequently the mixture was heatshocked at 42 °C for 30 seconds. The mixture was chilled on ice for another 5 min, followed by addition of 500 µL pre-warmed SOC media and incubation at 37 °C for 1 h at 200 rpm. The cells were spread under sterile conditions on pre-warmed LB Agar selection plates. Routinely, different dilutions of the cells were plated to obtain single colonies after ON incubation of the plates at 37 °C.

3.2.8 Colony PCR

This technique was used to screen colonies for the presence or absence of a cloned gene of interest. If the PCR product based on a picked colony corresponded to the expected gene size, the plasmid was sequenced to verify its flawlessness.
A colony was picked in a sterile environment and resuspended in 4.6 µL H_2O in a PCR tube. The polymerase and primers were added directly into this mixure. Therefore, five µL of 2X Maxima Hot Start Green PCR Master Mix (Thermo Scientific, California) as well as 0.2 µL of each, forward and reverse primer were used. This 10 µL PCR mixture was subjected to a PCR cycling protocol after manufacturer's recommendation. The remaining rest of the colony was transferred to an ON culture in 5 mL LB media plus proper antibiotic marker to extract the correctly cloned plasmid.

3.2.9 DNA Sequencing and sequence alignment

Sanger Sequencing was performed by Microsynth Seqlab in Göttingen. The DNA samples (plasmid and PCR products) were prepared according to the company's sample requirements. Sequencing results were provided by Microsynth Seqlab as fasta or ab1 files.
For a single sequence alignment SnapGene™1.1.3 function "align with a sequence

trace" was used. For Sequence alignment of multiple sequence files BioEdit Sequence alignment editor version 7.2.5 was used with is alignment function ClustalW alignment.

3.2.10 Primers

Primers used for this work are listed in the Appendix Table A.3. They were designed by using SnapGene™1.1.3. by considering general rules for primer design (bp length, GC amount, self hybridization). The molecules were synthesized by Thermo Fisher Scientific (Darmstadt) routinely using a synthesis scale of 25 nmol and the purification option of desalted.

3.3 Media and buffers

3.3.1 Media

In the following different media will be listed used to perform this study.

3.3.1.1 LB medium

Lysogenic broth (LB medium) was used for shake flask *E. coli* cultivation. It is a defined medium with following composition: 10 g/L tryptone, 5 g yeast extract and 10 g NaCl. To obtain solid LB plates, additionally 15 g/L agar was added. The medium was sterilized at 121 °C for 20 min. If needed, antibiotics were added to the media after sterilization. Because of their temperature instability they were sterilized by filtration (0.22 µm cut off). Antibiotics were used in following concentrations: Ampicillin: 100 µg/mL, kanamycin: 30 µg/mL and chloramphenicol: 34 µg/mL. When using two antibiotics simultaneously, concentrations were reduced by half to minimize selection pressure.

3.3.1.2 High cell density medium for E1α-E1β

For the fermentation of E1α-E1β in a high cell density culture a protocol from Prof. Dr. Kai Tittmann group (Dept. of Molecular Enzymology, Georg-August-Universität Göttingen) was kindly provided. Some components of the media needed to be sterilized separately before the fermentation. This was 30 g glucose in 100 mL ddH_2O, 4.08 g $MgSO_4$ in 50 mL ddH_2O and 66 g K_2HPO_4 in 100 mL ddH_4O. They were added under sterile conditions to the bioreactor before its inoculation with *E.coli* culture. In the bioreactor, 300 g Yeast extract, 3 g

Ammonium chloride and 1 mL antifoam were autoclaved in 5.75 L of ddH_2O. The feed medium was composed of 600 g yeast extract and 500 mL glycerol, dissolved in a final volume of 2 L ddH_2O.

3.3.1.3 SOC medium

Super Optimal broth with Catabolite repression (SOC) is prepared from Super Optimal broth (SOB) stock medium as shown in 3.5. The SOB stock medium was prepared as shown in Table 3.4.

Table 3.4: Preparation of SOB stock medium.

Yeast extract	0.5 %
Tryptone	2 %
NaCl	2 10 mM
KCL	2.5 mM
bring pH to 7	1 M NaOH

Table 3.5: Preparation of SOC stock medium.

Autoclaved SOB stock	97 % v/v
1 M MgCl (sterile)	1 % v/v
1 M Glucose	2 % v/v

3.3.2 Buffers

Different kind of buffers were employed to perform experiments. They will be listed in the following.

Table 3.7: Lysis and equlibration buffer for hPDC purification.

Phosphate buffer pH 7.4	50 mM
NaCl	200 mM
β-Mercaptoethanol	5 mM
PI cocktail (Roche, Basel)	1 tablet in 50 mL

Table 3.8: PBS buffer composition.

NaCl	140 mM
Phosphate buffer	10 mM
KCl	3 mM
adjust pH to 7.4	

Table 3.9: 50x TAE stock buffer.

Tris base	242 g/L
EDTA pH 8	50 mM
Acetic acid	57 g/L

Table 3.10: 4x laemmli buffer (SDS-PAGE loading buffer). Aliquot in 1 mL amount and store at -20 °C.

100 % glycerol	10 mL
β-Mercaptoethanol	5 mL
SDS	2 g
1% Bromophenol blue	1 mL
1.5 M Tris-HCl (pH 6.8)	4 mL
MilliQ water	to 20 mL

Table 3.11: 10x SDS-PAGE running buffer.

Tris base	250 mM
Glycine	1.92 mM
SDS	1 % w/v

3.4 Protein expression and purification

In the following sections, protocols for the processing of hPDC, GFP and gamS proteins will be presented.

3.4.1 Protein expression hPDC subcomponents

Proteins expressed in *E.coli* BL21 (DE3) and variants were treated according to Guo et al., 2017. A five mL culture of the protein to be expressed was grown ON in LB medium with appropriate antibiotics at 37 °C and 220 rpm shaking. The subcomponents E1α-E1β and E2-E3BP encoded on pETDuet-1-N6HisE1E1α-E1β and pETDuet-1-E2-N6HisE3BP were grown with 100 µg/mL ampicillin and for E3 encoded on pET28a$^+$-N6HisE3 34 µg/mL kanamycin was added. The inoculation of the expression culture in shake flask was carried out around noon in LB medium to a ratio 1:200. After growing at 37 °C and 220 rpm to an OD_{600} of 0.6 – 0.7 the cells were induced in the late afternoon by adding 0.4 mM IPTG. In addition, for E2-E3BP cultures 50 µg/mL lipoic acid after Wei et al., 2003 and for E1α-E1β cultures 50 µg/mL TPP was added simultaneously to induction. The expression took place ON at 15 °C, 200 rpm.

3.4.2 E1α-E1β expression in high cell density culture

A preculture of E1α-E1β was set up like described above. A 2 L glass bioreactor from Bioengineering was sterilized with medium (see section 3.3.1.2) and inoculated at ratio 1:15 with fresh preculture and addition of proper antibiotics. Temperature was kept at 37 °C and pH was maintained at 7.0 by addition of 10 % v/v ortho-phosphoric acid or 10 % w/v KOH. The stirrer speed was adjusted to maintain pO_2 at 25 % and in addition, the air flow was increased from 0.8 vvm to 2.6 vvm over time. When OD_{600} reached around 10 - 15 (after 3 to 4 h) feeding was started with a slow pump rate to maintain oxygen supply and adjusted with increasing OD_{600}. The feed composition is noted in section 3.3.1.2. When OD_{600} reached between 30 to 35 (around 5 to 6 hours after inoculation) temperature was decreased to 25 °C and induced by addition of 1 mM IPTG. Around 11 h after inoculation OD_{600} reached around 60 and the culture was harvested by centrifugation at 4 °C. The obtained cell pellet was frozen at -20 °C until further downstream processing.

3.4.3 Protein purification hPDC subcomponents

For the purification of expressed proteins all work was conducted on ice or at 4 °C. Cells were harvested at 6000 g and 10 min centrifugation. The supernatant was discarded and the pellet was washed twice in cold PBS buffer. After, pellet weight was determined and each of gram pellet was resuspended in five times volume binding buffer consisting of 50 mM Potassium phosphate buffer at pH 7.4, 150 mM NaCl, 5 mM β-Mercaptoethanol and protease inhibitor cocktail (from Merck KGaA, Darmstadt). Subsequently, cell disruption was performed by sonication with pulsed cycle at 20 % and amplitude setting at 50 %. The sonication process was run three minutes in 20 second on and 20 second off intervals and then chilled on ice for a few minutes, this process was repeated three times. The resulting cell debris was separated by clarification at 14000 rpm, 30 min at 4 °C. The supernatant was subjected to IMAC chromatography columns and purified according to the manufacturer's recommendation. Depending on the culture size, either His SpinTrap columns from GE Healthcare (0.5 to 3 mL) or gravity columns Protino Ni-TED 2000 from Macherey-Nagel (3 to 9 mL) were employed. For large scale purification Ni-NTA column with five mL bedvolume (GE Healthcare) was used with a FPLC system Bioline from Knauer.

Small scale purification via His SpinTrap columns from GE Healthcare with total protein capacity of 750 µg was performed according to manufacturer's recommendation. Usually, a 50 to 200 mL culture was loaded onto one column. The washing step was performed with 20 mM imidazol and elution with 500 mM in a total volume of 400 µL.

For intermediate scale purification Protino Ni-TED 2000 columns from Machery-Nagel with total protein capacity of 5 mg were employed. Routinely, 200 to 600 mL culture was loaded onto one column. Washing step was performed with 20 mM imidazole and elution with 250 mM imidazole in a total volume of 9 mL.

The large scale purification of hPDC subcomponents was performed by FPLC. Therefore, a cell pellet of 5 g or more was processed at a FPLC system Bioline from Knauer. There, the process control was realized through UV-light detector and conductivity detector. A HisTrap™HP 5 mL column from GE Healthcare was mounted with a maximal protein binding capacity of 40 mg. Its maximum pressure limit is 5 bar and it was equilibrated for five column volumes (cv) with filtered binding buffer: 150 mM NaCl, 50 mM Phosphate buffer at pH 7.4, 20 mM imidazol and 5 mM β-Mercaptoethanol. After, cells were disrupted, the lysate cleared and filtered like described above, the crude protein suspension was loaded onto the column and washed with binding buffer for another 15 cv

at 5 mL/min. Subsequently, an linear elution gradient was started over 20 cv at a flow rate of 2.5 mL/min to end up with a final imidazole concentration of 400 mM. The E1α-E1β proteins eluted around 25 to 55 % of the gradient progress of the elution buffer. Fractions of 5 mL volume were collected and cooled immediately. Whilst the first fractions contain some impurities, the later fractions were highly pure. After loading fractions on a SDS-PAGE to check for target protein presence and purity the fractions of choice were pooled together.

If necessary, a polishing step via size exclusion chromatography (SEC) using a Superdex 200 10/300 GL column from GE Healthcare following manufacturer's instructions was performed. Finally, buffer exchange and concentration of proteins were performed with 10 kDa cutoff Amicon® ultra centrifugal filters from Merck KGaA. The final storage buffer contained 50 mM potassium buffer, 150 mM NaCl, 2 mM DTT and 5 % glycerol at pH 7.4.

3.4.4 GFPuv expression and purification

A derivative of Green florescent protein (GFPuv) was expressed in *E. coli* BL21 (DE3) and treated using the following protocol adapted from Vessoni Penna et al. (2004). A 5 mL culture was grown ON in LB medium with the appropriate antibiotic at 37 °C and 220 rpm shaking. This starter culture was used to inoculate the expression culture in shake flask at a ratio of 1:150. After growing at 37 °C and 220 rpm to an OD_{600} of 0.5 – 0.7, the cells were induced with 0.4 mM IPTG. The expression took place overnight at 20 °C and 220 rpm (16 to 18 h). Cells were harvested by centrifugation at 6000 g, 4 °C for 10 min. The supernatant was discarded and the pellet was washed twice in cold PBS buffer (see Table 3.8). Pellet weight was determined and resuspended in a 1:1 ratio with binding buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0). Cell disruption was performed with bead beater FastPrep-24™Classic Instrument from MP biomedicals at 6 m/s, 3 cycles with 30 sec processing time. The obtained suspension was cleared at 12,000 g for 30 min at 4 °C. The supernatant was filtered through a PES 0.22 µm cut-off syringe filter and subjected to IMAC chromatography columns (Protino Ni-TED from Macherey-Nagel). The protocol was performed according to manufacturer's recommendation. Deviant, the washing step was performed with an imidazole concentration of 5 mM, the elution at 25 mM imidazole. Buffer exchange and sample concentration was performed with 10 kDa cutoff Amicon® ultra centrifugal filters from Merck KGaA. The final storage buffer contained 10 mM Tris-HCl, 10 mM EDTA at pH 8.0, and samples were stored at -80 °C until usage. Total protein content was measured through Bradford method (micro-assay procedure from BioRad, see section 3.5.2). An aliquot of the GFPuv

sample was loaded on a 12% SDS-PAGE (see section 3.5.1) to determine purity of the target protein. By using ImageJ, densiometry of the SDS-PAGE lanes was performed to calculate target protein concentration.

3.4.5 gamS expression and purification

The protein purification was adapted according to Sun et al. (2014). A 500 mL LB-culture was inoculated with 50 mL of preculture, grown at 37 °C, 220 rpm and induced at OD_{600} of 0.5 to 0.6 by addition 0.25 % arabinose. Induction was performed at 25 °C, 220 rpm for 4 h. After harvesting the cells the protocol for hPDC purification above was followed. In short, cell debris via sonication and Ni-NTA purification using Protino Ni-TED 2000 columns from Machery-Nagel after manufacturer's recommendation was performed. Additionally, 0.1 % Triton was added to the buffers and the washing step of the columns were performed with 25 mM imidazole. The final storage buffer contains: 50 mM Tris-Cl pH 7.5, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 2 % DMSO. Protein purity was checked by SDS-PAGE, gamS protein has a size of around 13 kDa.

3.5 Protein analysis methods

3.5.1 SDS-PAGE

In order to detect proteins and determine purification success, Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) was employed. The gels consisting of a stacking and an resolving gel (12 %) were casted with specifications shown in Table 3.12. Protein samples were prepared by addition of 4x laemmli buffer (see Table 3.10) to approximately 5 – 15 µg proteins for purified and protein mixtures, respectively. Samples were boiled at 95 °C for 5 min and subsequently loaded up to a maximum of 20 µL per well. A protein standard was used to detect protein size, therefore 4 µL of PageRuler Prestained Protein Ladder from Thermo Fisher Scientific was loaded. The electrophoresis chamber was filled with running buffer (see Table 3.11) and voltage was applied (via PowerPac™Basic Power Supply from BioRad). A two step protocol was used, firstly 10 minutes at 80 V and subsequently 40 min at 160 V. The proteins were dyed by addition of Bio-Rad Protein Assay Dye Reagent Concentrate and incubation for at least 30 min at an orbital shaker.
In case of protein sample matrix interfered with the SDS-PAGE, acetone precipitation was conducted before loading proteins on the gel. Accordingly, a sample was mixed with ice cold acetone in a 1:5 ratio. The suspension was rested on ice for 5 min, followed by centrifugation at 12000 rpm for 5 min. Carefully, the

supernatant was discarded and the pellet was resuspended in laemmli buffer and boiled as described above.
To determine purity or content of a mixed protein sample the open source image processing program ImageJ was used and the densitometry analysis was performed.

Table 3.12: Stacking and resolving gel specifications for the preparation of four to six SDS-PAGE.

component	resolving gel (12 %) in µL	stacking gel in µL
40 % acrylamide (29:1)	3750	1000
1.5 M Tris-HCl, pH 8.8	3750	-
1 M Tris-HCl, pH 6.8	-	1250
H_2O	14535	7590
10% SDS	300	100
TEMED	15	10
10% APS	150	50

3.5.2 Bradford assay

To determine protein concentration, the microtiter plate protocol of the Bio-Rad protein assay was used. The lower boundary for total protein detection is 50 µg/mL. This colorimetric assay is read out at 595 nm spectrophotometrically, where the dye and the protein sample form a complex. As reference protein Bovine serum albumin (BSA) was employed and a standard curve was established for every measurement. The assay was performed according to manufacturer's instruction and in triplicates to minimize handling error.

3.5.3 hPDC activity assay

The molecular weight of the hPDC subcomponents were used for the ratio calculation of the complex. Following molecular weights were used: E1α-E1β heterotetramer 154.4 kDa, E2 58.8 kDa, E3BP 48.3 kDa and E3 homodimer 106.8 kDa. The assembly of the full hPDC is based on a stoichiometry of 30xE1, 40xE2 subunits plus 20xE3BP subunits for core assembly and 6xE3 subunits. When calculating the ratios based on the molecular weight, this equals E1:E2+E3BP:E3 of 1.396:1:0.193. For the *in vitro* activity assay measurements E1 and E3 subunits were employed in excess over the core, thus the ratio here

equals E1:E2+E3BP:E3 1.515:1:0.816. For *in silico* simulations the ratio equals E1:E2+E3BP:E3 0.3:1:0.1 (in molecules: 18:60:6).

Activity assay to determine overall functionality of hPDC was performed after Guo et al., 2017. Firstly, the purified hPDC subcomponents were assembled to the full complex. Therefore, for each assay 6.79 µg E2-E3BP was incubated with 5.54 µg E3 for 5 min at RT. Subsequently, 10.29 µg E1α-E1β was added and the volume was adjusted to 20 µL with 50 mM potassium buffer pH 7.4 and 150 mM NaCl. The amount of 170 µL of reaction buffer (see Table3.13) was added to 96 Well UV-Star® Microplatte from Greiner Bio-One and incubated with 20 µL of enzyme mixture for 10 min at 37 °C. By addition of 10 µL pyruvate (final concentration 2 mM) the reaction was initiated and monitored via a Multiskan Spectrum spectrophotometer (Thermo Scientific) immediately at 340 nm, 37 °C for at least 20 min. The initial slope was used to calculate enzyme activity whereas one unit describes the conversion of one µmol substrate per minute. Unit calculation proceeded as shown in equation 3.1. A standard curve of NADH over different substrate concentrations was recorded. The obtained slope s_{NADH} with unit A340/µ was determined. Next the initial slope (5 min) of the enzyme reaction s_e was determined with unit A340/min. To obtain the specific activity (U/mg) Units were divided by mg of protein.

Table 3.13: Overall hPDC activity assay reaction buffer pipetting instruction. Table shows amount for one activity assay performed in a 200 µL volume.

component	volume µL	final mM	stock mM
2x buffer A	100	50	100
TPP	8	2	50
$MgCl_2$	8	2	50
NAD^+	10	2.5	50
CoA	10	0.5	10
MilliQ water	34	-	-
buffer A:			
Phosphate buffer pH 7.4		50	
NaCl		150	

$$Unit = \frac{s_e}{s_{NADH} * V} \tag{3.1}$$

s_e	initial slope of enzyme reaction	[A340nm/min]
s_{NADH}	slope of NADH standard curve	[A340nm/µM]
V	reaction volume	[mL]

3.5.4 Dynamic light scattering

This technique is used to determine size or size distribution of proteins in a solution. Since this instrument is not available at the TUHH, the experiments were carried out at the EMBL Sample Preparation and Characterization Facility in Hamburg.
Protein samples were diluted with final storage buffer (50 mM potassium buffer, 150 mM NaCl, 2 mM DTT and 5% glycerol at pH 7.4) to 1 mg/mL and filtered shortly before the DLS measurement with 0.22 µm Ultrafree® MC GV 0.5 mL Centrifugal Filter Unit (Merck Millipore, Massachusetts, USA). After, 46 µL of a protein sample in a disposable cuvette was measured at a DynaPro® NanoStar® (Wyatt Technology, Santa Barbara, Canada) 30 times at 25 °C. The data was processed by the software Dynamics V7.0 software (Wyatt Technology, Santa Barbara, Canada).

3.6 Cell free protein synthesis (CFPS)

3.6.1 Cell extract production for CFPS

The protocol published by Sun et al. (2013) was used to prepare an *E. coli* based cell extract which will be used for CFPS. The procedure requires three days in a row and to minimize possible handling error, working in a group of two people is recommended. The crude cell extract was prepared from *E. coli* Rosetta strain. A 4 L batch was prepared accordingly to the protocol and resulted in around 11 mL of 28 mg/mL extract enough for around 2200 single 10 µL reactions. The extract was aliquoted and stored at -80 °C.
Next, the produced extract needs to be calibrated to find the optimal concentrations of potassium (20 to 40 mM), magnesium (0 to 6 mM) and DTT (0 to 3 mM). Therefore, different CFPS reactions were carried out like described in the protocol. As DNA template eGFP was used on a pET-duet-1 backbone, promotor region called OR1-OR2-Pr and a UTR1 were cloned in front of the eGFP gene, the sequence is given in the Appendix (A.1.1). The reactions were carried out as described in the following and the readout happened with the flow cytometer described in the section 3.8.
The experimental set up of a TX-TL reaction consists of three parts. One part is the extract that, together with the buffer, makes up 75 % of the CFPS

reaction volume. The second part is the buffer which contains among others the amino acids, energy source and nucleotides. The third part consists of DNA template and possible other additives. If necessary the reaction volume was adjusted to 10 µL with MilliQ water. An excel spread sheet was used to calculate volumes and concentrations. As DNA template either a plasmid was used at a concentration of 10 nM or LET with 16 nM if not otherwise stated. The ingredients were mixed together at RT, spun down shortly and incubated at 29 °C for up to 8 °C. Until the subsequent readout, usually few hours later, the produced protein samples were stored at 4 °C.

3.6.2 LET production

Production of LETs were performed at two different paths. As a first approach, LETs were generated by linearization of a plasmid (OR1-OR2-Pr-eGFP plasmid) through proper restriction enzymes. Therefore, sites were chosen that allow to leave a 3' and 5' overhang to the gene of interest between 250 to 500 bp (Sun et al., 2014). The digested plasmid was loaded on an agarose gel, the band with the expected size was excised from the gel and purified with the PCR and gel clean up kit from Machery Nagel. Different elution methods were tested to achieve highest possible LET concentrations.
As this method depends on a plasmid source, tedious cloning technologies that are not compatible with automation preceded. To tackle this drawback, another method was tested to generate LETs from scratch. Therefore, an assembly PCR is chosen due to its flexibility and feasibility with automation. Based on Stemmer et al. (1995) and Sequeira et al. (2016) a two-step PCR reaction was established and optimized based on TerMaat (2011). The gene of interest plus upstream and downstream regulatory parts where fragmented to overlapping 40mers, this was done by using the online tool of Rouillard et al., 2004. Each oligo overlapped with the neighboring oligo by around 20 bp. Like primers, the oligos were purchased at 25 nmol and desalted from Thermo Fisher Scientific (Darmstadt)(see Appendix Table A.1). For the first assembly PCR reaction, all oligos were used in equimolar concentration and dissolved in an oligo master mix in which each primer was present in 0.5 µM to assemble the full length sequence. In the second PCR reaction only the two outermost primers were used with a stock solution of 5 µM to amplify the final product. For both PCR reactions the Phusion™High-Fidelity DNA Polymerase from Thermo Fisher was used. The pipetting and cycling instructions for both reaction can be found in Table 3.14 and 3.15. Since this LET production is organized modular, targeted mutations could be introduced by exchanging only one set of oligos in the MM by modified oligos containing a point mutation. Therefore, the aa to be exchanged was usually

modified by studying the codon usage table and switching one or two bases for one aa change. It is possible to mutate several aa on one oligo simultaneously. To verify presence and purity of the generated LET an agarose gel was employed. A list of all oligos and modified oligos can be found in the Appendix (A.1).

Table 3.14: Pipetting instruction for the two-step PCR to generate LETs. Phusion™High-Fidelity DNA Polymerase from Thermo Fisher was used.

Reaction	Component	volume
1st PCR	Polymease 2x	7.5 µL
	primer mix (0.5 µM)	1.5 µL
	mutation primer for (1.5 µM)	0.5 µL
	mutation primer rev (1.5 µM)	0.5 µL
	proceed with mutation primer for	0.5 µL
	proceed with mutation primer rev	0.5 µL
	MilliQ water	fill to 15 µL
2nd PCR	Polymerase 2x	30 µL
	primer 3' (5 µM)	6 µL
	primer 5' (5 µM)	6 µL
	DNA (1st PCR)	1.2 µL
	MilliQ water	fill to 60 µL

Table 3.15: Cycling instruction for the two-step PCR to generate LETs. Phusion™High-Fidelity DNA Polymerase from Thermo Fisher was used. Expected fragment length is close to 1.3 kbp.

	1st PCR			2nd PCR		
	Temp.	Time	Cycles	Temp.	Time	Cycles
Initial denat.	98 °C	15 sec		98 °C	15 sec	
Denaturation	98 °C	1 sec		98 °C	1 sec	
Annealing	58 °C	10 sec	27 x	68 °C	10 sec	35 x
Extension	72 °C	20 sec		72 °C	20 sec	
Final extension	72 °C	120 sec		72 °C	120 sec	
Storage	4 °C	hold		4 °C	hold	

3.7 Bio-layer interferometry (BLI)

BLI measurement was performed with the instrument Octet® RED96 System from Pall FortéBio. It is an optical, analytical technique to measure biomolecule interaction without labeling. Eight distinct spectrophotometers are built in, so different test conditions could be investigated simultaneously (Shah and Duncan 2014). Appropriate Biosensors were purchased from the company, for His-tag based capturing the Anti-Penta-HIS (HIS1K) biosensors were used. HIS1K consists of a highly specific and affine Penta-His antibody which binds to the His-tag of recombinant proteins, even in a crude protein lysate. First, the assay was set up with the data acquisition software 10.0. The assay steps (loading, association and dissociation) and parameters like incubation time, temperature and shaking speed were defined. Until not otherwise stated, a general assay was performed at 30 °C, 1000 rpm shaking and consisted of following order: 1. baseline step, 2. ligand loading to the biosensor, 3. second baseline step, 4. analyte association, 5. baseline dissociation. Usually, one biosensor was sacrificed as negative control. Next, the assay was prepared in the wet lab. All necessary dilutions of loading sample and analyte sample were pipetted into black, flat bottom 96 well plates (Greiner Bio-One), assay volume per well was 200 µL. The assay was run fully automated and the progress could be tracked online. Finally, the measured raw data was evaluated with the Data analysis HT 10.0 software from Pall FortéBio. Here reference senors or reference samples were referenced, curves were smoothed and association and dissociation rates were determined using appropriate binding model. This was the case for matching fitting curves with appropriate X^2 and R^2 values.

Following standard conditions were applied to BLI measurements if not stated otherwise:
In case of hPDC core components utilized as analyte: Loading protein concentration of E1: 27.5 nM, loading protein concentration of E3: 48 nM, loading duration: 1000 s, analyte protein concentration: 80 nM, association step duration: 250 s, dissociation step duration 10000 s. Reference sample was GFPuv.
In case of hPDC core components utilized as ligand: Loading protein concentration: 5 to 40 nM, baseline step: min. 100 s, loading duration: 250 s, analyte protein concentration of E1: 2 to 50 nM in two dissociation steps, analyte protein concentration of E3: 2 to 50 nM in three dissociation steps, association step duration: 400 s, dissociation step duration: 500 s.

3.8 Flow cytometry

The possibility to measure fluorescent signals was offered by the CytoFLEX system of Beckman Coulter. Flow cytometry is used to measure cells/particles and determine their relative size and shape but also detect fluorescence based biomarkers. In this study, CFPS represents the base to express GFP, and some of its variants, as model proteins. To still enable the CyctoFLEX working mechanism, the non-present cells were mimicked by non-functionalized 0.909 µm polystyrene beads from micro particles GmbH (Berlin) which triggered measure events according to Wurm et al. (2019).
The fluorescent CFPS samples were measured in crude lysate background. Since the usual sample volume of CFPS is 10 µL, the volume of the protein crude solution was not enough for a CytoFLEX measurement. Therefore, the samples were diluted 1:4 in sheath fluid. Into this solution also the polystyrene particles were added to a final dilution factor of 1:20000 from a 5 % w/v stock. A blank measurement was performed with beads in only sheath fluid for the single bead population gating strategy. The reading was done with 50000 events and a flow rate of 30 µL/min. The gain settings were adjusted to 1000 FITC, 52 PB450 and 27 KO525 and 95 PE. The threshold was set manually to FSC-Height 10000. Evaluation of the raw data was performed with the built in software CytExpert as well as calculation spread sheet. Positive and negative controls (PC and NC) were employed to detected a homogeneous single bead population required for experiment evaluation. Firstly, the population representing single beads was gated. After, the fluorescence signal of this population was used in different channels against the total counts to determine fluorescence intensity of each sample. To detect for GFP, the FITC channel (ex: 488 nm, em:525/40 nm) was used, to detect for variants that shifted their wavelength spectrum to lower (blueish) KOH channel was employed (ex: 405 nm, em: 525 nm/40 BP) and to detect variants with a wavelength shift to higher wavelengths (yellowish) PE (ex: 488 nm, em: 585 nm/42 BP) was used. If not stated otherwise, the confidence interval was indicated as error. It was calculated using variation of coefficient by using a 99% confidence level ($\alpha = 0.01$), the standard deviation and the population size.

3.9 Simulation of biochemical network via ODE model

hPDC network reactions were simulated based on ODEs with the open source software COPASI. A kinetic model of mammalian PDC with a set of parameters were obtained from Modak et al. (2002); Zeng et al. (2002). The species

concentrations were plotted via the time-course simulation with implemented numerical Livermore Solver for Ordinary Differential Equations (deterministic LSODA) solver. Furthermore, the parameter scan option with the time course task was employed to make a scan over varying initial concentrations of species or global quantities.

3.10 Assay automation

Automating process steps can save time and enhance reproducibility. For this work the Biomek 4000 laboratory automation workstation from Beckman Coulter was available. It is equipped with different pipetting tools enabling a liquid transfer in the range of one to 220 µL. Furthermore, a gripper tool for the transfer of plates can be selected. Two positions at the deck are connected to a thermostat "Minichiller 300" from Huber to allow tempering of plates and tubes in the range of -20 °C to 80 °C. One position is connected to a thermoshake from Inheco allowing to control temperature and shaking of plates and tubes. Also, a thermocycler "TRobot" from Biometa is integrated to the deck enabling automated PCRs. Consumables and other labware to run assays at the liquid handling software were purchased from Beckman Coulter.
The device was controlled by a build in Biomek software. This software was utilized to program methods by specifying single steps and adjusting different parameters. First, the instrument setup was adjusted specifically for the assay to be conducted. Different labware like pipet tips, tube holder, reservoirs or plates were assigned to selected positions. Next, single steps were specified, which were either pipetting, transfering or incubating steps. Within each step volumes, place or time was specified. For each assay the liquid type was adjusted by defining flow velocity, prewetting tips or touching the side of wells/tubes to make sure also small volumes or viscous liquids were transferred properly and no bubbles were created.

Assays to be automated were first tested by manual handling and evaluation. Subsequently, the method instructions were transferred to the Biomek software. For each assay specific adjustments were conducted to optimize the assay workflow. To enable flexibility regarding number of reactions, loops were programmed via the software, or a txt-file specifying source and destination as well as volumes of each step was read in.

Chapter 4

Novel protein manipulation, expression and detection workflow

In this chapter a novel and automated workflow for fast protein modification and screening will be presented. This comprises the establishment of flexible and fast DNA generation via assembly PCR, protein expression in cell free manner and the detection of enzyme characteristics. To provide a proof of concept for this automated and connected process chains, GFP will be utilized as role model. Moreover, a final section will show first efforts to adopt hPDC into the presented work flow. In future, this project-part will be embedded in an algorithm based automated and semi-rational protein optimization loop.

4.1 Introduction

The field of protein engineering plays an important role for industrial biotechnology applications and comprises a broad variety of methods and techniques (Steiner and Schwab 2012). Screening for enzymes with new or enhanced function is a key node which can be accomplished rationally (low throughput) or randomly (high throughput). Combing these concepts makes use of both advantages, i.e., smaller library size, less screening rounds and thus, less labor time (Steiner and Schwab, 2012). By using the large number of so far elucidated protein structures an array of promising mutation points can be predicted with computational methods. Screening them experimentally for their properties involves tedious cloning work, expression of the protein and monitoring its function. Therefore, flexible and fast methods are demanded.
Producing linear DNA templates bypasses cellular cloning, and can be simply

accomplished by a time effective PCR. These products can be expressed in cell free open environment within few hours and modified or monitored easily. In this way, each "evolution round" is accelerated and only few rounds are required since rational design thoughts narrow the library size.

In this chapter the idea of an automated, feedback guided protein optimization loop will be presented. This loop consists of several parts shown in scheme 4.1. Firstly, rational protein design will lead to the identification of promising sequences sites which are correlated to enzyme function. By using this kind of computational method the number of promising mutations to screen is reduced and thus only a small library size, in contrast to the random mutagenesis approach, needs to be realized. The second step consists of automated and fast production of the enzymes and their variants. This is going to be realized by *de novo* gene synthesis and coupled cell free protein synthesis (CFPS). Subsequently, the produced enzyme will be evaluated which leads to the third step: intermediate throughput screening. Adequate sensing methods with the potential to be automated need to be built up. Lastly, data evaluation techniques

represent the crucial point for closing the loop, as the data gathered in step three must be evaluated and judged, generating the starting point for the next cycle of the loop. These steps are to be repeated till the targeted enzyme enhancement is accomplished. The tasks performed in the framework of this thesis will focus on the enzyme production and the screening method.

4.2 Establishing cell free protein synthesis and the generation of linear expression template

CFPS offers many advantages over conventional cell based protein expression procedures. This can be i.e. economy of time and flexibility. Therefore CFPS represents a promising approach for fast enzyme production (Schinn et al., 2016). Due to the open environment of CFPS, the protein synthesis process, even of toxic proteins, can be directly manipulated and monitored (Carlson et al., 2012). To establish CFPS in this laboratory a class of well-studied fluorescent proteins, namely GFP variants, will serve as role model. For the production a cell lysate containing all necessary components for the translation and transcription machinery (coupled mode) will be employed. In the following section the establishment of a cell lysate for CFPS is presented.

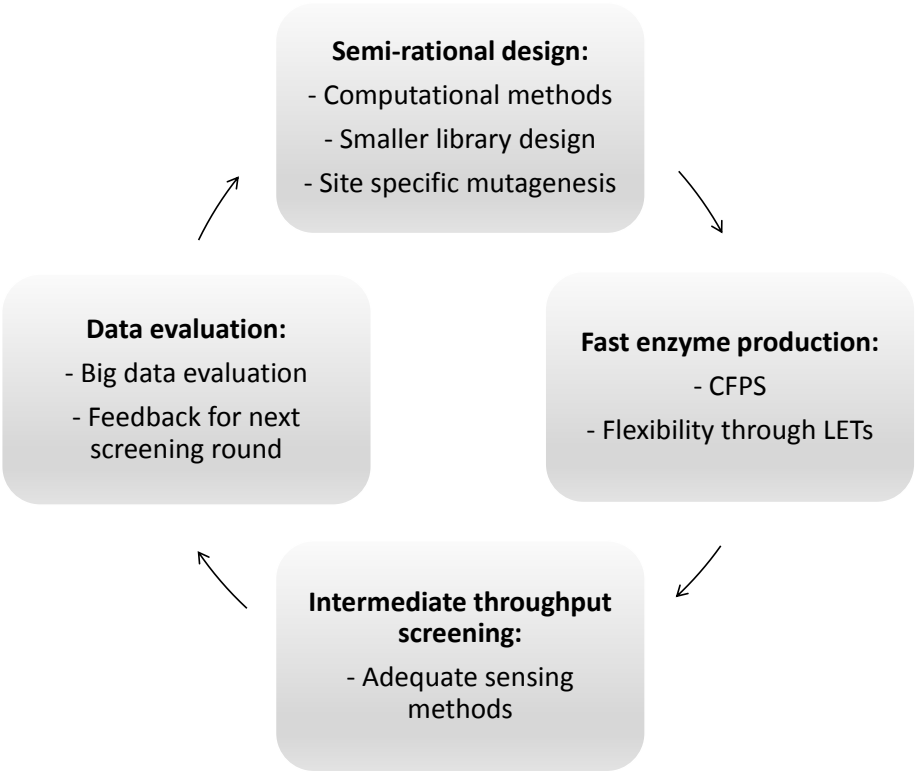


Figure 4.1: Scheme of a feedback guided protein optimization loop pursued within this study. Within the framework of this thesis, the enzyme modification and production work flow as well as the detection method will be treated.

4.2.1 Home made *E. coli* based cell lysate

Different protocols for the production of cell lysates for a variety of systems are available. *E. coli* was chosen as a production system, since it is the host for the conventional, cell based production in this laboratory, and the codons of the genes are optimized for this host.

Two well studied protocols were tested to serve as the base for CFPS establishment in this laboratory. One protocol is based on Kigawa et al. (2004) and represents a T7 RNA polymerase based S30 cell lysate. Here, next to the paper also instructions from the EMBL Protein Expression and Purification Core Facility were considered which are based on Kigawa et al. (2004) and promise 500 µg/ml of protein within two hours. The second, very detailed protocol is published by Sun et al. (2013) and called TX-TL cell free expression system. It is an endogenous core RNA polymerase and housekeeping sigma factor 70 *E.coli* coupled transcription and translation based lysate and promises good yields (750 µg/mL) and low costs.
By following the instructions accurately, this lysates were produced with the help of Dr. Anibal Jose Mora Villalobos and calibrated according to the instructions. To test the performance of the TX-TL lysate a plasmid as described in the next section was created.

4.2.1.1 S30 T7 RNA polymerase based cell lyaste

The lysate after Kigawa et al., 2004 was tested by using plasmids encoding for hPDC subcomponents (pETDuet-1-N6HisE1α-E1β, pETDuet-1-E2-N6HisE3BP and pET28a$^+$-N6HisE3). A plasmid without a gene insert was used as negative control. Creatine phosphate was used as substrate and the reaction was run for two hours at 37 °C. The protein samples were loaded onto a SDS-PAGE, but no target protein was visible. The lanes which should include a target protein showed the same pattern like the negative control (NC) (data not shown). Because of this negative result, the TX-TL lysate presented in the following subsection 4.2.1.2 was set as alternative lysate for protein production.

The T7 RNA polymerase lysate was omitted because no target protein was detected on a SDS-PAGE. Firstly, the crude cell free background could interfere with the protein of interest bands. Furthermore, large molecular weight proteins coexpressed on one plasmid were used for a first testing. Resources need to be shared for the coexpression of two proteins, and additionally large molecules are challenging as well. The chosen proteins of interest were not the right choice for the first time establishment and detection of proteins with the T7 based lysate.

However, after gaining more experience in the field CFPS, this lysate stored for around 1.5 years at -80 °C was tested again with the simpler eGFP protein facilitating the detection. In this case target protein production could be verified and compared to the TX-TL lysate (see Figure 4.2). Reaction conditions were same for both lysates: 30 µL scale, incubation at 29 °C for 2 h at 300 rpm.

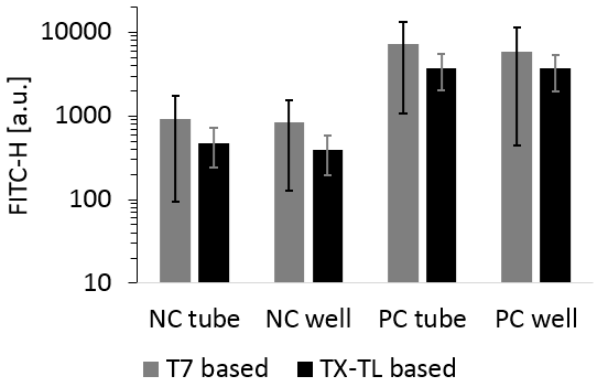


Figure 4.2: Comparison of fluorescence intensity of eGFP produced by T7 RNA polymerase lysate vs. constitutive promotor OR2-OR1-Pr-UTR1 in TX-TL reactions. Same experiment was performed in 30 µL scale once in a tube and once in a well of a 96 well plate. Error bars represent SD, number of events are 50000.

Table 4.1: Ratio of signal intensity PC to NC of T7 vs. TX-TL lysate.

	ratio
T7 tube	7.8 ± 1.2
T7 well	7.1 ± 1.3
TX-TL tube	7.8 ± 0.7
TX-TL well	9.4 ± 0.7

The T7 based lysate exhibited a more crude background compared to the TX-TL lysate. This was seen in the modified flow cytometry method for the fluorescence intensity detection (presented in the next section, 4.3). The baseline level for the NC is elevated in comparison to TX-TL. The population detected was very scattered and thus caused large error bars when calculating the standard deviation (see Figure 4.2). Nevertheless, the performance of the stored T7 lysate (and energy solution) equaled the TX-TL performance when calculating the ratio of PC to NC (see Table 4.1). Also, for the reaction scale of 30 µl a transfer from 1.5 mL Eppendorf tube to a well of a 96 well plate was trouble-free. For the TX-TL lysate a slight increase of performance using the 96 well plate was successful. This might be due to a better surface area to volume ratio of the well, and thus better air input in that scenario (Fuchs et al., 1971). This effect might be more pronounced for larger reaction volumes.
By optimizing magnesium, potassium and plasmid concentration, protein yield obtained from T7 based plasmid might be enhanced (Shin and Noireaux, 2012; Sun et al., 2014).

4.2.1.2 TX-TL endogenous *E. coli* transcription-translation lysate system

This home made cell lysate was prepared after Sun et al. (2013). It utilizes an endogenous *E. coli* transcription-translation (TX-TL) mechanism and therefore, the DNA template should contain optimized regulatory parts. In the promotor region the lambda repressor Cro promoter OR2-OR1-Pr and the UTR1 (untranslated region) containing the T7 g10 leader sequence for highly efficient translation initiation is proposed, for termination a T500 transcription terminator (Shin and Noireaux, 2010). To meet these requirements, a plasmid was constructed with the help of Dr. Jose Anibal Mora Villalobos. The plasmid pETDuet-1 was used as a backbone. Its promoter region was exchanged by the OR2-OR1-Pr and the UTR1. In addition, eGFP was inserted to the plasmid. The resulting construct is named pETDuet-1-OR2-OR1-Pr-UTR1-eGFP, short GFP-PC and the sequence can be found in the Appendix section A.1.1. This plasmid was successfully employed in CFPS to produce eGFP and thus is used as reference (positive control) in future. The TX-TL lysate was capable of producing eGFP (see Figure 4.4B), and therefore was set as the lysate for further studies.

In order to perform CFPS, a suitable DNA template needs to be provided. In general the template can be a circular plasmid or a linear sequence. The protein synthesis from a plasmid results in higher protein yield compared to linear expression template (LET), but LET offers flexibility and is less time intense to be constructed (Schinn et al., 2016). In this study flexibility as well as an automated work flow is essential, therefore LETs are to be used. Nevertheless, to establish and characterize the new method of CFPS at the laboratory, a plasmid based characterization shall follow. Since fluorescent proteins are quite easy detectable they were set as the role protein to study CFPS. The detection and quantification of the fluorescent proteins are described in section 4.3.

4.2.2 Batch to batch variations

During the time course of this work two different batches of cell lysate were required and multiple PC plasmids were purified. Thereby batch to batch variations were revealed and investigated more carefully (see Figure 4.3 A).

Positive control reactions varied in the fluorescent signal, one possible explanation for this might be DNA template quality (Sun et al., 2013). Plasmids were obtained in small batches described in section 3.2.1. Several batches were needed to characterize TX-TL reactions. During this study different yields for the positive control reaction led to a closer plasmid investigation. Different plasmid

batches were loaded onto an agarose gel to check for impurities whereas some showed a smear at lower bp size. This indicates RNA impurity. Subsequently, a TX-TL reaction was carried out with the same conditions for DNA template with and without contamination. In both cases eGFP is expressed, whereas the fluorescent signal for the DNA template without contamination is higher.
Another reason for varying protein yield might be RNAse contamination. RNAse degrades RNA and thus no template remains to be translated to protein. Since RNAse are omnipresent, and there is no RNAse free working station available, it can be assumed that CFPS reaction performance is decreased (Scheele and Blackburn, 1979).

For the two lysate batches a large variation is revealed. The second batch shows only around half activity of the first batch. This shows the need of internal standards and relative comparison. Although both batches were carried out with the same carefulness, there must be either one step at the protocol that was insufficient or one chemical is limiting. Since there are a lot of steps for lysate preparation, energy supply preparation and buffer preparation with very sensitive chemicals, it is difficult to track back where the limiting factors is.

4.2.3 Reaction scale

The volume of one TX-TL reaction recommended by Shin and Noireaux (2010) is 10 µL. For some analysis techniques this volume is not sufficient, so the scaling up by a factor of two, four and eight was investigated. As shown in Figure 4.3 B and C, the 10 µL reaction exhibited the highest florescence signal. The larger the volume gets, the lower the absolute fluorescence intensity was for the investigated samples. The fluorescence intensity of the NC also decreases with increasing reaction volume, but in comparison to the signal height its influence is low. The NC was referenced from the PC, and the 10 µL reaction was set to 100 %. Therefore, the 20 µL reaction showed 96 % fluorescent intensity, the 40 and 80 µL volume reactions were to 82 % and 84 %, respectively.
These samples were visualized in their reaction tubes before they were measured. In Figure 4.3 C, it can be seen clearly that the 10 and 20 µL reaction volume is homogeneously glowing, whereas a gradient with high fluorescence intensity at the air surface area to low fluorescence intensity at the bottom of the tube was visible for the 40 and 80 µL volumes.
The lower CFPS performance of higher reaction volume might be reasoned to limitation by diffusion. If the surface area to volume ratio gets lower, the oxygen input gets inefficient as described already above (Fuchs et al., 1971). By continuous agitation or change of geometry this limitation might be improved. The oxygen limitation is probably concerning both, CFPS performance and

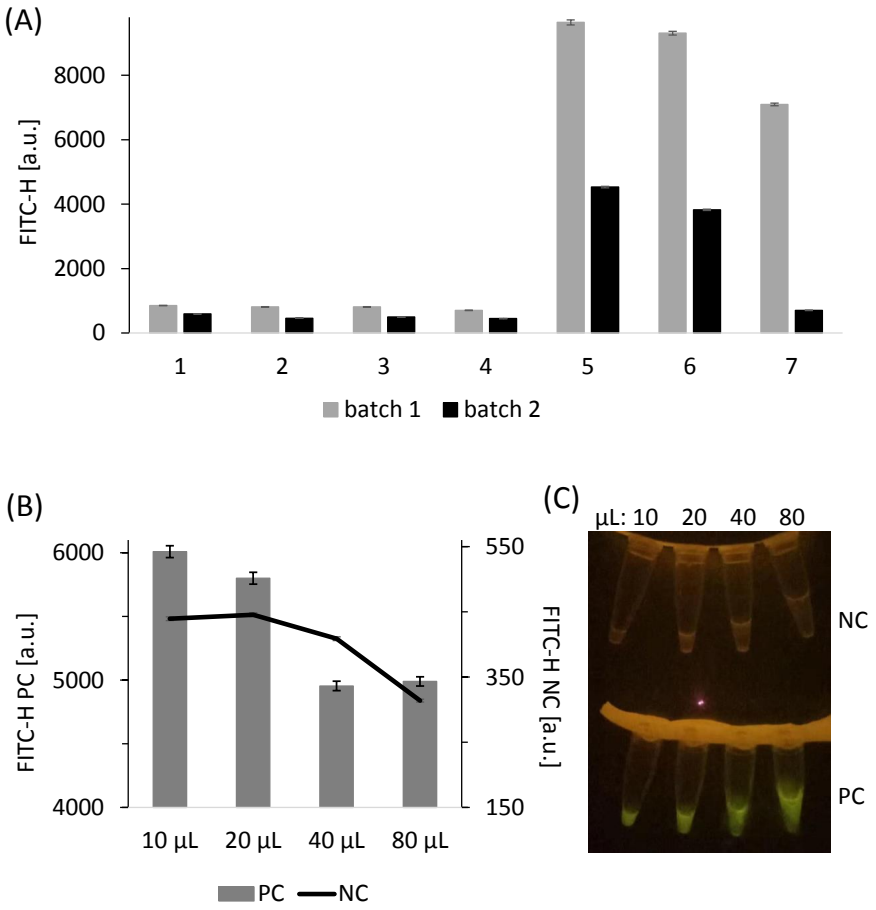


Figure 4.3: A: Comparison of TX-TL cell lysate batches. Three different plasmid batches were used to study TX-TL batch 1 (grey bars) to batch 2 (black bars) variation, as well as plasmid batch to batch variation. 1, no plasmid; 2, empty plasmid a; 3, empty plasmid b; 4, empty plasmid c; 5, GFP-PC a; 6, GFP-PC b; 7, GFP-PC c. Error bars indicate 99% confidence interval. B: Comparison of TX-TL reaction scale of 10, 20, 40 and 80 µl. NC: no plasmid, PC: eGFP. The median of 50000 events measured by flow cytometry in FITC-H channel is visualized. Error bars represent a confidence level of 99 %. C: Picture represents evaluated samples of B before their measurement, visualized at a blue light source. Upper part represent NCs whereas lower part show PCs.

oxygen dependent fluorophore formation of GFP (Teerawanichpan et al., 2007). In so far published CF protein expression performed by TX-TL lysate, routinely a volume of 5 to 20 µL reactions were used (Marshall et al., 2018; Rustad et al., 2018). As for this work no further scale up is required for the high throughput reactions, no work was continued in this route.

4.2.4 Investigation of DNA templates

4.2.4.1 Plasmid characterization

Although described by Shin and Noireaux (2010) that TX-TL S30 lysate can work with exogenous T7 RNA polymerase, the attempt to run such a CF reaction with the TX-TL lysate showed no detectable positive results. A T7 promoter based plasmid was employed with 8 U of T7 RNA polymerase. As a positive control a constitutive promotor (OR2-OR1-Pr-UTR1) GFP-PC was employed. After eight hours of incubation at 29 °C, the fluorescent signal obtained from the T7 promoter based reaction equaled the NC (no plasmid) as can be seen in Figure 4.4. The slight background signal seen in the NC could be due to a surface effect at the liquid-air interface, or a component in the E. coli lysate has its absorption maximum near that of the light source. The constitutive promoter based PC reaction in combination with the TX-TL lysate exhibited a clearly detectable fluorescent signal, thus CFPS was successful.

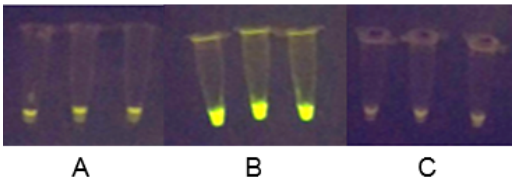


Figure 4.4: Comparison of eGFP yields with T7 based promoter to constitutive promotor OR2-OR1-Pr-UTR1 in TX-TL reactions. A: T7RNA polymerase + IPTG + T7 promotor. B: constitutive promoter and no additions. C: NC (no plasmid). This reactions were performed in triplicates. The light source (470 nm, blue) is able to visualize eGFP with its ex. 488 nm (blue) and em. 509 nm (green).

Based on these results, it was decided to continue with a DNA template containing a constitutive promoter by utilizing the TX-TL lysate. This combination was set as standard for further work. Nevertheless, there is still potential to run the TX-TL lysate employing a T7 based promoter as described by Shin and Noireaux (2010), however this route requires more optimization work.

The amount of DNA template in the CFPS influences the final protein yield. To test for saturating DNA amount, plasmids with varying concentrations were tested in CFPS reaction. A range of concentrations were chosen: 10 nM, 20 nM, 40 nM and 80 nM (data not shown). After 8 h run time at 29 °C all signals were similar. This demonstrates that already 10 mM plasmid is sufficient for

plasmid based TX-TL reaction and is in good agreement with Sun et al. (2014). Most probably other limiting factors would need to be overcome to convert more DNA into protein. The concentration of 10 nM is set as standard for following experiments.

Although clear fluorescence signal of the CFPS sample is detected, no expressed proteins could be identified on a SDS-PAGE (Figure 4.5). PC and NC show the same lane pattern, no additional eGFP band at around 27 kDa is visible. The dye used to visualize the SDS-PAGE is called Roti® blue quick from Carl Roth GmbH + Co. KG. The vendor specifies a sensitivity of protein amount detectable per SDS-PAGE lane, which is 10 ng/lane. For the presented SDS-PAGE 2.5 µL CFPS reaction mixture was loaded into each well, this calculates to theoretical eGFP concentration of less than 4 µg/mL because it is not detectable. On the other hand, the eGFP band might be overlaid with the other protein bands in the cell lysate since a yield of 0.75 mg/mL for eGFP is described by the author of the CFPS protocol (Sun et al., 2013). Hence, for the detection of expressed protein a functionality testing is inevitable.

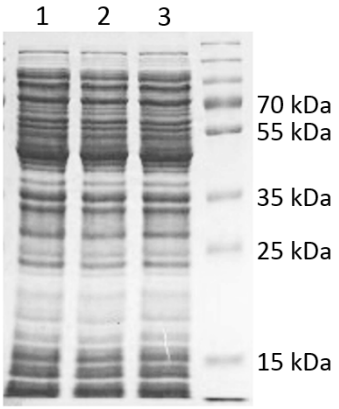


Figure 4.5: SDS-PAGE of CFPS samples after 8 h expression time. A volume of 2.5 µL is loaded on a SDS-PAGE after acetone precipitation. PC GFPuv size is around 27 kDa. 1, NC (no plasmid); 2, NC (empty plasmid); 3, PC (GFP-PC).

For the eGFP quantification the flow cytometer method was applied. Here, the PC from TX-TL reactions usually exhibited a signal ranging from 5000 to 9000 FITC-H units when NC is subtracted. In addition, the reactions are diluted 1 to 4 with sheath fluid and beads. When calculating eGFP yield by using crude GFPuv calibration curve (see section 4.3, Figure 4.11), this equals a CFPS yield of eGFP of around 40 to 90 µg/mL (for 27 kDa protein this results in 1.38 to 3.1 µM). This yield is lower than described in the protocol of Sun et al. (2013) where around 750 µg/mL eGFP is reached. Further optimization approach could target to the energy balance of the CFPS system. Produced inorganic phosphates that originates from dephosphorylation reaction for the production of energy may inhibit some enzymatic reactions (Kim and Swartz 1999; Whittaker 2013). For TX-TL reaction 3-PGA is the primary energy donor. Increasing protein yields produced by TX-TL lysate might be possible by addition of Maltose or Maltodextrin. Caschera and Noireaux (2014) reported that energy

recycling can be performed by recycling inorganic phosphate accumulating in the CF reaction. Maltose as well as Maltodextrin can activate metabolic reactions involved in glycolysis performed by endogenous enzymes in the cell lysate. This route showed its laxative effect only at an elongated incubation of nine to ten hours, but did not influence initial rate yields. So its application is limited when incubation time cannot be extended.

The low yields of produced eGFP in this work might also be reasoned to the extract preparation. The protocol used for generating the TX-TL extract by Sun et al. (2014) claims a single centrifugation step after cell debris at 12000 g for 10 min to remove cell debris. Liu et al. (2005) claims that a second centrifugation step enhances protein yield in the end. Although the underlying mechanism is not known, they suggest that inhibiting factors might be depleted in the second centrifugation step.
Finally, a calibration curve with eGFP protein instead GFPuv would be more accurate. Since the brightness of fluorescent protein is strongly dependent of wavelength maximum and quantum yield, there can be large variations between different fluorescent proteins. Thus, the transferability of a standard curve set up with one protein is only partially applicable to the quantification of another, although similar protein.

4.2.4.2 LET generation by plasmid digestion

The sequence information of LETs is the same of that of plasmids, the only difference is the not circularized form, and thus LETs contain two ends. This property leads to lower protein yield of LETs compared to plasmid since this ends are vulnerable to exonuclease activity. Fast degradation rates of double stranded DNA to single stranded DNA of multiple nucleotides per second are reported at some conditions (Lee et al., 2011). According to Sun et al. (2013), non-coding sequence overhangs at this ends will sterically protect the coding sequence for the gene of interest.

In a first attempt to generate LETs, the positive control plasmid was digested with restriction enzymes to obtain a linear variant. Because of limited restriction sites available, a fragment with following properties was produced: ApaI was employed to cut around 900 bp before the gene, ScaI was employed to produce around 1500 bp overhang at the 3' end of the gene. The gene itself plus regulatory parts contain around 800 bp, thus altogether a final LET size of around 3200 bp resulted. This fragment was loaded onto an agarose gel and purified. Because of the large size of the fragment only low concentration of the LET was obtained after purification procedure.

A concentration of 16 nM LET was reported to be optimal (Sun et al., 2013), the concentrations obtained were around 4 to 7 nM. Nevertheless, this LETs were employed in CFPS. As visible in Figure 4.6, a low yield for the LETs (4 and 5) in comparison to plasmid (3) is visible after eight hours of reaction time. When using NC (1) as 0 % yield, for 5 nM LET an eGFP yield around 2 % and for 7 nM LET a yield of around 8.5,% was achieved. Moreover, the PC had a higher concentration of 10 nM, therefore the lower concentration LETs are underestimated. Although this yields are not significant, they indicate possible eGFP expression in a cell free manner from LETs. Furthermore, the concentrations of the LETs were too low and should be enhanced to obtain higher protein yield. Since the LET approach for CFPS offers a flexible platform to obtain proteins of interest and still exhibit potential to be optimized this route is pursued further. Moreover, in contrary to plasmids, LET production is easily automatable.

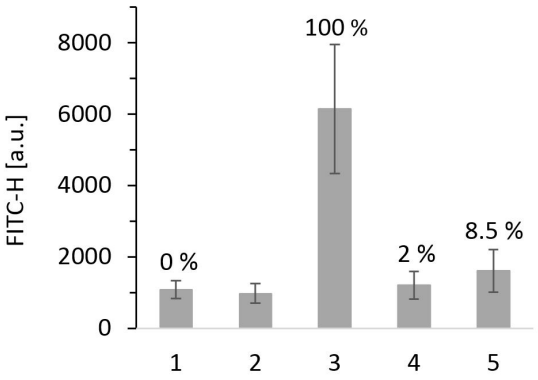


Figure 4.6: Fluorescent signal of eGFP produced by CFPS. As DNA template a digested and purified PC plasmid was used. 1, no plasmid; 2, empty plasmid; 3, PC 10 nM; 4, LET 5 nM; 5, LET 7 nM. Error bars indicate SD.

Table 4.2: LET yield in relation to plasmid, in percentage. Three independent LET batches with indicated concetrations were employed in CFPS at different temperatures.

	24 °C	29 °C	34 °C
1. batch, 4.5 nM	4.9	3.2	4.5
2. batch 4 nM	5.6	1.9	7.6
3. batch 4.5 nM	4.7	3.4	18.6
average	5.0	2.9	10.2
standard error	0.4	0.7	6.0

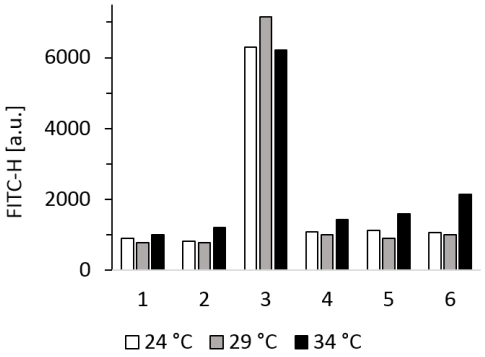


Figure 4.7: Fluorescent signal of eGFP produced by CFPS. As DNA template LET (digested and purified PC plasmid) was used. Reaction time is 8 h at three different temperatures. 1, no plasmid; 2, empty plasmid; 3, PC 4.5 nM; 4, LET 4.5 nM; 5, LET 4 nM; 6, LET 4.5 nM.

One parameter to enhance CFPS yields is the reaction temperature. Enzyme reaction rates are influenced by temperature, also those of Ribonucleases (RNAses). A temperature screening of CFPS based on LETs was performed to check for optimal condition for eGFP. As shown in 4.7, the usual temperature of 29 °C and temperature of ± 5 °C was tested. A lower temperature could lower RNAse and exonuclease activity, thus protein yield could increase. This holds true for 24 °C. For the low DNA template concentrations of 4 to 4.5 nM, a decrease in temperature led to significant increase of protein yield produced by LET. When respective positive control is set to 100 % yield, LETs run at 29 °C show 2.9 % ± 0.7 and LETs run at 24 °C show an average of 5.0 % ± 0.4 yield. Increasing the temperature to 34 °C led to the highest yields detected of averaged 10.2 % ± 6.0 whereas the standard error is high. This yields were achieved for the maximal running time of eight hours, for shorter reaction times lower temperature might decrease protein yield. Thus a temperature screening for shorter times with LETs will follow.

4.2.4.3 Enhancing LET based protein yields by gamS

A protein called gamS which is a RecBCD inhibitor of bacteriophage lambda was included to CFPS to protect linear DNA from degradation caused by exonucleases. The plasmid was purchased from Addgene and expressed as described in section 3.4.5. It was added to CFPS in a recommended concentration of 3.5 µM (Sun et al., 2014).

As shown in Figure 4.8, a test was set up without (left side, A) and with (right side, B) gamS addition. NCs (empty plasmid and no plasmid, 1 and 2) as well as two different concentrations of plasmid based PC and same two concentrations of LETs produced by digestion and purification from the PC plasmid were investigated. The visualization via blue light source showed less fluorescence intensity for the PC with gamS addition (B1, 3 and 4) in comparison with the PC without gamS (A1, 3 and 4). This result is surprising since no effect of gamS to plasmid based CFPS is described by Sun et al. (2014). However, the blue light source showed higher signal intensity for plasmid based samples without gamS, whereas the flow cytometry measurement showed higher fluorescence intensity results for samples with gamS addition. The interference of gamS with the flow cytometer measurement is going to be discussed in the following paragraphs. The flow cytometer method is presented in section 4.3.

When evaluating CFPS samples by flow cytometry, the NCs exhibit a low basal signal. Both plasmid based PCs exhibit a clearly distinguishable fluorescence signal compared to the NCs. It is remarkable that the DNA concentration has a

big impact on the FITC-H yield, both concentrations (4.5 and 8 nM) are lower than the usually used 10 nM plasmid DNA, so not at saturating levels yet.

For the LET based CFPS samples the addition of gamS showed an enhanced

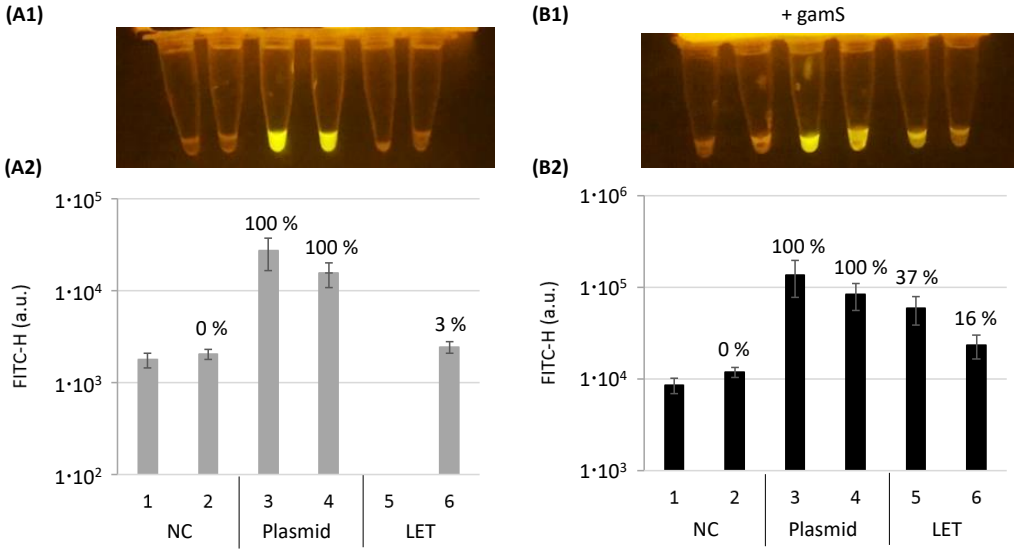


Figure 4.8: TX-TL reactions visualized by blue light source (A1 and B1) and measured by flow cytometry (A2 and B2). A1 and A2 (grey bars) correspond to conventional TX-TL measurements. B1 and B2 (black bars) represent the same set up, only gamS protein was added to a final concentration of 3.5 µM. The order of the tubes in the pictures correspond to the order of the bars in the graph. Legend: 1, NC no plasmid; 2, NC empty plasmid; 3, PC 8 nM; 4 PC 4.5 nM; 5, LET 8 nM digested and purified; 6, LET 4.5 nM digested and purified. Please note the difference of one magnitude in the FITC units, gamS addition to the samples interferes with the flow cytometer method. The sample no 5 in A2 was not recorded.

signal at the blue light source, even distinguishable from the NC sample by eye. The negative control with the higher basal fluorescence level (in this case empty plasmid) was set as 0 % eGFP yield and the respective two PC plasmids were set as 100 % yield. In respect to that, the yield of the LETs were calculated. The result from the previous section match very well with the yields here. LET based CFPS without gamS additions showed a very low yield of about 3 %. For the same template amount of 4.5 nM and with addition of gamS the LET to plasmid yield increased to 16 %. Also the concentration of the LET is crucial, for a higher concentration of 8 mM the yield of the LET to plasmid with gamS addition reached around 37 %. This is still low compared to plasmid (around one third) but equals published yields. Sun et al. (2014) reports less than 2 % LET yield to plasmid without gamS addition and 37.6 % LET yield to plasmid with gamS addition for DNA template provided with 16 nM.

This numbers are very similar for the yields achieved in this study although the DNA template concentration here is 8 nM. Because protein yields from plasmid run faster into saturation for quite low plasmid yields compared to LETs, the higher the LET concentration the closer it gets to the respective plasmid yield.

Although gamS addition might not be advantageous for plasmid, for LETs it displays an positive effect. Protein gamS, a truncated variant of lamda gam could successfully inhibit RecBCD (exonuclease) activity (Sun et al., 2014). Another effect observed is that gamS might exhibit surface activity. By flow cytometry measurement, events/second are displayed. For samples without gamS, and by using a flow rate of 30 µL/min and a particle dilution of 1:20000, a range of 4000 to 6000 events/second is reached. For samples with gamS and the same set up, this number rises from 1000 to 17000 events/second. According to literature research, no evidence of fluorescence activity of gamS could be found. This hints towards gamS influencing the bead based flow cytometer method. One possible starting point could be the interaction of gamS with the beads. A concentration dependent flow cytometry experiment of gamS was launched to have a closer look at this effect. Different concentrations of gamS were added to an empty plasmid control. The fluorescence signal increases with increasing gamS concentrations (see Table 4.3). 7 mM of gamS protein already enhanced the FITC signal by a factor two compared to no gamS addition. Since beads are employed to trigger measurement events, they are needed, but should not interfere with samples. gamS should not affect plasmid performance (Sun et al., 2014), and so the FITC yield should stay stable. However, when assuming that gamS influences the measurement of all samples in the same manner, these samples can still be judged relatively to each other.

Table 4.3: gamS influence on flow cytometer method.

gamS [mM]	FITC-H [a.u.]	rCV
0	3739	102 %
1.75	4122	82 %
3.5	6021	73 %
7	7538	65 %

4.2.4.4 Gene synthesis - LET generation by assembly PCR

So far LETs were always obtained by plasmid digestion, which has two drawbacks. Firstly, there is still a dependence of plasmid generation and secondly, the obtained LET concentrations are too low to make full use of the TX-TL

potential yields. Therefore, next approach for LET production was by *de novo* gene synthesis using a two-step PCR method. The DNA template free method described by Sequeira et al. (2016) offered a platform to realize a flexible DNA sequence creation and in contrary to the digestion of plasmid, high DNA concentration. Since the DNA template is generated by simple PCR it is also compatible with automation.

In short, the method is based on a fragmentation of the double stranded sequence into adjacent oligomers which overlap with each other. By an assembly PCR this oligonucleotides (oligomers) assemble back to a full sequence. The thermodynamics of the overlapping oligomers are crucial for a successful assembly reaction, they need to match to ensure a specific hybridization.

Firstly, the polymerase chain assembly method without using a template DNA, and additionally containing some nucleotide gaps between neighbouring oligomers was tested according to Sequeira et al. (2016). The gap represents a part of the sequence that is only covered by one strand instead of a double strand. It was intended to use these gaps as sites to introduce modifications later on. Furthermore, the longer the gaps, the less oligomers are needed to cover the sequence of interest. Oligomers were designed by cutting the sequence exactly at 40 nucleotides. After running polymerase assembly reaction, either no gene product or product with wrong size, thus unspecific gene product was obtained. Also after some optimization work this problems could not be fixed, therefore the strategy was changed.

Rouillard et al. (2004) describe a polymerase assembly method in which there are no gaps between oligomers, but instead more freedom of nucleotide length to keep thermodynamic properties of all oligomers in the set consistent. This is probably the reason the LET generation after Sequeira et al. (2016) did not work in the case of GFPuv. Because of the strict length of oligomers, but different GC content and thus melting temperature the assembly was disrupted at critical sequence parts. Taking this into account by setting constraints regarding maximum number of nucleotides allowed to deviate from the average nucleotide gene length and also keeping melting temperatures within an narrow range, the previous problems can be overcome. An oligomer designer tool "gene2oligomer" based on Rouillard et al. (2004) was used to design a new set of oligomers.

A fragment of 1152 bp was created which consists from 5' to 3' end of 5' overhang 159 bp, promoter region 88 bp, GFPuv 714 bp, terminator region 42 bp and 3' overhang 143 bp. This fragment was assembled out of a set of 56 primers which have an average hybridization unit of 20 bp ± and average T_m of 41 °C for the Taq polymerase (min = 39.3 °C, max = 47 °C). The oligonucleotide sequences for the GFPuv protein can be found in the Appendix A.1.1. GFPuv was used as template because it can be quantified with the calibration curve established in section 4.3.

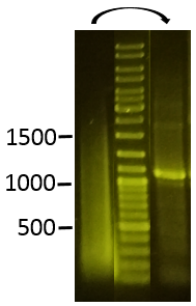


Figure 4.9: Agarose gel of two-step assembly PCR fragments. Left lane indicates first PCR product assembled out of 56 oligomers annealing to a GFP gene variant. The right lane shows the subsequent second PCR product in which the desired fragment of 1152 bp is amplified.

Two subsequent PCRs were performed. The first PCR reaction involves all 56 oligomers in low concentration and serves as assembly reaction to anneal all oligomers to the full sequence. As can be seen in Figure 4.9, this reaction results in a smear of DNA ranging from few bp to around 1.5 kbp. Afterwards, the product of the first PCR served as DNA template for the second PCR. Here, only the outermost forward and outermost reverse primer of the initial set of 56 primers were employed to amplify and obtain the desired fragment size of 1152 bp.

For these two subsequent PCR reactions different temperature profiles with the Phusion® High-Fidelity DNA Polymerase were tested to obtain high quality LETs important for the CFPS. Since no cellular repair or correction mechanisms come to game for the LETs produced by PCR, high fidelity of the assembled sequence must be ensured before CFPS. Since PCR amplifies exponentially, this would result in a high error rate and mixed sequence population. Random mutations caused by the polymerase can change enzyme properties and result in contradictory outcome. Also, high concentration of LET is desirable and also quality and purity of the sequence, this includes no unspecific bands. Potapov and Ong (2017) showed that the fidelity of Phusion polymerase is 39 times higher than the fidelity of the well-known Taq polymerase (the error rate of Phusion polymerase is given with 1.8×10^{-6} sub/base/cycle), so also having less cycles will influence the fidelity (and concentration) of the product.

A range of temperatures for the first as well as the second PCR reaction was tested, most lead to a gene band with varying intensity at the correct size. Reducing the temperature could lead to unspecific hybridization and thus a mixed population, a temperature too high would result in no product (Product information of Thermo Scientific Phusion flash High-Fidelity PCR Master Mix). Some products were send for Sanger sequencing which verified a flawless LET sequence. No errors were detected for annealing temperature of 58 °C for the first PCR and an annealing temperature of 68 °C for the second PCR reaction. The obtained concentration in a range from 100 - 300 ng/µL are more than sufficient and will contribute to enhance LET based CFPS yields. These LETs are to be characterized within the frame of fluorescent protein modification in the following chapter.

4.2.5 Conclusion and outlook

Two different but *E. coli* based cell lysates were produced to establish CFPS in this work. Both lysates showed activity with small fluorescent proteins. GFP variants (eGFP and GFPuv) were chosen as role models on which this work is based on. For both lysates, a T7 RNA polymerase based and an endogenous polymerase based, lower eGFP protein yields were achieved in comparison to yields reported in literature. Here investigation into protein folding and solubility could follow by supplementation of additives like chaperons. Because the TX-TL lysate showed activity immediately, it was set as the working lysate and used as a basis for this study. Nevertheless, in future depending on the protein, a T7 based strong promoter might be the better choice over the endogenous core TX-TL transcription and should be tested and compared. Furthermore, the majority of the plasmids utilized in this lab are T7 promotor based, therefore a direct testing of a protein of interest in CF manner would be accelerated, without the need of previous cloning steps.

A transfer of CFPS from a tube to a well of a 96 well plate was shown to be trouble-free for a reaction scale of 30 µL, but first indication of oxygen limitation was seen for the geometry of a tube. For volumes of 40 and 80 µL a decrease of almost 20 % was observed with a clear gradient in fluorescence intensity increasing towards the air surface area. For further applications, if a higher reaction volume is required, action should be taken according to scale up criteria. Agitation and reaction vessel geometry are two parameters to investigate.

The investigations in this section showed a large influence of each CFPS component on the CFPS performance, which became apparent in batch to batch variations. These components include the quality of DNA template, additives and

cell lysate batches. Therefore, care should be taken to ensure reproducible and accurate handling and thus allow for efficient protein production. Furthermore, the concentration of the DNA template displayed a crucial influence on the CFPS yields. While for plasmids used as DNA templates already low concentrations around 10 mM are sufficient to achieve a saturation effect and maximum protein yields, this is not sufficient for LETs (the optimum is around 16 nM). The performance of latter in comparison to plasmid based CFPS could be enhanced from no significant protein yield to around one third of plasmid based yield in this study. This improvement was enabled by addition of noncoding overhangs on both LET ends and employing the exonuclease inhibitor protein gamS in reactions. An influence of gamS to the flow cytometer method was seen. Here the underlying mechanism is not clear and could be investigated further.

The transfer from plasmid based to LET based CFPS was successful. Because the obtained protein yields are strongly dependent on the LET concentration, high quantity as well as quality was aimed at. Finally a fast and flexible way of LET generation was established by a two-step assembly PCR method. The gene of interest was produced by using a high fidelity Phusion polymerase and tested for optimal PCR conditions. In this way LETs with flawless gene sequences were obtained in a high concentration range.

4.3 Development of a highly sensitive and robust method for detection of fluorescent proteins in crude cell free protein background [1]

To detect and characterize cell free (CF) produced proteins sensitively, a robust detection method is indispensable. In this study, so far, fluorescent proteins were employed to provide a proof of concept for the new protein manipulation and detection process chain. Consequently, flow cytometry can be utilized for the screening of the fluorescent protein variants. A flow cytometer can quickly read out a huge population by delivering discrete intensities of fluorescence and statistical evaluation. Since the flow cytometer, dependent on laser and filter settings, is only measuring fluorescence in a defined range, background proteins without intrinsic fluorescence, as it is the case for CFPS, should not influence the measurement. However, fluorescent proteins produced in a CF manner lack particles/cells which are crucial to trigger measurement events at the flow cyctometer by scattering light. Therefore, a new strategy was developed in which the particles should be mimicked by unfunctionalized 0.909 µm polystyrene beads. These are large enough to scatter the light and enable fluorescence measurements, at the same time being small enough to not occupy the entire measuring chamber and so allow the fluorescent bulk solution surrounding the beads to be measured (Wurm et al., 2019). This method was adopted to fluorescent proteins expressed in CF manner and characterized within this work to enable the detection of fluorescent protein in crude suspension. Two different fluorophores, fluorescein and GFP, served for these purposes.

Firstly, pure fluorescein and GFPuv were spiked into matrix (sheath fluid and 10 mM citrate buffer, 10 mM EDTA, pH 7.0) to establish calibration curves. For detailed experiment set up see material and method section 3.8. Additionally, this experiment was duplicated at a conventional spectrometer to compare the sensitivity of these methods (see Figure 4.10A and B). The emission signal at 525 nm of purified GFPuv spiked into 0.001% PEG containing buffer and beads was recorded via flow cytometry at the CytoFLEX flow cytometer (Beckman Coulter, Germany) and via spectrometry at Tecan Safire 2 (Thermo Fisher Scientific, Germany). A broad range of GFPuv concentration was covered (0.0001 to 100 µg/mL, see Figure 4.10B). Thereby the flow cytometer measurement was approximately ten times more sensitive than spectroscopy (lower limit

[1] Part of this work is based on M. Wurm, S.Ilhan, U. Jandt and A.-P. Zeng, 2019, Analyt. Biochem.

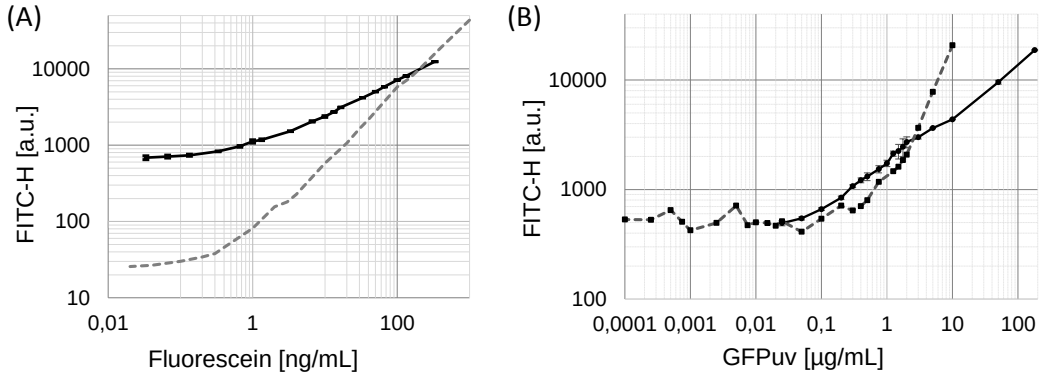


Figure 4.10: (A) Fluorescein calibration curve detected via different methods. Black line: measured at a flow cyctometer using 909 nm polystyrene beads. Grey dashed line: measured at a Tecan Safire 2 spectrometer. Both methods show comparable sensitivity, the limit of detection is below 1 ng/mL. (B) Calibration curves for purified GFPuv. Black line: GFPuv measured at a flow cytometer using 909 nm polystyrene beads and PEG-containing buffer. The sensitivity limit is around 0.05 µg/mL. Grey dashed line: GFPuv measured at a Tecan Safire 2 spectrometer. The sensitivity limit is in the range of 0.2-0.5 µg/mL. Figure adapted from Wurm et al. (2019).

around 0.5 µg/mL) and so the better choice to detect low fluorescent protein yields (Wurm et al., 2019). Fluorescein measured via flow cytometry (and addition of polystyrene beads) or spectroscopy showed similar sensitivities for both methods of less than 1 ng/mL. These curves show a linear relationship between concentration and FITC-signal strength over a broad range of fluorophor concentration, thus the new flow cytometer method is easily applicable for purified fluorophores in a defined bulk solution.

A further GFPuv calibration curve with the newly established flow cytometer method was set up in a crude CL background to judge about its applicability with samples produced by CFPS and without any purification steps in between (see Figure 4.11 C). To ensure comparability to purified GFPuv, this curve is shown in a black solid line (containing PEG). GFPuv spiked into a crude cell lysate background was measured with and without PEG, these measurement points overlay (grey dotted line). PEG is known to block unspecific protein adsorption as it might occur for GFPuv at the bead surface, this phenomenon will be investigated in more detail in the next section 4.3.1. The sensitivity of GFPuv in a defined background is higher (0.05 µg/mL to 0.2 µg/mL), but requires the addition of PEG to avoid concentration depletion effect on the bead surface. Although the lower detection limit of GFPuv in crude CF lysate background is less sensitive, no bead surface effects were recorded here. Furthermore, the

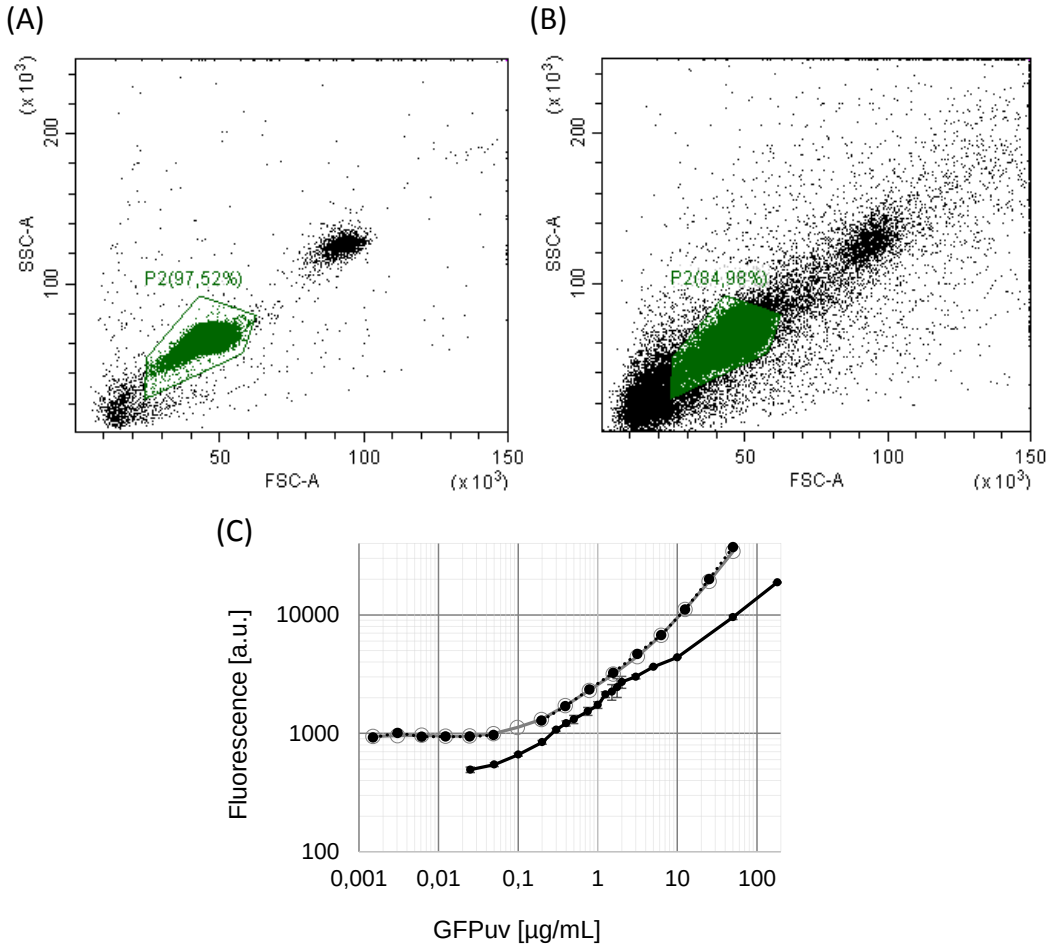


Figure 4.11: (A) and (B): dot plot of GFPuv samples in pure (A) or protein lysate background (B). Both plots represent 5.2 µg GFPuv spiked into matrix. The SSC is plotted over FSC to resolve different populations. Settings are: 50000 events, 30 µl/min, threshold trigger level: channel FSC 10000. The gated population (P2) represent single ploystyrene beads, the fluorescence in the bulk solution surrounding them are detected in the FITC channel and plotted in (C). Here, protein GFPuv was spiked into crude lysate (grey bars), measured with and without PEG addition which led to overlapping curves. For comparison purified GFPuv with PEG addition is demonstrated in black solid line. The sensitivity of this measurement for fluorescent proteins in a crude background is around 0.2 µg/mL. Non-functionalized 909 nm polystyrene beads were added to trigger measurement events.

addition of PEG to the crude background measurements had no influence on the results (see Wurm et al., 2019). Both effects can be explained by the crude background which is rich in unspecific proteins and indeed, already contain PEG (2 % in the final CFPS reaction, 0.5 % in the prepared sample for the flow cytometer measurement, see protocol Sun et al., 2013). Therefore, the crude background blocks unspecific bead surface interaction and makes this method highly robust and well suited for the detection of fluorescent protein in crude CL background.
However, compared to measurements in a defined matrix, the crude background containing different proteins, salts and additives lead to an enhanced portion of background particles triggering the measurement events. Figure 4.11A shows a dot plot of pure GFPuv measured in buffer. The green gated area marks the single bead population which makes up around 97.5 % from which the fluorescence signal is detected. In comparison, Figure 4.11B shows the same GFPuv concentration spiked into a crude CL background. The dot plot is scattered over a broad range because large particles in the CL which are close to the polystyrene bead size also trigger measurement events. When applying the same gate as for (A) to detect the single bead population, its portion on the overall events drops to 85 %. To ensure defined and comparable measurement of each sample, the threshold settings were adjusted (ass denoted in the Figure caption) to filter out smaller particles. Additionally, a blank measurement with only beads in sheath fluid was performed for each experiment to detect and gate the single bead population.

4.3.1 Protein to bead surface effects

Proteins tend to adsorb at different solid to aqueous interfaces and form agglomerates at surfaces, which might lead to their inactivation (Pai et al., 2010). To study this behavior of possible agglomeration of GFPuv at the bead surface, experiments with the addition of PEG, which is known to block proteins, were conducted. In case of fluorescein no effect of PEG is expected since it is chemically stable. In case of GFPuv, which is known to form dimers (Topell et al., 1999) the behavior in respect to the bead surface will be tested.

As can be seen in Figure 4.12, three different GFPuv concentrations in a time series from 0 min to 21 h were tested. Here, the protein was first incubated for the time denoted in presence of beads and with (black bars) or without PEG (grey bars) addition (of 0.001 % w/v). Afterwards, the beads were transferred to

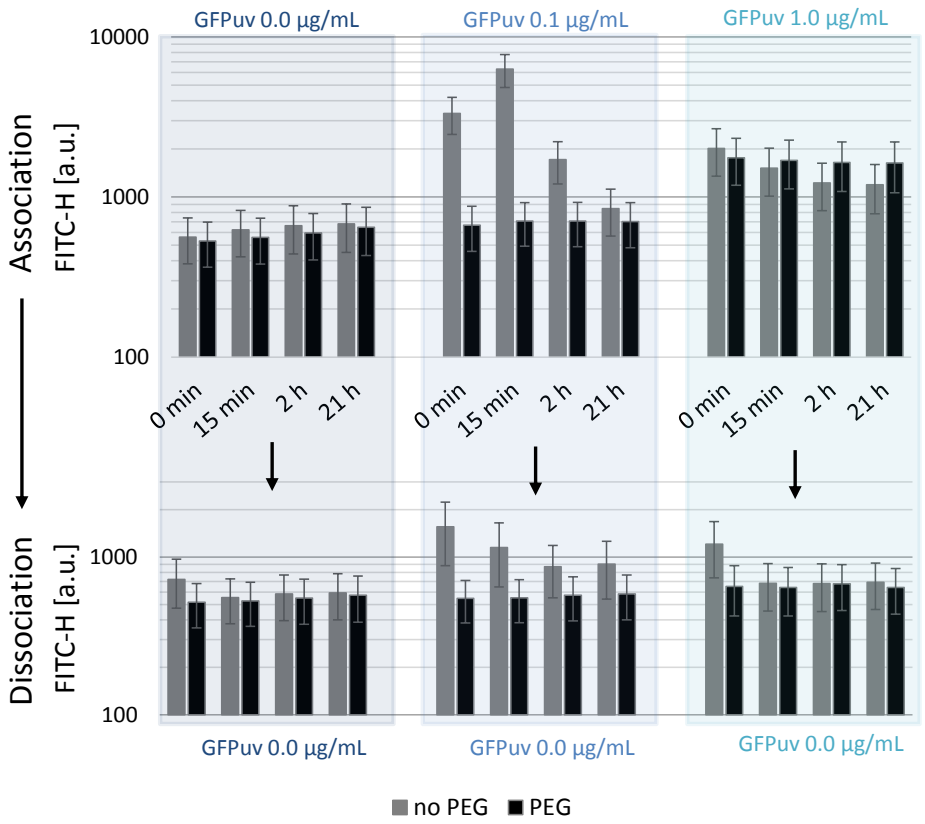


Figure 4.12: Measurement of time dependent fluorescence signals for GFPuv with (black) and without (grey) PEG in the buffer. Error bars denote the CV. The beads were incubated for the denoted times (association) with the given GFPuv concentration (indicated in blue colour above the graphs) prior to measurement. Afterwards, these samples were centrifuged to replace the fluorophore containing buffer by GFPuv-free buffer (dissociation). Figure adapted from Wurm et al. (2019).

protein free solution by buffer exchange via centrifugation steps. This method will clarify whether the proteins stick to the bead surface and if this is time dependent.

Because of the difference of the results for the reaction series without PEG addition (grey bars) to with PEG addition (black bars), it can be concluded that PEG is indeed suppressing the GFPuv bead adsorption. During the incubation phase with PEG the fluorescence signal strength for each GFPuv concentration remains constant over time. After GFPuv removal, the fluorescence signal of the beads adjusts to the negative control of 0 µg/ml GFPuv. This leads to the conclusion that the fluorescent signal in case of PEG addition must be attributed to fluorescence from the bulk solution. In case of no PEG addition

and 0.1 µg/mL GFPuv, an increase of the fluorescence signal is seen until 15 min, and thereafter a sharp decline until 21 h. To a lower extend this can be also seen for 1 µg/mL GFPuv, whereas due to the 10 times higher concentration the maximum fluorescence signal might be reached in a time range of seconds and therefore difficult to capture. This behavior can be explained by a concentration depletion effect, which takes place at the bead surface and is blocked by PEG. So in case of protein GFPuv a clear concentration dependent effect of surface agglomeration is shown.

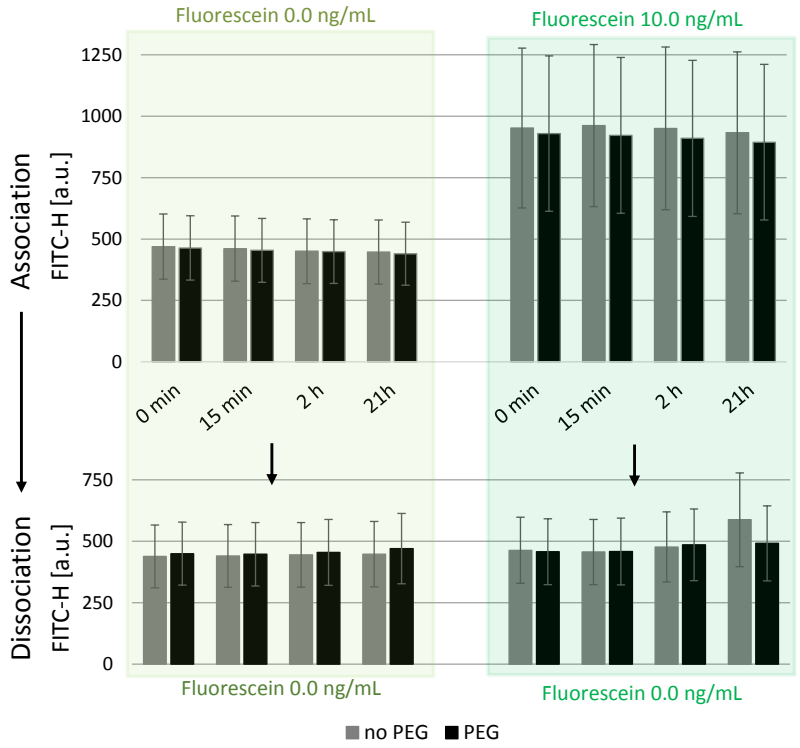


Figure 4.13: Measurement of time dependent fluorescence signals for fluorescein with (black) and without (grey) PEG in the buffer. Error bars denote the CV. The beads were incubated for the denoted times (association) with the given GFPuv concentration (indicated in blue color above the graphs) prior to measurement. Afterwards, these samples were centrifuged to replace the fluorophore containing buffer by GFPuv-free buffer (dissociation). Figure adapted from Wurm et al. (2019).

The analogue measurement was performed with fluorescein as fluorophore (see Figure 4.13). The results for fluorescene signals in presence and absence of PEG are highly similar. This concludes to no concentration depletion effect for fluorescein in a bead solution. After removal of the fluorophore from the bulk solution, the remaining fluorescence in the bead solution equals the negative control signal strength. Therefore the measured fluorescence signal for 10 ng/mL fluorescein must be assigned to the bulk solution, demonstrating that fluorescein is not binding to the polystyrene beads.

4.3.2 Conclusion

A modified flow cytometer method was developed for the purpose of fluorescent protein detection in a crude CF protein rich background. Therefore, two different fluorophores (fluorescein and GFPuv) were utilized to characterize this new method. Unfunctionalized 909 nm polystyrene beads were added to the samples of interest to trigger measurement events and allow the fluorescence in the bulk solution surrounding the beads to be analyzed. This new method is highly robust due to the high number of measured events (50000-100000), while being 10 times more sensitive compared to conventional spectrometer analysis. The lower limit of quantification of GFPuv in a crude background is around 0.2 µg/mL, enabling a certain detection of lowest expression levels. A calibration curve of GFPuv was established, measured with crude CF background. This was utilized to evaluate the eGFP produced by the TX-TL lysate on base of the PC plasmid. The yield of CF produced fluorescent protein is roughly estimated to be around 40 to 90 µg/mL.
A concentration dependent influence of GFPuv interacting with the bead surface was found. This interaction could be successfully suppressed by PEG addition to the samples, which allowed for a clear measurement of the fluorescence only in the bulk solution. Furthermore threshold settings and gating strategy was adjusted to a crude protein rich background. Since protein samples produced in a CF manner already contain PEG, no further addition for these samples are required, making this new flow cytometer method suited ideally for fluorescent proteins in a crude bulk solution.

4.4 Proof of concept - modification of fluorescent proteins

4.4.1 Creation of GFP variants

As already mentioned, GFP is a well-studied protein and different variants of it are known. These variants do have a shift in wavelength spectrum and thus do not emit at green light but at longer or shorter wavelength (yellow/red or blue/violet). Because green and blue, but no violet laser is available at the flow cytometer in this laboratory, mutants that can probably be detected with these available laser and filter settings were selected. A shift of GFP to yellow or to blue wavelength is chosen. A literature survey with the aim to select sequences of eGFP, eYFP and eCFP was performed. Using Bioedit 7.2.5, a sequence alignment of different published florescence proteins was performed, and the most common mutation points which lead to eCFP and eYFP were identified. On the basis of GFPuv, three mutations points were selected to obtain an eCFP variant (eCFPv): N147I, S176G, H232L. To obtain a eYFP variant (eYFPv) two points were identified: S66T and T204Y. In addition to these two points, a variant containing one more mutation site (V68L) was included (see Table 4.6). The oligomers containing these mutation points were exchanged by a set of new modified forward and new modified reverse oligomers. They were designed by exchanging usually one or two aa of the codon considering the codon usage of *E. coli*. All Oligomers are listed in Appendix Table A.1.

LETs for GFPuv, eCFPv and eYFPv were created by the two-step assembly PCR as described in section 3.6.2. The set for each assembly consisted of 56 primers. To obtain eCFPv, six oligomers of the initial set for GFPuv construction were altered. To create eYFPv containing a 2-point mutation, four oligomers from the initial set were exchanged by modified oligomers (see Table A.1). All variants were obtained by the two-step assembly PCR method and the sequence was verified by Sanger sequencing. The mutation points inserted were all present. Thus, these new LETs were employed in TX-TL reactions and analyzed in the following.

4.4.2 GFP variants produced by PCR based LETs and expressed via CFPS

4.4.2.1 LET purity and protease inhibitor

The quality of the DNA template influences the subsequent protein yield. LETs serve as DNA template, which are produced in two subsequent PCR reactions. So far these gene products were purified afterwards and employed for the protein expression. In regard of automation of the LET generation and subsequent CFPS an experiment was launched to test LETs in PCR buffer and polymerase background in comparison to purified LETs. This question is crucial to answer because the liquid handling system available is not equipped to perform gene purification. In addition, protease inhibitor was added to this reaction series to see whether it helps to enhance protein yields (see Figure 4.14).

GFPuv as well as eYFPv and eCFPv LETs were produced by a two-step PCR. A part was not purified and used in crude form (no. 3, 4 and 5), the second part was purified (no. 6, 7 and 8). The concentration of the LETs were adjusted to 16 nM, in the case of unpurified LETs the target gene concentration was approximated based on the agarose gel and might be a source for signal variation. There was a clear signal visible for all GFPuv and eYFPv based CFPS. The eCFPv based protein expression exhibited a fluorescent signal similar to NCs (no. 1 and 2). Probably no signal was seen here because the FITC channel doesn't have the optimal excitation and emission wavelength to detect the cyan wavelength (see Figure 4.17). In general, the purified LETs show better performance than unpurified LETs. In case of GFPuv the signal intensity of purified LET is 2.5 times higher than for crude LET, for eYFPv this is 2.2 times. Different reasons may account for this. Firstly, the concentration of the unpurified LETs might be judged incorrectly, thus the LET concentrations for the unpurified LETs might be not accurate. More important, the crude background might interfere with the CFPS to some extent. The crude background contains the PCR reaction mixture, which includes magnesium for optimal polymerase activity. The TX-TL lysate is calibrated for a specific magnesium concentration at which it exhibits its best performance, usually between 1 to 6 mM (Sun et al., 2013). The PCR buffer contains a final 10 mM of magnesium in the PCR reaction. The DNA amount in the final CFPS can make up to one third of the reaction volume, thus up to 3.3 mM additional magnesium. This magnesium concentration in the final CFPS doesn't match the calibrated concentration anymore and so lowers the protein yield. Another effect was expected for the LETs in crude background which might contain some incomplete PCR product. It is reported that "sacrificial DNA" enhances CFPS yield. The idea behind is

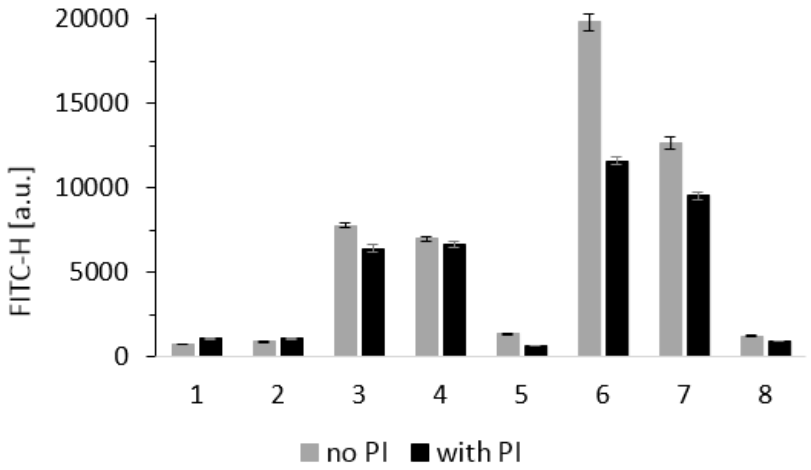


Figure 4.14: CFPS setup to test pure vs. crude LETs and in addition influence of protease inhibitor (PI). Reaction conditions: 10 µL reaction volume, 8 h run time at 29 °C. Grey bars represent the usual set up, black bars contain PI as additive. Legend: 1, NC no plasmid; 2, NC empty plasmid; 3, LET GFPuv crude; 4 LET eYFPv crude; 5, LET eCFPv crude; 6, LET GFPuv purified; 7, LET eYFPv purified, LET eCFPv purified. Error bars represent a 99 % confidence level.

to add non coding dsDNA to the CFPS, so this DNA gets degraded as well and increases the half time life of the desired LET (Broadbent, 2016). Probably the extent of non-coding dsDNA in the second PCR product is too low exhibit an effect. Although not purified, the lane for the second PCR product contains only one prominent band at the size of the desired LET on the agarose gel, thus is probably already highly pure.

Furthermore, protease inhibitor (PI) was added to the reaction solution. At a run time of eight hours already produced proteins could be degraded by endogenous proteases present in the *E.coli* lysate. Therefore, EDTA free protease inhibitor cocktail from Roche was added in recommended amount to the CFPS. As shown in Figure 4.14 (black bars), no enhance in fluorescence signal was visible for the reactions with PI. Usually, degradation of proteins in cells take place during cell regulation (Cooper, 2000). Because this cellular regulation mechanisms doesn't account for a cell lysate, there might be no evident effect of PI for CFPS. Stech et al. (2014) reported that a broad-spectrum protease inhibitor cocktail, such as the one used in this study, is not having any effect and protein degradation is already visible after five hours of incubation. Instead, caspases (cysteine proteases with aspartate-specificity) prevent proteins produced in CL to be cleaved, but also endogenous proteins like translation

factors which are essential for protein expression. However, for very short incubation times (less than 4 h) protein degradation might be only a minor problem.

For further automation work LETs will be generated at a liquid handling system and thus will not be purified. Although this leads to a decrease of protein yield by more than a factor two, the obtained signal intensity is still sufficient for the purpose of mutant analysis.

4.4.2.2 Temperature testing

As can be seen in Figure 4.15, the crude LETs were tested for a short incubation time of only two hours at different temperature. A lower temperature of 24 °C showed in general the lowest fluorescence intensity, 29 °C mostly an intermediate intensity and 34 °C mostly the highest fluorescence intensity. This is likely due to time limitation. Lower temperature lowers reaction rates which cannot be compensated for short running times as in Figure 4.7. Therefore, the temperature for short incubation times should be high enough to allow for a complete turnover of resources, but low enough to allow the proteins to be soluble and functional.

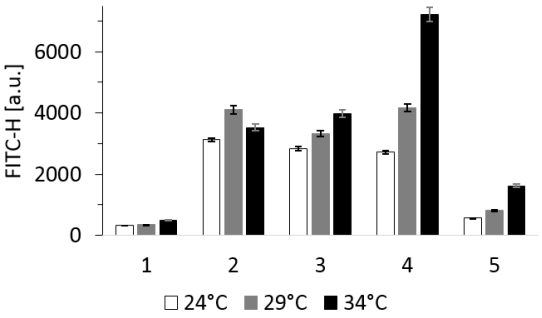


Figure 4.15: CFPS set up to test crude LETs incubated for 2h at different temperature. Reaction conditions: 10 µL reaction volume, 2 h run time at 24 °C, 29 °C and 34 °C. Legend: 1, NC no plasmid; 2, LET GFPuv; 3, LET eYFPv-2 point; 4, LET eYFPv-3 point; 5, LET eCFPv. Error bars represent a 99 % confidence level.

In case of GFPuv with no additional mutation, 29 °C showed higher intensity over 34 °C. This might be due to low stability of the GFPuv for higher temperatures. The mutation point of S65T is reported to be important for better folding (the chromophore maturation is faster) and show better solubility at 37 °C and thus also fluorescent brightness (Arpino et al., 2012). This mutation point is governed by the eYFPv mutants. Furthermore, the eYFPv owns T203Y mutation, which is responsible for the shift to longer wavelengths of GFP variants due to the additional polarizability of the π-stacked Tyr203 near the chromophore (Watcher et al., 1998). For the eCFPv, the replacement by the hydrophilic Ile (N146I) allows for flexibility close to the chromophore. This should stabilize CFP and enhances its quantum yield as well as more efficient fold at 37 °C (Lelimousin et al., 2009). Furthermore, the mutation point of S175G leads to cerulean (Park et al., 2016) and improved maturation due to more flexibility at that region (Rekas et al., 2002). These

might be reasons why eCFPv exhibits highest fluorescence intensity also at the highest tested temperature of 34 °C.

The detection of CFP variant with given laser and filter settings is not optimal (see Figure 4.17), and even aggravated by the low quantum yield and thus brightness of this variant. To overcome these drawbacks a CFP variant could be tested with enhanced properties, namely SCFP3A (F64L S65T Y66W S72A N146I H148D M153T V163A S175G A206K) and thus might ameliorate the CFP variant detection (Kremers et al., 2006). Here investigations were made next to chromophore formation improvement (F64L, V68L), the protein folding (S72A, V163A, S175G), and solubility (M153T, V163A). A206K hinders the fluorescent proteins to dimerize, this mutation was not tested within the scope of this work.

4.4.3 Automation

The so far obtained work flow of DNA generation, CFPS reaction and analysis of enzymes is just one part embedded in the frame of a feedback guided automated protein optimization loop system presented at the beginning of this chapter. Joining all these parts together in automated way will create a powerful platform for protein engineering purposes. Therefore, the beforehand described individual processes will be jointed in an automated work flow at a liquid handling system. To cover all tasks of the feedback guided protein optimization loop the liquid handling station will be connected to an analysis instrument and a computer which is the heart piece of this ensemble. For the case of fluorescent proteins a flow cytometer will be connected to the liquid handling station.
The Biomek 4000 Automated Workstation from Beckman Coulter serves as the liquid handling station (see Figure 4.16). It is equipped on the left side with holders for different single tip or eight tip pipette tools as well as a gripper for the transfer of plates. In the middle part 10 positions follow, whereas the lower 5 positions are variable and the upper five positions are fixed. The two rightmost positions are connected to thermostats with a temperature range from -20 °C to 80 °C. On the upper right corner a Thermocycler TRobot from Biometa is installed to perform PCR or incubation reactions in PCR plates. In the lower right corner a Thermoshake from Inheco allows to cool, heat and shake plates.

Firstly, the LETs were generated. Therefore all oligomers to be used were filled into 96 well plates. All necessary components like DNA polymerase and water will be provided on the deck. First, a master mix of all jointly used oligomers was produced. This was then aliquoted for each individual reaction. Next, the variant specific oligomers were pipetted directly into the designated reaction to create the different LET variants. Finally, all remaining ingredients for the

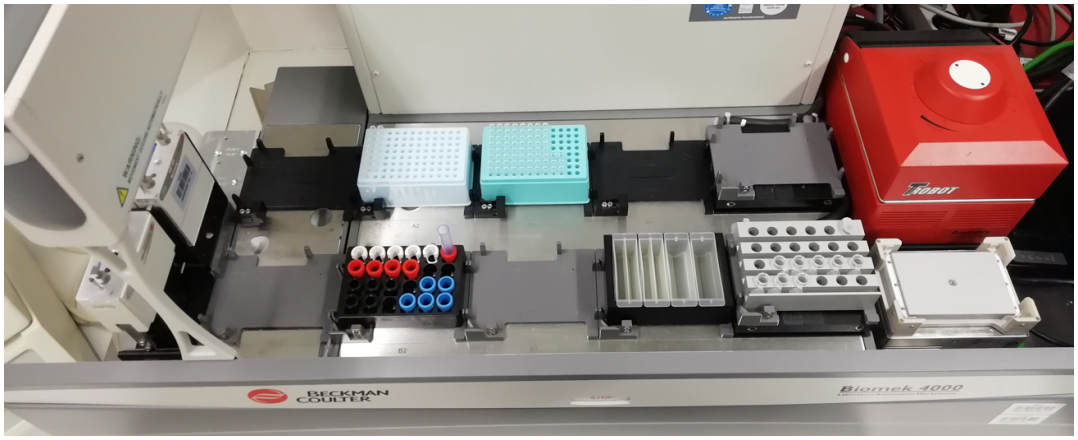


Figure 4.16: Liquid handling station "Biomek 4000" from Beckman Coulter used for automation work. In addition to the deck positions, this liquid handling system is equipped with a Thermocycler TRobot from Biometa (red instrument) to perform PCRs.

PCR were added on top, and the plate transferred into the TRobot for the temperature cycling. After its completion, the plate is transferred to a free position, and a second PCR is launched. Therefore the first PCR product is used as DNA template. As soon as the DNA template preparation is accomplished, the assembly for a CFPS starts. A master mix was prepared and aliquoted. By addition of DNA template, the reaction started and was incubated for few hours at 29 °C. Subsequently, each protein lysate was transferred into an Eppendorf tube and mixed with polystyrene beads for the flow cytometer readout. This automated process cascade starting from LET production until the flow cytometer measurement takes less than six hours for six mutants to complete, including a three hour incubation time for protein expression.

Difficulties arose during the transfer of PCR from stand-alone Thermocycler to the TRobot built in at the liquid handling station. One problem was noticeable evaporation of the PCR mixture during the cycling reaction due to not sealed PCR-plates. Another problem was the inefficient temperature ramps and heat transfer of the TRobot. Combining these effects, the yields and purity of the LETs produced via PCR in the TRobot were reduced due to changing concentrations in the reaction (because liquid evaporated) and not reached temperature settings. Furthermore, each action of the liquid handling system was checked and problems because of small sizes and volumes were revealed. Therefore, each protocol was adjusted in meticulous work to avoid bubbles or enable correct viscous liquid transfer from tip to tube.
The evaporation problem could not be solved by sealing of the PCR-plate because

this action cannot be performed by the liquid handling station, thus interferes with automation. Finally, sealing pads were purchased which were attached to the inner lid of the TRobot. This action helped to ameliorate the evaporation problem, but did not solve it completely. Another measure to tone down the evaporation problem was by scaling up the very little PCR mixture volume around three times. In this way the evaporation volume relatively to reaction volume could be lowered to around 10 %. The problem of slow temperature ramps could be solved by manual calculation of times for the cycling procedure by considering maximal temperature ramps of the TRobot. Thus, a cycling protocol with extended time specifications resulted. The cycling protocol for the TRobot is given in Table 4.4 and differs slightly from the cycling protocol of the stand-alone Thermocycler (see Table 3.15). This measure helped to produce the correct PCR fragments, which could be verified by applying the fragments onto an agarose gel. Although the problems were not solved completely, they were improved to a level which allows to operate the intended reactions without shortcomings for the results.

Table 4.4: PCR cycling instructions optimized for a two-step PCR of fluorescent variants using TRobot.

	1st PCR			2nd PCR		
	Temp.	Time	Cycles	Temp.	Time	Cycles
Initial denat.	98 °C	38 sec		98 °C	38 sec	
Denaturation	98 °C	8 sec		98 °C	8 sec	
Annealing	58 °C	25 sec	27 x	68 °C	20 sec	34 x
Extension	72 °C	23 sec		72 °C	22 sec	
Final extension	72 °C	240 sec		72 °C	240 sec	
Storage	10 °C	hold		10 °C	hold	

4.4.4 Setting up a cost function

To evaluate and judge about the experimental performance of different GFP variants, an algorithm was implemented to the protein optimization loop. This algorithm is dependent on a cost function which determines the error of the predicted to the actually measured values (Kamil K., 2018). In this way the algorithm can improve its prediction for the future, also referred to as machine learning. Within this project, a shift in the wavelength spectrum of fluorescence proteins is aimed to, therefore a ratio of the fluorescence intensities of each protein variant in different channels will serve to establish the cost function. The flow cyctometer employed is not able to measure the whole fluorescent

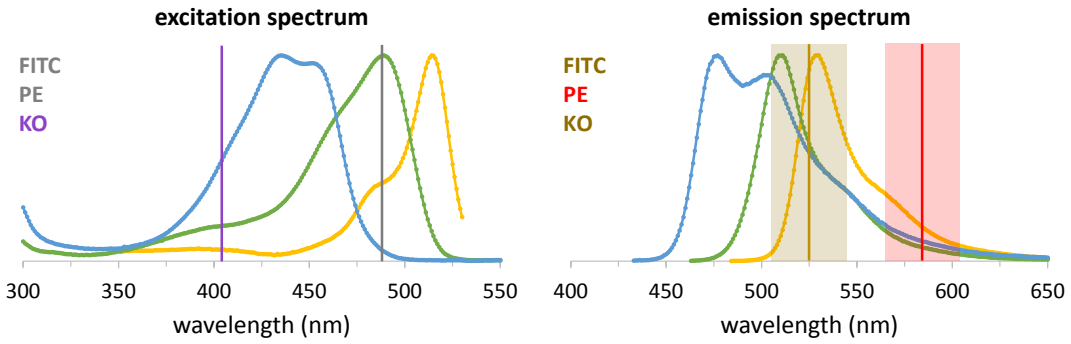


Figure 4.17: Spectral excitation and emission data for eGFP (green), eCFP (blue) and eYFP (yellow) obtained from fpbase.org. Indicated are laser (excitation) and filter (emission) settings of the flow cytometer available. The bandwidth of the filters are indicated in the emission spectrum. In general CFP exhibits a shorter wavelength than GFP, whereas YFP exhibits a longer wavelength than GFP.

spectrum, but is equipped with distinct lasers and filters. A laser focuses sharply, so the excitation wavelength is precise at the designated wavelength. The flow cytometer in this laboratory is equipped with Fluorescein isothiocyanate (FITC) and R-phycoerythrin (PE) channels, both excite at 488 nm. Additionally the Krome Orange (KO) channel can be used to excite the fluorophore at 405 nm (see Figure 4.17). The filter to detect the emission of fluorophores is at 525/40 bandpass (BP) for FITC and KO channel, for PE at 585/42 BP (an overview is presented in Table **??**). Since these channels doesn't exactly match the spectral maxima for excitation and emission of the produced GFP variants, a theoretical outcome by considering non optimal wavelength detection as well as quantum yield of different GFP variants was calculated as presented in Table 4.5.

The calculated theoretical values can only provide a trend for the cost function. This is because the fluorophores used for theoretical calculations are optimal eYFP and optimal eCFP, the fluorophores worked with in this study only contain some mutation points to shift the wavelength maximum to a desired direction. Moreover, experimentally other factors like fluorophore maturation time and pH value influence the signal intensity of the GFP variants. Nevertheless, from a theoretical point of view, by using the laser and filter settings available the desired target mutations are detectable. Furthermore, by using ratios of one sample measured in different channels, problems like batch to batch variations are omitted and thus a robust cost function is established.

Table 4.5: Spectral data for eGFP, eCFP and eYCFP were obtained from fbbase.org. The data represent relative fluorescent intensity at different wavelengths. For the channel excitation and emission the relative maximal intensity of the indicated fluorophor (also by considering the range of the bandpass filter for emission) was extracted. Multiplying these two values led to total channel values in which the theoretical maximum intensity of each fluorophore in the given channel is indicated. Because each fluorophore exhibits a different quantum yield, the molecular brightness was obtained from fpbase.org and is calculated as the product of extinction coefficient and quantum yield. By considering the brightness, the channel ratio (cost function) for different fluorophores was calculated.

channel	excitation	emission	
FITC	488 nm	525/40 BP	
PE	488 nm	585/42 BP	
KO	405 nm	525/40 BP	
excitation maximum	eGFP	eYFP	eCFP
FITC	0.998	0.380	0.054
PE	0.998	0.380	0.040
KO	0.173	0.057	0.509
emission maximum	eGFP	eYFP	eCFP
FITC	1	1	0.891
PE	0.148	0.327	0.176
KO	1	1	0.891
total channel	eGFP	eYFP	eCFP
FITC	0.998	0.381	0.048
PE	0.148	0.125	0.009
KO	0.173	0.057	0.453
brightness	eGFP	eYFP	eCFP
	34	45	13
total * brightness	eGFP	eYFP	eCFP
FITC	33.939	17.140	0.625
PE	5.0399	5.607	0.124
KO	5.889	2.574	5.896
cost function	eGFP	eYFP	eCFP
FITC/PE	6.734	3.057	5.063
FITC/KO	5.763	6.659	0.106
KO/PE	1.168	0.459	47.719

4.4.5 Results on wavelength shift of fluorescent proteins

After transferring the protocols of LET generation without purification of genes, CFPS and sample preparation for flow cytometer read out, different fluorescent variants are produced and evaluated in different channels to detect their shift in wavelength. Ratios of different channels of a sample are set as cost function to evaluate each mutant. Firstly, a time series was set up to see how much the incubation time for CFPS, which is currently the most time intense factor for the automated wet lab part, can be shortened to still allow good resolution and distinctness of the variants via ratios. As can be seen in Figure 4.18 A, the FITC/KO ratio of the three variants tested remain quite stable over a time course of two to eight hours. The time point of four hours even show a better resolution than eight hours, so it is even advantageous to shorten the incubation time to avoid interference or decay effects over time.

A 2-point mutation to obtain eYFPv and a 3-point mutation for eCFPv led to successful production of GFP variants, by using a ratio of two channels (FITC ex: 488 nm, em: 525 nm to KO ex: 405 nm, em: 525 nm). The relative signal ratio averaged over time is calculated (see Figure 4.18 B). A yellow variant is visible in the FITC channel, but not in the KO channel, thus its ratio should be the largest. The green variant is visible on both channels, whereas the cyan variant is only detectable in KO and thus a low ratio should result. This trend was stable and well pronounced and the different mutations were significantly distinguishable. This ratio corresponds well to the theoretically calculated cost function.

As can be seen in Figure 4.18 the theoretical cost function is not reached for the eCFPv mutant for the FITC/PE ratio as well as for KO/PE ratio. The laser and filter settings of the flow cytometer measurement for a variant shifting to blue wavelength direction is not optimal. Furthermore, the blue variant has a low brightness in comparison to the green or yellow fluorescence variant. Probably this 3-point eCFP variant is not ideal yet, therefore more optimization work is required.

4.4.5.1 Additional fluorescence protein variants

Since this project is going to be embedded in a protein optimization loop with a smart algorithm learning and giving feedback about mutation points, more information and variety is desirable. Furthermore, a fluorescent protein shifting to blue wavelength and nicely detectable is missing so far. Therefore, additional variants for each variant, green, blue and yellow (see Table 4.6 5, 6 and 7)

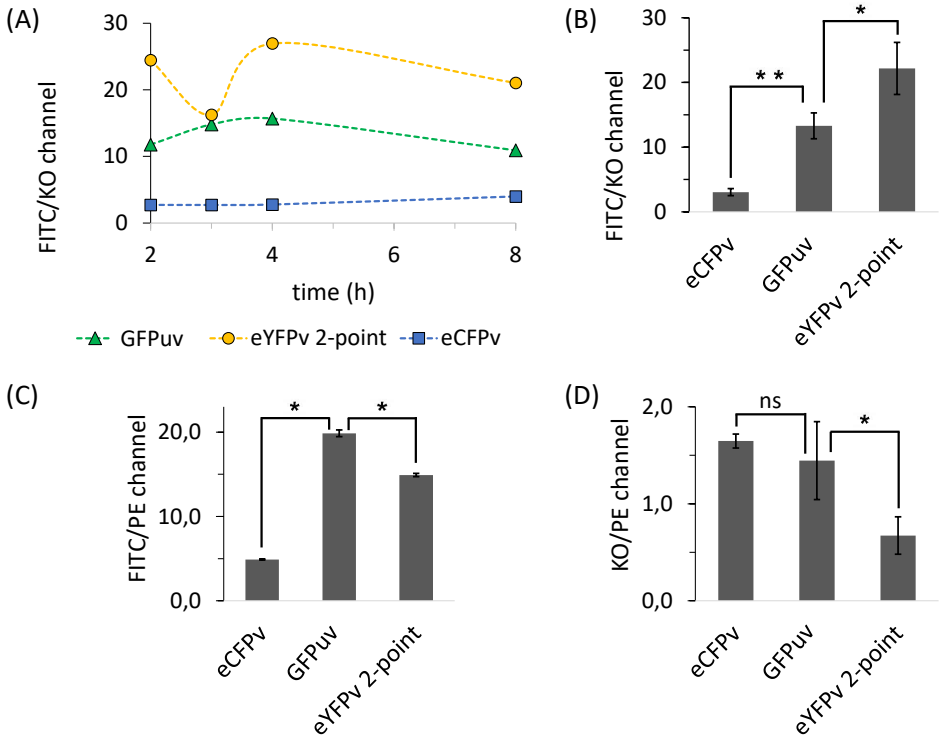


Figure 4.18: (A) Relative ratio of fluorescence signals in FITC to KO channel over CF protein expression time series. Shown are three different fluorophores: the starting variant (GFPuv), an eCFPv and an eYFPv 2-point mutant. (B) Evaluated cost function of channel ratios. Average of time series is plotted for each variant, error bars indicate SD of time series measurement. The variants are significantly distinguishable (eYFPv 2-point variant p-value = 0.005, eCFPv variant p-value = 0.039). (C) and (D) same data like B but evaluated in FITC/PE and KO/PE channel ratios. No significant difference is detected for eCFPv to the starting point of GFPuv in the KO/PE channel. Error bars indicate SD of time series measurement.

Table 4.6: Legend for the fluorescent protein variants used in this experiment.

mutant no	trivial name	mutations
1	GFPuv	none
2	eYFPv 2-point	S65T, T203Y
3	eYFPv 3-point	S65T, V68L, T203Y
4	eCFPv	N146I, S175G, H231L
5	-	F64L, S65T, R80Q, T153M, A163V
6	-	S65G, V68L, S72A, T153M, A163V, T203Y
7	-	F64L, S65T, Y66W, N146I

were set up, according to promising mutation points based on literature research (UniProtKB - P42212 (GFP-AEQVI)). Each mutant contained up to six mutation points, nevertheless, all were produced in a good concentration via two-step PCR. All mutants were purified and sequenced once to verify correct change of aa, which was successful. All crude LETs with variations were used in a TX-TL reaction with maximum possible volume allowed for DNA template. The reaction was run for three hours at 29 °C.
To compare the performance of the old and the new mutation points, all LETs were purified and employed in a TX-TL reaction with approximately 65 nM and incubated for 3 hours at 29 °C, the results can be seen in Figure 4.19 A.

The new mutants (5,6 and 7, see Table 4.6) match the theoretical cost functions with their order in each channel ratio. This is because the intended new green variant (mutant 5) shows enhanced performance over GFPuv (mutant 1). Because of this, also the FITC/PE ratio matches the cost function, which is not the case for the previous mutants (2, 3 and 4). GFPuv already contains mutations which lead to the assumption that these mutations may influence brightness and wavelength maximum, maturation time and stability at 37 °C. In comparison to eGFP, the GFPuv contains seven point mutations (V1A, L64F, T65S, Q80R, F99S, M153T and L231H). The mutation point of Q80R is reported in an azurite variant of BFP (Park and Rhee, 2012). In comparison to GFP this mutation shifts the spectrum to shorter wavelength and also exhibits lower brightness (14.41, from fbbaseorg). For GFPuv a brightness of 23.7 is given which is still below eGFP brightness of 33.54 (data from fpbaseorg). This might be the reason why mutant 5 shows higher intensities in FITC channel than GFPuv.

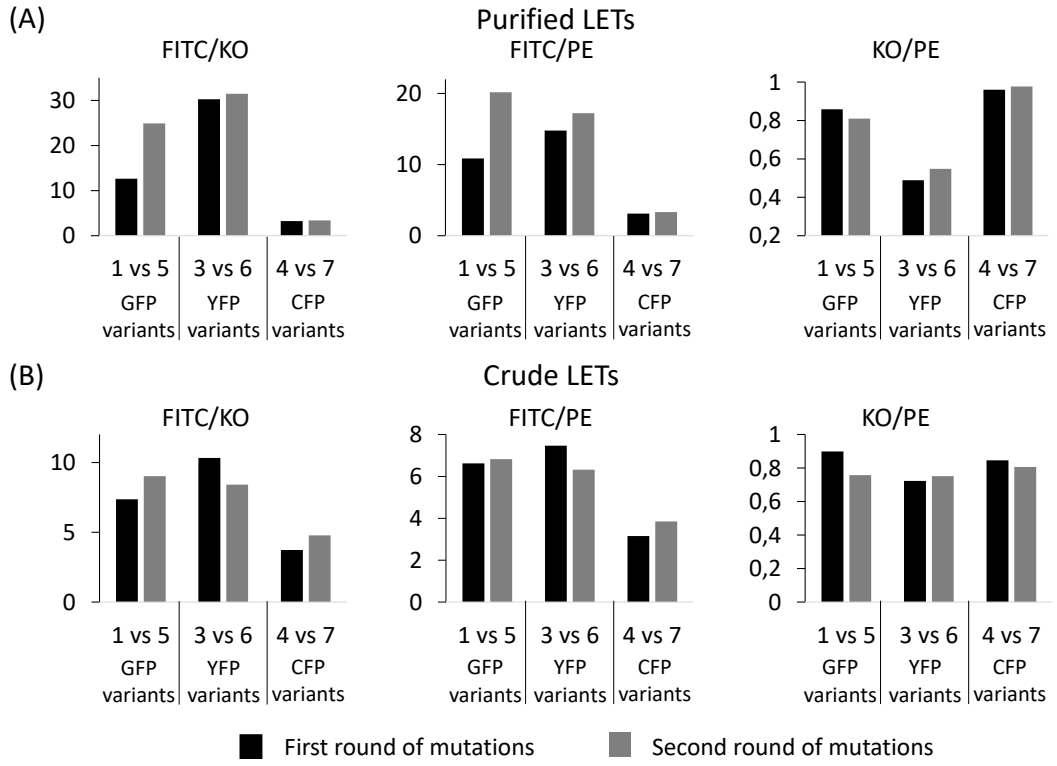


Figure 4.19: A: Result of different purified LET variants (65 nm) after 3 h incubation at 29 °C employed in TX-TL reaction. Shown is each variant in different channel ratios, whereas the units are a.u. The black bars indicate the first round of GFP variants, whereas the grey bars indicate the same GFP variant with additional mutations points in a second round of mutants; 1: GFPuv; 2: S65T, T203Y; 3: S65T, V68L, T203Y; 4: N146I, S175G, H231L; 5: F64L, S65T, R80Q, T153M, A163V; 6: S65G, V68L, S72A, T153M, A163V, T203Y; 7: F64L, S65T, Y66W, N146I. B: Result of different unpurified LET variants employed in TX-TL reaction for 3 h at 29 °C. The legend is the same as for A and can be seen additionally in Table 4.6.

4.4.6 All fluorescent protein variants produced from crude LETs

After these very promising results of the new mutants, all the LET variants were used in crude form for CFPS (see Figure 4.19 B). The different variants are distinguishable in the different channels of FITC, KO and PE. Nevertheless, the obtained signals are depending on the concentration of LET used. Here fluctuation between different LETs might exist because crude LETs with approximated concentrations are used. This challenge has been addressed by saturating concentrations of LETs for each case. Furthermore, since the LETs were not purified and tested, there might be some spontaneous mutations within one LET which could alter the property of a part of the LET population, this will influence the measurement outcome at the flow cytometer. By forming ratios for each mutant in different channels (see Figure 4.19), a relative investigation is possible and thus circumvents the (concentration-based) biased absolute intensity.

The results of the variants 2, 3 and 6 are noteworthy. All of them contain the same mutation points S65T and T203Y. According to literature, these should shift the wavelength spectrum of the fluorescent protein to the yellow direction. One mutation point more, namely V68L for mutant 3 shows an enhancement over mutant 2, which can be seen on the FITC/KO ratio and also the absolute intensities in Figure 4.19. The combination of these mutations was reported to exhibit brighter intensities by the flow cytometer method (Cormack et al., 1996). Further mutations as for mutant 6 (S65G, V68L, S72A, T153M, A163V, T203Y) did not lead to an enhancement of ratio or signal intensity, although also for mutant S72A a stabilizing effect on GFP variants at 37 °C is reported (in combination with S65 mutation, Cormack et al., 1996; Teerawanichpan et al., 2007) .

Based on GFPuv also a wavelength shift to the blueish emission wavelength was pursued. Therefore, mutant 4 (N146I, S175G, H231L) and mutant 7 (F64L, S65T, Y66W, N146I) were intended. Both show a slightly enhanced signal in the KO channel in comparison to the other mutants (except for GFPuv) which allows to draw conclusion on a shift to the shorter wavelengths, as it is the case for blueish GFP variants. For mutant 7, additionally a relative high signal is measured in PE channel, which hints towards a broadened wavelength maximum or an extended shoulder towards longer wavelengths. This phenomena can be described by the mutation point Y66W. It shifts the wavelength spectrum to the longer wavelength, by also decreasing the level of fluorescence (Lelimousin et al., 2009). The further mutations of M153T and V163A are found to be close to the chromophore. Although for the pure LETs the theoretical cost

function was met by these new mutants, in case of unpurified LETs this is not the case for the FITC/PE ratio. This result shows that problems might occur because of overlying effects of crude LETs with the final fluorescence outcome. A promising mutant could be missed, omitting it from closer investigation. Multiple measurements of each mutant with freshly prepared LET might ameliorate this problem.

The variants 2 and 3 contain S65T which is described to be increasing thermostability and higher brightness in CFP. The mutant 6 in contrary contains S65G which is described in yellow mutants, also increasing brightness and thermostability. Nevertheless, mutants 2 and 3 show better performance on the PE channel than mutant 6. Obviously, the combination with other mutants do not allow to nail an effect down to a singular mutation. The mutation points of T153M and A163V maybe decrease the fluorescence intensity of YFP variant. This mutation point is also there for mutant 5 which shows in general a good performance, thus these mutation points are dependent on synergistic effects with other mutation points.

4.4.7 Conclusion and Outlook

The creation of LETs as DNA template for CFPS via an assembly PCR approach enable easy gene manipulation. Exchanging individual oligonucleotides in an established set for the purpose of mutant generation was accomplished reliably and trouble-free. This modular set up offers the flexibility needed for point mutation generation to create variants without any need of cellular cloning procedures in high yield. Although the GFP variants expressed from purified LETs exhibit significantly higher fluorescence intensities, also the GFP variants expressed from unpurified LETs resulted in signals distinguishable from each other. A broad spectrum protease inhibitor could not help to enhance protein yields obtained by the cell lysate. Nevertheless, if longer incubation times for protein expression are applied in future, caspases might be employed to protect the protein of interest from degradation.
Furthermore, the incubation time for protein expression via CFPS can be shortened to two to three hours, which is sufficient to still measure significant differences between the individual mutants. Here, for some GFP variants an increase of temperature to 34 °C is advantageous for short incubation times. This is to be evaluated for each protein of interest, since properties like protein folding and solubility are influenced by the temperature as well.

The whole set up of LET generation, CFPS and sample preparation for flow cytometer readout could be established at a liquid handling system and is accomplished fully automated within less than six hours. In future an algorithm should evaluate the results and give suggestions for a next protein optimization round. To be able to judge the protein performance, for the role model of fluorescent proteins a cost function was established. This calculates ratios of each mutant in different channels and so decision can be made about their spectral shift.

Various mutation points where tested to shift the spectrum of the fluorescent proteins to shorter or longer wavelengths according to literature. This was successful for the yellow variants, but more troublesome for the blueish variants. Probably the generally decreased quantum yield for the shorter wavelength variants is troubling here. Nevertheless, proteins expressed from purified exhibited a fluorescent behaviour as described in literature and met the cost function. The same proteins expressed from unpurified LETs exhibited fluorescent signals which did not match the cost function in some cases. Some mutation points exhibited a severe impact on the spectral protein properties. The results were strongly influenced by the combination of different mutations, as synergistic effects of mutation points make the evaluation more complex. In future this should be taken into consideration.

This proof of concept for a new, fast and automated protein optimization work flow with the role protein GFP can be applied to other proteins of interest in future. Therefore, fusion proteins could be expressed and tested for fluorescence intensity to test for highest expression. Usually a N- or C-terminal fluorescence tag doesn't affect the fluorescence and the fused protein retains its functionality (Palm et al., 1997). For functionality testing of protein to protein interaction a set up using FRET events would be conceivable.

4.5 hPDC expression in cell free manner

The advantages of CFPS as described in the introduction will not only be used for a proof of concept study with GFP variants, but will help to investigate proteins of industrial relevance. The multimeric hPDC offers properties like cofactor regeneration or shielding of intermediates which can be exploited for enzymatic reaction cascades in another context than usual hPDC reactions. Therefore, producing hPDC subcomponents fast and flexible by CFPS will offer a versatile platform for protein engineering and verification.

4.5.1 Preliminary work using CFPS kits and CL

Coexpression of the hPDC subcomponents causes low protein yields as described in Appendix A.2.1. In order to ensure feasibility of hPDC subcomponent production in a cell free manner, preliminary work was carried out. Therefore, two different commercially available CFPS kits were chosen which are known to produce proteins in high yield, which is enough for simple SDS-PAGE detection. One is Expressway™Cell-Free *E.coli* Expression System from Invitrogen and one is PURExpress® In Vitro Protein Synthesis Kit from NEB. Both systems are *E.coli* based but do have different features. The later NEB system is based on Shimizu et al. (2001), all components needed for the CFPS are purified from *E.coli* and reconstituted to form the so called PURE system. The Invitrogen kit is based on a conventional *E.coli* lysate containing all the lysate background.

The Invitrogen kit was used according to manufacturer's instructions in feed mode and run for six hours at 37 °C. The reaction was successful for the positive control protein CALML3 (19.5 kDa) but no target proteins could be detected for hPDC subcomponents. Employed plasmids were pETDuet-1-N6HisE1α-E1β, pETDuet-1-E2-N6HisE3BP and pET28a$^+$-N6HisE3. One reason might be the large size of the hPDC proteins (36 to 59 kDa) in contrary to the positive control. Larger proteins usually result in lower protein yield in CFPS. By reducing temperature to 30 °C this obstacle might be ameliorated (Invitrogen, 27. September 2011). Moreover, the protein yield is divided over two subcomponents for the coexpression plasmids, which aggravates their detection in addition. Probably, increasing template DNA amount could also help to increase protein yield. Another source of error might be contaminated DNA template with salts, ethanol or RNAses. Here a fresh plasmid preparation would lead to success. However, latest error source is improbable, since plasmid was checked for contamination, plasmid purification was performed carefully to avoid ethanol contamination and RNAse inhibitor was employed. Protein aggregate formation, which is assumed for hPDC components, is generally a problem for

protein synthesis (Kopito, 2000). Here, decreasing temperature and in addition employing chaperons or mild detergents will emend the performance of CFPS and could lead to detectable protein synthesis (via SDS-PAGE) for the Invitrogen kit.

In contrary, the NEB kit is used in batch mode according to manufacturer's specifications. Here, target proteins were expressed and detected by SDS-PAGE (see Figure 4.20). The reaction duration was tested for 2, 3 and 4 hours, whereas no obvious differences in protein yield was detected by SDS-PAGE (data not shown). This indicates a limitation of substrates, DNA template or accumulation of inhibitory side products within the first two hours of reaction time, which could be overcome in a feed mode reaction.
By applying BSA in known amount onto the SDS-PAGE, a standard curve was established. The concentration of the positive control protein DHFR was calculated to be around 200 µg/mL. hPDC subcomponents are also detected but with lower yields. The bands for the subcomponents are marked with small blue arrows in the Figure 4.20. E3 exhibited a prominent band which was able to be quantified to around 120 µg/mL. Also E1α, E1β and E3BP are present, but at a noticeably lower yield compared to E3, thus they could not be quantified reliably. The subcomponent E2 was not visible at the SDS-PAGE. This phenomenon is also observed with conventional protein expression: E3BP yields are multiple times higher than E2 amount for the coexpression (this is discussed in the Appendix A.2.1 as well as in section 6.2). Here other detection methods like a highly specific antibody based western blot or an autoradiogram would increase protein signal intensity and could help to visualize expressed proteins with their sizes, also in a crude environment.
The PURE system exhibits advantages over a CL when it comes to controllability. To prevent insoluble protein aggregates, chaperones are added and reaction conditions can be tailored as needed for the proteins of interest (Shimizu et al., 2005). Furthermore, this system contains less nuclease and protease activity than a CL which increases protein yields of LET based CFPS (Tuckey et al., 2014). This might explain the immediate expression of hPDC subcomponents by PURE system without any optimization of the plasmids beforehand.

The results of the above mentioned test kit verified expression of hPDC subcomponents by CFPS. The *E.coli* lysate kit from Invitrogen did not show target proteins on a SDS-PAGE, although the positive control worked out. This problem was not investigated further because the PURE system from NEB expressed the

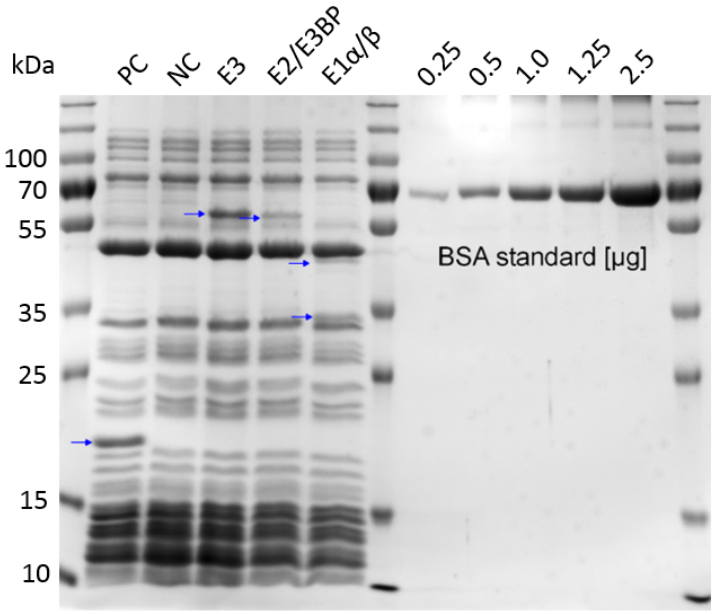


Figure 4.20: SDS-PAGE of hPDC proteins produced in cell free manner by using PURExpress® In Vitro Protein Synthesis Kit from NEB. The positive control (PC) is DHFR. E3 represents the hPDC subunit E3 (∼55 kDa). E2/E3BP represents the coexpression of E2 (∼70 kDa) and E3BP (∼55 kDa). E1 represents E1α (∼40 kDa) and E1β (∼36 kDa). On the right side a BSA standard is objected for protein quantification.

target proteins. As a next step, this proteins need to be validated for functionality. Moreover, the option to be expressed with an endogenous promoter as used for the TX-TL lysate requires further research.

4.5.1.1 hPDC subcomponent expression with different host

Because the CFPS established at this laboratory so far delivered protein yields in the low range of µg/ml, which is too low for activity testing, a cooperation with an institute specialized in CFPS was launched (Department of Cell-free and Cell-based Bioproduction, Branch Bioanalysis and Bioprocesses, Fraunhofer-Institute for Cell Therapy and Immunology, Potsdam-Golm). There, two cell lysates, *E. coli* and Sf21, were tested with all hPDC subcomponents. No proteins could be detected for *E. coli* lysate, very low yields were detected in case of Sf21 lysate. Based on these findings an optimization of codons of the hPDC genes to eukaryotic lysate was suggested. Moreover, the production of proteins was studied in a continuous CFPS mode to enhance protein yields. Hence, two problems arise, firstly there is a need to work with plasmids, a transfer to LETs is unlikely. Secondly, setting up an eukaryotic based lysate in a large amount for

a continuous mode CFPS complicates the process and is not simply feasible with laboratory capacity. Producing a complex protein like hPDC in a CF manner is challenging, even for specialized institutes with capacity and equipment. So far, for a reconstituted and thus defined cell free protein expression system, hPDC subcomponent expression could be shown, but no reliable activity could be detected (see section 4.5.3). Also, yields were too low to work further with these proteins.

4.5.2 hPDC expression at permanently induced conditions

As already mentioned, co-expressing hPDC subcomponents is challenging. This can be seen for low protein yields, but also special expression hosts used. For expression usually LysY cells were employed. These are cells that are regulated stricter than e.g. BL21 (DE3), so basal lever of target protein expression is suppressed and thus toxic proteins can be produced with that system (New England Biolabs Inc. 23.11.2017). To make sure that hPDC proteins can be produced with the TX-TL lysate, which is not utilizing the inducible T7 but an endogenous "constitutive" promoter, a test was set up. A conventional cell based expression culture was grown like described in material and method 3.4. Additionally, a second culture was inoculated and induced by IPTG simultaneously to see if this cells can reach a cell number similar to the conventional procedure. Specifications can be seen in Table 4.7.

Table 4.7: Culture properties of conventional induced and "permanently" induced cells, harbouring the plasmids for E1α-E1β and E2-E3BP coexpression. Cultures were inoculated 1:250 and expressed at 20 °C.

	OD_{600} induction	expression for x h	OD_{600} harvest	pellet weight [mg]
E1α-E1β induced	0.82	16	2.9	53
E1α-E1β permanent	0.04	18	2.6	52
E2-E3BP induced	0.92	16	2.2	65
E2-E3BP permanent	0.06	18	1.8	46

The permanently expressed plasmids were inoculated and induced at an OD_{600} of less than 0.1. After 18 h of expression they yielded in similar (E1α-E1β) or slightly reduced (E2-E3BP) cell density and pellet weight of that from the conventionally produced proteins (see Table 4.7). This shows a tighter regulation of expression like it is for LysY cells are better for co-expression of hPDC

subcomponents. Nevertheless, the cells are able to grow under the pressure of constant protein production, so hPDC subcomponents might be expressed in presence of the constitutive TX-TL promoter.

4.5.3 Functionality testing

Obtaining proteins and detecting them e.g. on a SDS-PAGE doesn't provide information about protein functionality. Therefore, an activity assay was performed to determine about their functionality. In order to test the functionality of hPDC subcomponents produced through cell free protein reaction by using PURExpress® In Vitro Protein Synthesis Kit from NEB, an overall hPDC enzyme activity assay was performed. The amount of enzyme expressed was calculated by densitometry of the SDS-PAGE (Figure 4.20). Since the enzyme amount produced is not sufficient for a conventional activity assay, for this set up lower enzyme amounts were used. A comparison between the usual set up and the used one is given in Table 4.8. However, the total protein amount in the assay with CF produced hPDC subcomponents is multiple times higher than stated for target protein content. The crude cell lysate contains already cell based proteins, and this was used for the activity assay measurement. All cell free produced subcomponents that can be seen in Figure 4.20 are reconstituted to form the complex. A graph results which has a high absorbance already at time point zero and shows a positive slope (data is presented in detail in the Appendix section 1.2.3). Due to missing protein negative control (crude lysate background) it cannot be said if the indicated slope results from enzyme activity or background/side reactions. The absorbance jump at time point zero is probably caused by the high total protein content in the sample which causes adsorption at 340 nm.
To reduce complexity because of uncertainties concerning different proteins, only one parameter will be changed at a time by keeping other parameters stable to receive better resolved information. Because E2-E3BP is the core part of the enzyme complex that is to be modified, this subcomponents were chosen for more intense investigations.

The overall hPDC activity assay is repeated again with only E2-E3BP produced by PURExpress® In Vitro Protein Synthesis Kit from NEB, the other subcomponents remained conventionally produced in cells. In this case, hPDC assembly for functionality testing consist of conventionally produced and purified E1α-E1β and E3, the E2-E3BP component was produced in a cell free manner by using the NEB kit. For the negative control, the E2-E3BP component was replaced by the DHFR positive control of the kit since it is assumed that this

Table 4.8: Target and actual hPDC subunit masses used for overall hPDC activity assay. All subcomponents were produced by PURExpress® In Vitro Protein Synthesis Kit from NEB.

	target value		actual value	
	µg/assay	ratio	µg/assay	ratio
E1α-E1β	10.29	1.86	2.40	1.32
E2-E3BP	6.79	1.23	1.80	0.99
E3	5.54	1.00	1.82	1.00

protein is not interacting with hPDC subcomponents (data and Figures with more details can be found in the Appendix section 1.2.3.2). Both curves start at a quite high basal absorbance of around 0.45 compared to the purified hPDC or negative control curves (around 0.1). This is probably due to side reactions of the protein rich background of crude E2-E3BP. Also, both curves (positive and negative controls) show a flat slope within the first minutes, which could indicate an enzyme activity. Because the negative control overlays with the mixed reconstituted hPDC, the slope originates from unspecific protein reactions. A flat slope was expected because E2-E3BP amount is lower than used in a conventional assay. The content of a 65 µL CFPS reaction was used after concentration with ultracentrifugal unit to 50 µL. The amount of E2-E3BP is roughly estimated to around 1.3 µg (usually 6.79 µg). The set up of the activity assay does not allow to add more than 10 % of reaction volume as protein content, so there is an upper limit of volume of subcomponents to add (which was exceeded in this assay by a factor of 2.5 already). In addition, it was not possible to measure the exact content of target protein in the crude background, so it is difficult to assign an accurate target protein amount.
In this context also a negative control reaction of a TX-TL reaction was added. An unusual curve with a steep slope for the beginning but also a steep negative slope after few minutes of assay time resulted. There is an interference of the TX-TL background with the absorbance activity measurement. It is thus not possible to subtract background reactions from a likely TX-TL produced hPDC subcomponent reaction.

Functionality of hPDC subcomponents produced in a cell lysate could not be demonstrated with an overall hPDC assay. Firstly, so far the enzyme amount is not sufficient due to low protein yield of CFPS. Furthermore, the crude protein background produces noise which cannot be filtered out with proper negative controls and overlays with possible occurring enzyme activity. Notable here is that the problem of uncontrollable background reactions during activity assay is not seen for the PURE system (data can be found in the Appendix 1.4).

This system is reconstituted out of purified components for translation and transcription machinery and thus has a defined component background, which does not show the undesirable activities of a CL (Tuckey et al., 2014).
More investigations on crude background in hPDC overall activity assays were made, thereby uncontrollable background reactions of crude cell lysate was revealed (data can be found in Appendix 1.6). Therefore, functionality testing of hPDC subcomponents in an overall complex assay is not possible with the set up of crude cell lysate environment. For a reconstituted cell lysate system such as the NEB kit this problem of uncontrollable background noise is not present, but still no significant active proteins were detected.

4.5.4 Conclusion and Outlook

By performing preliminary experiments with hPDC subcomponent plasmids, a production of corresponding proteins via a reconstituted, so called PURE system was verified. For a CL based system no proteins could be detected via SDS-PAGE. Next to the presence, hPDC functionality was investigated as well. Here an influence of the crude CL background for an hPDC activity assay was revealed. This prohibits the detection of functionality of hPDC produced in CF manner. Either a purification of the CF produced proteins need to be accomplished in future to perform the overall hPDC activity assay, or the change to a PURE system should be pursued since it triggers no background reactions.

The CFPS for hPDC subcomponents was simultaneously tested by a cooperation partner. They achieved low yields (few µg/mL) for hPDC subcomponent production via an insect cell lysate system. A change of the lysate system to an eukaryotic host could ameliorate the protein yields of the membrane associated hPDC. Already via conventional, cell based, protein production the coexpression of E1α-E1β and E2-E3BP is challenging, therefore more optimization efforts for the transfer to a CF expression must be made. These efforts can include the addition of chaperons or mild detergents, or a CF lysate with vesicles for the membrane associated proteins.

4.6 Short summary

Overall, a home made *E. coli* based cell lysate was established and characterized for the purpose of CFPS. To allow higher flexibility and save time, LETs were utilized as DNA template and produced via a fast and reliable assembly PCR method. Since fluorescent proteins served as role model to study different aspects of LET generation, CFPS and protein screening, a modified flow cytometer method was established allowing sensitive and robust detection of fluorescent proteins in crude bulk solution. This whole process chain was automated at a liquid handling station and thus can be used for a feedback guided, fully automated protein optimization loop. The production of hPDC subcomponents could be verified for a PURE CFPS kit, so far no activity could be detected for the hPDC proteins produced in CF manner.

Chapter 5

Nonlinear dynamics in hPDC

Herein, an unusual phenomena regarding hPDC function will be investigated: its nonlinear dynamic behavior at distinct reaction conditions. Therefore, the hPDC subcomponents as well as the substrate pyruvate were chosen as parameters for kinetic exploration. The results were linked to a mathematical model for the simulation of subcomponents, substrates and product to obtain a holistic understanding of the hPDC reaction mechanism *in vitro*.

5.1 Introduction

In biological systems the occurrence of periodic phenomena in enzyme catalyzed biochemical reactions are known at different levels and with different stimuli. Certain enzymes under certain conditions clearly indicate oscillatory behavior. Walter (1966) considered different reaction mechanism and evaluated them *in silico* after their ability to develop oscillations. Although screening for a large number and range of initial conditions and individual rate constants, no evidence for nonlinearity exists for simple one product, enzyme and substrate reaction, but for more complex cases. In reactions involving multiple enzyme intermediates, occasions for oscillation are evident. This rhythmic behavior can either be triggered by endogenous circumstances like enzyme regulation, or originate from external periodic processes like for the circadian clock (Goldbeter, 2013). One known class is the glycolytic oscillation discovered in yeast cells with an amplitude of five to ten minutes. Although glucose is fed at a constant rate, it is transformed in a periodic manner through phosphofructokinase in the metabolic pathway (Goldbeter, 2013). The enzyme class investigated in this study is the pyruvate dehydrogenase complex which represents a system consisting of

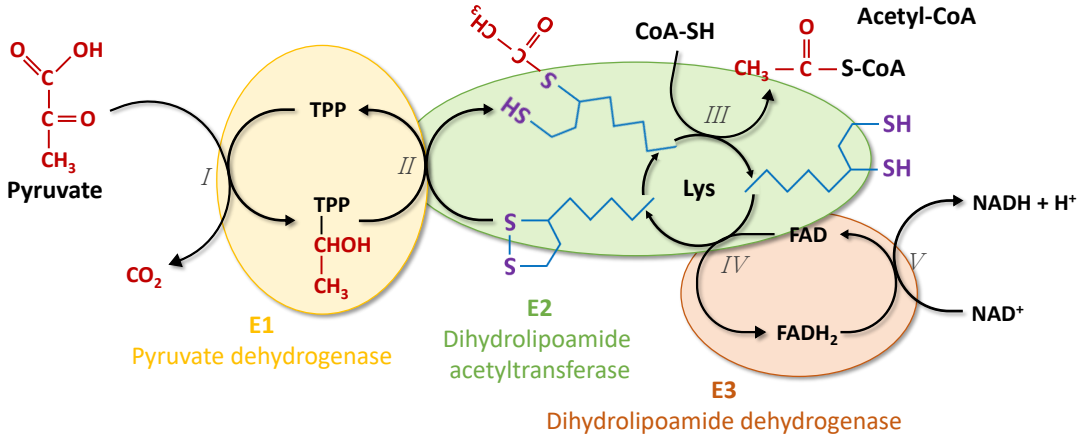


Figure 5.1: hPDC reaction mechanism. E1 (yellow), E2 (green) and E3 (red) work sequentially to convert pyruvate to CO_2 and acetyl-CoA, whereas E1 is the rate limiting step. I: Pyruvate reacts with TPP bound to E1 and undergoes a decarboxylation reaction. CO_2 is produced and released. II: E1 transfers the acetyl group of TPP and two electrons to the lipoyllysin group covalently linked to E2 to form an acetyl thioester. III: Transesterification reaction at E2, the acetyl group from the lipoyl-arm is transferred to the thiol of CoA. Thereby, the lipoyl group of E2 gets fully reduced and acetyl-CoA is formed and released to subsequently enter the citric acid cycle. IV: FAD mediated reoxidation of the lipoyl group by E3. V: The reduced $FADH_2$ is oxidized by cofactor NAD^+ by forming NADH + H^+. Figure adapted according to Nelson et al., 2008.

coupled dynamic enzyme reactions, and therefore is a likely candidate to develop nonlinear behavior. Its usual function is described and shown in Figure 5.1.

hPDC is a key node in central metabolism, and thus its regulation is crucial for stable metabolic flux. It is influencing the metabolic route in the respiratory pathway up and downstream of itself (Zeng et al., 2002). Goldstein and Selivanov (1993) described first unusual kinetic behavior for mammalian PDC as an exponent for α-keto acid dehydrogenase complexes. They set up a first kinetic model describing distinct states of the covalently bound lipoyl moieties. Later, Selivanov et al. (1994) showed theoretically undamped and time dependent oscillation. Menzel et al. (1998) studied E1 activity under different steady states in an anaerobic *Klebsiella pneumoniae* continuous culture. At a critical high dilution rate of 0.15 h^{-1} they described the phenomena of oscillation, at which E1 is utilized more efficient than at steady state conditions.
Zeng et al. (2002) extended the *in silico* model and investigated problems occurring from nonlinear dynamic behavior such as oscillations and multiplicity. Different steady states are reached for continuous cultures in dependence of

substrate concentration, whereas a lower pyruvate concentration is related to a more efficient decarboxylation rate of pyruvate. But also a time delay in enzyme synthesis leads to kinetic oscillation (Zeng et al., 2002). This refined mathematical model was not verified by aerobic experimental data so far. Since analytic resolution, process control and throughput via automation increased steadily over time, experimental data can now be collected to complement the digital twin. Considering hPDC consists of different subcomponents, they are worthwhile being investigated. Furthermore, the state of the substrate concentration (pyruvate) and cofactor concentration (CoA, TPP, NAD^+) could influence the dynamic behavior. Selivanov et al. (1994) and Zeng et al. (2002) suggest that this phenomenon can be seized experimentally by appropriate conditions. This might be lowering substrate concentration usually used in excess, and high resolution of the time course with oscillation occurring theoretically in a period of around 10 seconds.

Since the base of this work is built on theoretical investigation of hPDC metabolic control analysis, also a model based simulation of the complex was intended. For this work the open source software COPASI was utilized with its time course function and a built in LSODA solver for the ODE based dynamic simulation of the hPDC network. The model based work shall be validated experimentally, which will enable a fast screening of variable parameters with broad ranges.

5.2 *In vitro* hPDC assembly

The theoretical assumed ratio of the subcomponents for the hPDC assembly (E1:E2+E3BP:E3 1.515:1:0.816, see methods section 3.5.3) is probably not reliable for *in vitro* experiments. This is due to coexpression and purification of E2-E3BP. As can be seen on SDS-PAGE Figure 5.2, the ratio of E2 to E3BP is unbalanced and shifted towards E3BP. Therefore, the *in vitro* stoichiometry is not exactly equal to the theoretical calculations. The fraction of the theoretically assumed active trimer (consisting of $E2_2$ and $E3BP_1$) is probably present only to some extend within the purified E2-E3BP protein batch.
The His-tag utilized for protein purification is located N-Terminal at E3BP. Subunit E2 contains no His-tag and is co-purified because of its interaction with the core partner E3BP. Accordingly, only interacting E2 is purified, whereas in case of E3BP also inactive, single or agglomerate fractions are present and shift the ratio of the core subcomponents. As Vijayakrishnan et al., 2010 report, E3BP remains largely monomeric in solution and doesn't spontaneously self-assemble

like E2. In addition, the active protein yield is reduced since active E2 trimers (Prajapati et al., 2019) are missed completely.

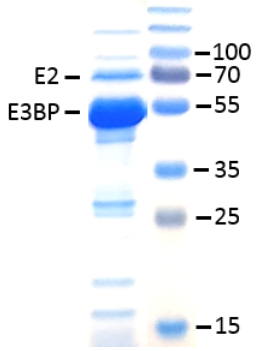


Figure 5.2: SDS-PAGE of IMAC purified hPDC core consisting of E2 and E3BP as depicted in the picture. Ratio of E3BP to E2 is not balanced because E3BP contains the His-tag for protein purification.

Table 5.1: hPDC ratio used for activity assays to study nonlinear kinetic enzyme behaviour of the complex in the first row. Following rows represent ratio of the three hPDC subunits by varying the amount of only one subunit and keeping the other subunit amount stable.

	E1α-E1β	E2-E3BP	E3
usual ratio	1.52	1.00	0.82
½ E1	0.76	1.00	0.82
¼ E1	0.38	1.00	0.82
⅛ E1	0.19	1.00	0.82
1/16 E1	0.09	1.00	0.82
½ E2-E3BP	1.52	0.50	0.82
¼ E2-E3BP	1.52	0.25	0.82
⅛ E2-E3BP	1.52	0.13	0.82
¼ E3	1.52	1.00	0.21
1/16 E3	1.52	1.00	0.05

It can be assumed that a part of the core proteins used for activity measurements were not active because some trimers might have a composition containing less than 2xE2. Also, E1 and E3 were added in excess for the full complex assembly (see method section 3.5.3). Combining these two effects, the total amount of E1 needed for complex formation is reduced, a part of the provided E1 probably remains unused and thus lowers the specific activity (mU/mg), but not volumetric activity of hPDC. This is underpinned for the result of the specific activity in Figure 5.3 B, where specific activity rises with lowering E1 content by ¼ in the complex assembly.

5.3 Investigations on hPDC subcomponent and substrate variations

In the following multi-parameter screenings considering initial pyruvate concentrations and variations for each hPDC subcomponent will be presented.

5.3.1 Parameters E1 amount and pyruvate concentration

In this section, variable E1 stoichiometries in the hPDC were investigated: 1 x E1, ¼ x E1 and ⅛ x E1 (ratio in the full complex can be seen in Table 5.1). Thereby, the amount of E1 for the complex assembly was lowered, the amount of the other subcomponents remained untouched. It was seen that the volumetric hPDC activity decreases, whereas the specific activity with ¼ amount of E1 increases in comparison to the usual E1 ratio (Figure 5.3 A and B). For further E1 dilution of ⅛, the activities decreased again (this phenomenon is discussed in the previous section 5.2).

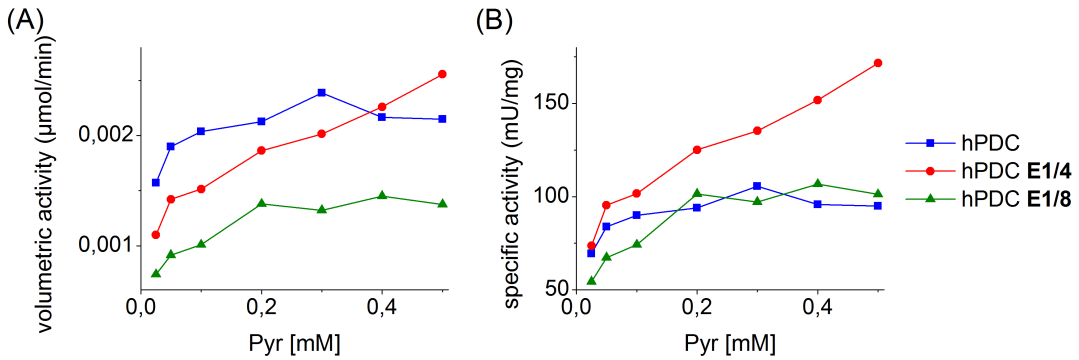


Figure 5.3: (A) and (B): volumetric vs. specific activity of hPDC over pyruvate range. Amount of E1 in the complex is decreased by ¼ and ⅛ respectively.

Thereby, different kinetic behavior can be seen for the activity assay plots sorted after E1 content in Figure 5.4 A, and the same data sorted after starting pyruvate concentration in Figure 5.4 B. For E1 in its highest amount (1x) quite sharp maxima at quite early time points were visible. For lower E1 amount in the full complex this maximum was smother and shifted to later time points. Furthermore, the number of periods of the transient oscillation increased for lower E1 amount. As shown in Figure 5.4 C, for a pyruvate starting concentration of 0.05 mM a full period is there for usual E1 amount. For ¼ E1 amount around two periods were present and for ⅛ E1 amount already three to four periods appeared. After these transient oscillations phased out, either immediately (high E1 ratio), or delayed (low E1 ratio) a decay of NADH started. This caused intersection of the lower E1 amount activity curves with the higher amount E1 activity curves.
The first period of the transient oscillation had a mean value in a range of few minutes and the amplitude leaked out fast, until it reached a new steady

state. When regarding the usual amount of E1 in the complex, hardly a transient oscillation was visible for pyruvate concentrations above 0.3 mM. Below 0.2 mM pyruvate, transient oscillation appeared and shifted to earlier time points for each decreasing pyruvate concentration. This behavior changed for E1 dilutions. Here, transient oscillations were also visible for more than double pyruvate concentrations (up to 0.5 mM initial pyruvate concentration was tested).

For E1 in usual amount the initial slope (0.0311 to 0.0275 A340 nm/min) is similar for a broad range of pyruvate (0.5 to 0.025 mM, see Figure 5.4 A). For ¼ E1, the initial reaction rate over the pyruvate dilution series diverges, which indicates E1 as the limiting factor. In case of ¼ E1, the initial slope ranged from 0.0371 to 0.0159 A340 nm/min and for ⅛ E1 from 0.0200 to 0.0107 A340 nm/min, accompanied with a changing activity curve shape. For high E1 amount at which pyruvate was a limiting factor, oscillation appeared quite early. For E1 ¼, when E1 is the limiting factor, oscillation appeared at same time point or later than 1 x E1.

By fixing E1 amount at different levels and varying pyruvate concentration, the turning point over the course of time in enzyme activity could be investigated (Figure 5.4 B). In general, for usual pyruvate concentration of 2 mM a monotonic saturation like activity behavior was present. Lowering pyruvate amount led to activity curves which show an increase of NADH over time, reaching a maximum and then declining over time. This behavior cannot be described by a monotonic function anymore. This turning point can be either sharp and mark the maximum of the activity slope, or smooth and lead to a moderate decay of NADH over time (E1 vs. E1 dilutions). For higher E1 ratios, oscillation appeared with or after the maximum, for lower E1 amount the oscillation appeared before the maximum. The diagrams depict single measurements, replicates can be found in appendix (see Figures 1.7 and 1.8). The replicates were not overlaid since the oscillations would diminish due to their phase shifts. The phenomena of NADH decay was not seen during the usual hPDC activity assay with 2 mM pyruvate starting condition, but first discovered during the multi-parameter screening at lower pyruvate concentrations. Hereby, a E1 amount in excess (usual hPDC *in vitro* condition) a steeper NADH decay curve was visible compared to the complex assembled with lower E1 amount. A steeper initial enzyme rate is coupled to a stepper NADH decay rate in the end. When normalizing all starting points of the different reaction conditions (Figure see appendix 1.8), for the lowest pyruvate concentrations measured, the NADH state nearly went back to zero within 100 min assay duration.

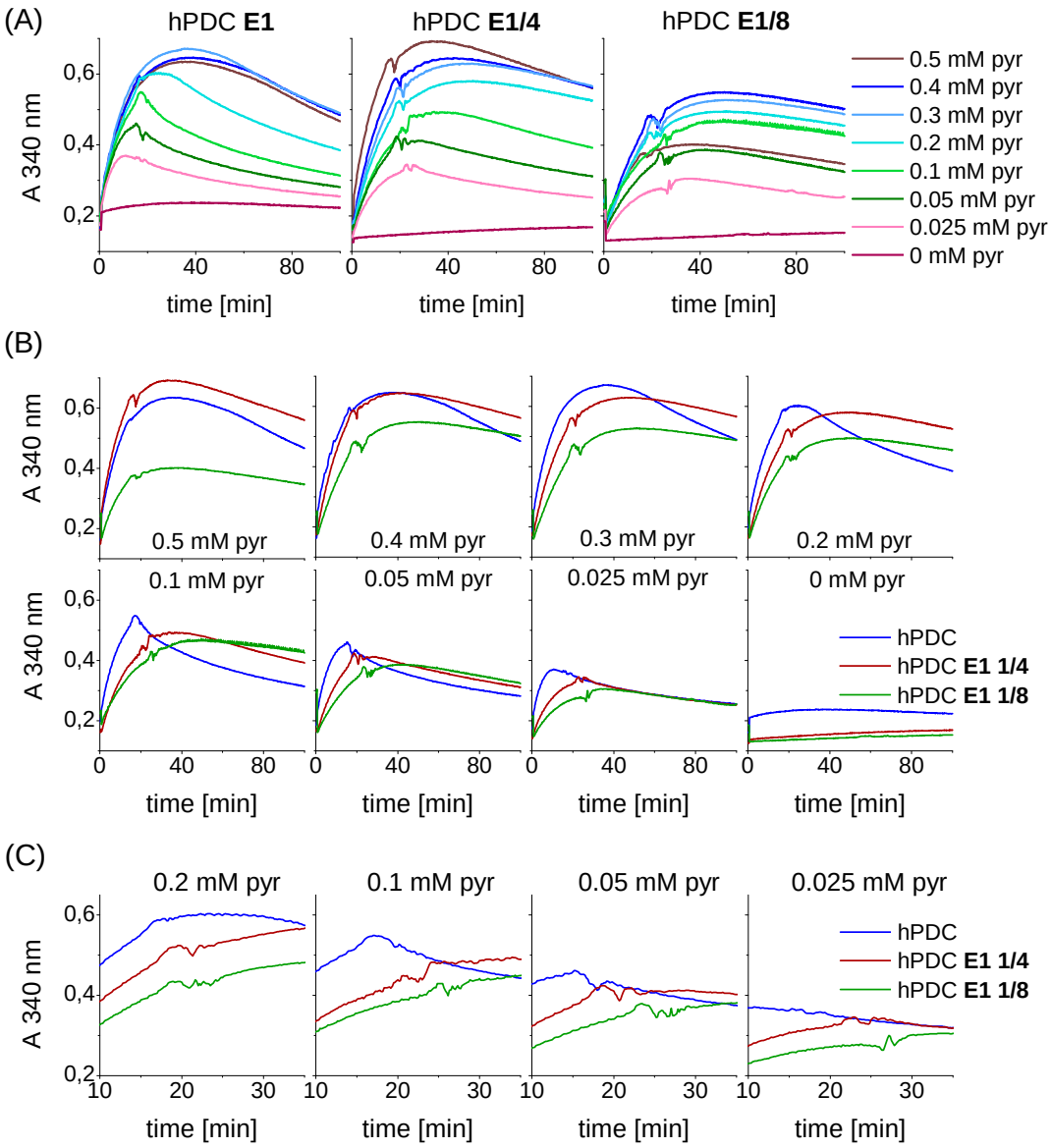


Figure 5.4: (A) and (B): hPDC activity assay curves. NADH absorption at 340 nm was plotted over time course. Two parameters were varied as depicted in the graphs: amount of E1 in the complex and initial pyruvate concentration. (A) is sorted by parameter E1 and (B) represents same data sorted by initial pyruvate concentration. (C): Some conditions from (B) have been enlarged to allow a better resolution of the transient oscillation. The conditions are depicted in the graph.

In these experiments, a decay of the reduced cofactor NADH measured at 340 nm over time was revealed for pyruvate concentrations lower than 0.2 - 0.5 mM . As shown in Figure 5.5, the maximum might indicate the point at which pyruvate is depleted, at least for the lower investigated pyruvate range. From this point on a baseline signal was expected, but instead NADH decays. There is evidence that NADH is unstable and decays already after short time when exposed to elevated temperatures or an unfavorable buffer composition or pH (Rover et al., 1998). Phosphate buffer thereby provokes a degradation of NADH because of adduct formation between phosphate and NADH. A 13 % degradation of NADH is reported after 40 minutes in phosphate buffer, pH 6.8 at 30 °C (Rover et al., 1998). This effect might contribute to the recorded NADH decay, but it is not clear whether there are other contributing factors as well. For a usual hPDC composition and low pyruvate concentrations, a decay of NADH of more than 50 % after 100 min at 37 °C is recorded. The NADH decay is only observed for physiological pyruvate concentration, for pyruvate in excess no global maximum is recorded. Since this maximum marks the point of NADH decay overtaking NADH formation (coupled to pyruvate consumption), this point must be shifted to late time points for higher pyruvate concentration. Probably other limiting effects come into play before the high pyruvate content is depleted, thus elongating assay time will not resolve better here. Nevertheless, at low pyruvate concentrations a correlation between signal height (which represents NADH) and indirectly pyruvate is there. The short reaction times allow to connect signal height to distinct pyruvate concentrations within a linear range up to 0.4 mM pyruvate. Afterwards, as mentioned, other limiting factors influence this correlation, which might be accumulating side products or upper threshold of the assay validity.

A relationship between the number of full transient oscillations and the E1 parameter was revealed. When the E1 proportion in hPDC decreased, the number of transient oscillations increased. As Zeng et al. (2002) showed, oscillation only occurs in a narrow range of pyruvate concentration. If E1 is not present in saturating amount, lower E1 amount in the complex corresponds to lower decarboxylation rate of pyruvate. So pyruvate is converted slower, which might lead to the effect of keeping the critical pyruvate range longer time and thereby elongating the range for oscillations. This can also be seen in the shift of the start point of the oscillations. The less E1 is there, the slower is the conversion rate of pyruvate and the later the critical pyruvate concentration for oscillation is reached.
An interesting point to notice is the reversed effect for E1 in saturating concentrations and pyruvate as limiting factor. For lower pyruvate concentration,

oscillation occurs earlier. Since the start pyruvate amount is lower and the initial decarboxylation rate for saturating E1 amount is stable, the critical pyruvate range for oscillation is reached earlier. For non-saturating E1 amount in the complex (e.g. ¼ or ⅛), the initial rate of the hPDC reaction is split up according to pyruvate concentration. Obviously pyruvate concentration is the driving force for the reaction rate and thus a predominant setscrew to influence hPDC reaction.

Lastly, in a closed experimental system (in which this work was conducted) transient, but no sustained oscillation can be detected. For an isolated enzyme reaction sustained oscillation can occur, but requires an open system in which the substrate of the regulated enzyme is provided in a constant rate (Goldbeter, 2013). A steady sate in substrate level would need to be realized to be able to keep undamped oscillatory behavior ongoing *in vitro*, this was shown for the *in vitro* model by Goldstein and Selivanov (1993) and Zeng et al. (2002). Utilizing a constant substrate level facilitates the detection of nonlinearities *in silico*. For a batch wise performance only transient oscillation can be realized.

5.3.2 Analysis of local and global maxima for E1 ratio and pyruvate concentration variations

As can be seen in Figure 5.5, activity assays were evaluated according to two characteristic points. One is the global maximum (glob) and one the first maximum of the transient oscillation (osc).
The time of their occurrence is plotted in Figure 5.5 A. Interestingly, the global maximum appeared stable over a broad range of pyruvate concentration at around 50 min for the low amount E1 content of ⅛. Only for pyruvate concentrations lower than 0.05 mM and higher than 0.5 mM the global maxima shifted to earlier time points. In contrary, the set up with 1 x E1 in the complex showed a linear relationship: increasing pyruvate concentration until 0.4 mM proportionally shifted the appearance of the maximum to later time points. After peaking at 0.4 mM pyruvate, the next global maxima at 0.5 mM shifted back to earlier time point. This might indicate a change in the steady state reaction. For E1 in a quarter amount the trend lies within the two preceding ones. It is noteworthy that the time point for the global maxima converges accurately at 0.5 mM pyruvate for all E1 dilutions. A similar trend was visible for the first maximum of the transient oscillation. For the higher pyruvate concentrations they appeared around 15 to 20 minutes of activity assay progress, which is clearly before the global maximum. However, with decreasing pyruvate concentrations the first maximum of the transient oscillation shifted to later

time points. Here, the different E1 levels split up, the lower the E1 content the later the maximum occurred. For the usual E1 amount this results in an overlay of global and oscillation maximum at 0.1 and 0.05 mM pyruvate, for the lowest recorded pyruvate concentration the global maximum was now followed by the maximum of the first transient oscillation. For E1 in quarter amount the maximum overlaid just at 0.25 mM and for the lowest recorded E1 amount the distinct maxima didn't diverge until 0.25 mM.
The transient oscillation is influenced by very low pyruvate concentrations (lower than 0.2 mM), and from here also from the E1 amount in the hPDC complex. For the usual amount of E1 the effect is minor, but very low pyruvate concentration in combination with very low E1 amount led to a late occurrence of the transient oscillations.

The signal intensity of the maxima followed a general linear trend, the higher the pyruvate concentration the higher the signal (see Figure 5.5 B). For the lowest pyruvate concentrations signal intensity was similar for global and transient oscillation maxima. For ¼ E1 and ⅛ E1 with increasing pyruvate concentration the global maxima exceeded the oscillation maxima in regard of signal intensity. For the usual E1 amount the two parameters were close together and reached a peak at above 0.3 mM pyruvate, at which the signal intensity was not rising anymore. From 0.4 mM initial pyruvate concentration signal height is decreasing, from this point forward pyruvate is not the limiting factor anymore.

To provide a rough approximation whether the global maximum indicated the time point at which pyruvate was depleted, the theoretical time point at which the initial pyruvate amount should be consumed was calculated. Therefore, the pyruvate concentration is divided via the initial NADH formation slope. The resulting time point should give a theoretical hint on when pyruvate is completely depleted. It should be noted that the theoretical value will be faster than the experimental value, because the reaction rate slows down over time and pyruvate is not consumed with the full speed of the initial rate. As can be seen in Figure 5.5 C, for the condition of 1xE1, theoretical and experimental values matched. As described above this is the condition at which pyruvate must be limiting. Thus the global maximum indicated the time point at which pyruvate was depleted. For the E1 dilutions theoretical and experimental values didn't match, experimentally more time was needed until pyruvate was depleted than suggested theoretically on the base of the initial rate. Remarkably, at around 0.4 mM pyruvate, the theoretical and experimental values for all E1 dilutions were close together.

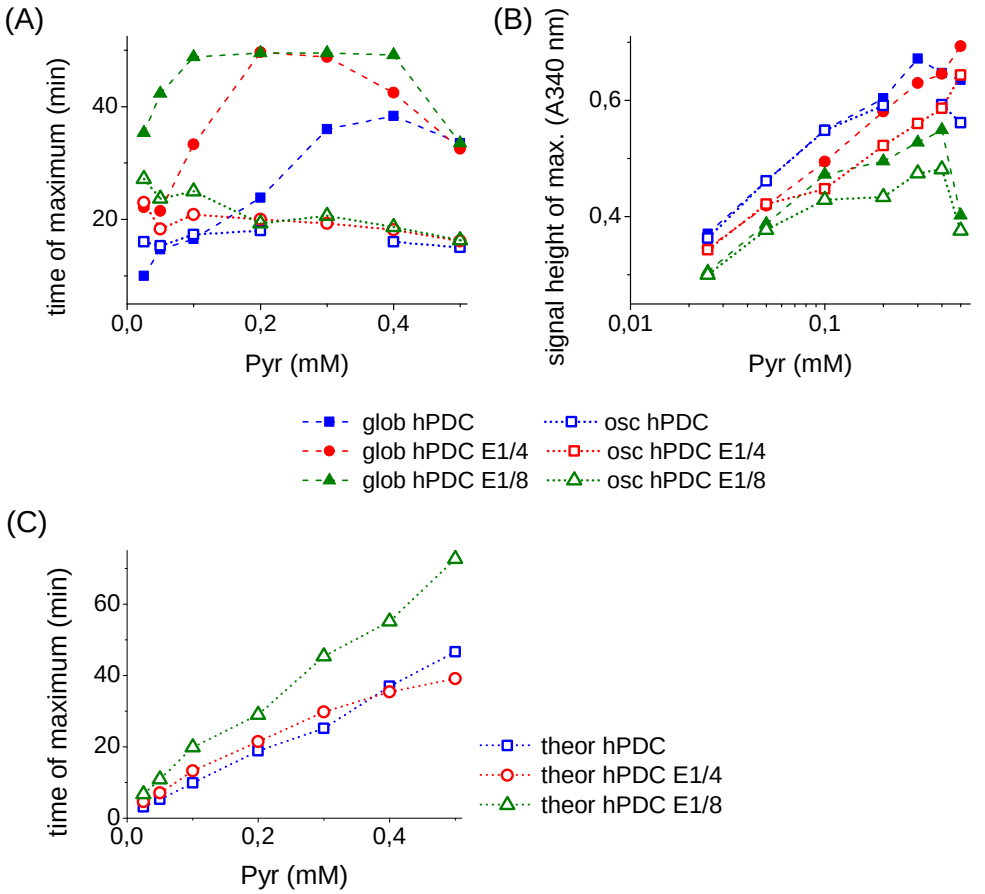


Figure 5.5: Evaluated results of hPDC activity assay with E1 and pyruvate dilution series. (A): Time in minutes at which a maximum during the activity assay occurred. The global maximum is abbreviated with glob, the first maximum of the transient oscillation is abbreviated with osc. (B): Signal intensity at 340 nm at which the maxima occurred over logarithmic pyruvate concentration. (C): Theoretical time of pyruvate depletion (theor). This was calculated by using initial pyruvate concentration via initial NADH production rate.

Overall it is seen for pyruvate concentrations below 0.2 – 0.5 mM that previous saturation like behavior for product formation of the hPDC activity assay changed to a NADH decay for the late time points. Usually, 2 mM pyruvate was used for the activity assay. However, the oscillations were observed for 0.4 mM pyruvate or less in case of the usual hPDC ratio. For lower amount of E1 in the complex, substrate concentration can be enhanced by keeping oscillatory behavior. When studying the appearance of maximum some effects were seen at 0.4 mM pyruvate concentration. Here, theoretical and experimental pyruvate depletion matches for all E1 dilutions. This could hint to a threshold concentration at which the hPDC mechanism changes from one steady state to another steady state, at least for the 1xE1 ratio, the for the other E1 dilutions this threshold is likely shifted to higher initial pyruvate concentrations. For *E.coli* an average intracellular pyruvate concentration of 390 µM is stated (Albe et al., 1990). This supports the pyruvate area around 0.4 mM as a likely threshold concentration for hPDC regulation. The transient oscillation marks a transition state at which the hPDC reaction rate changes. Before, the reaction rate is faster than after the transition. According to Zeng et al. (2002) high pyruvate state results in an less effective decarboxylation rate and vice versa. Since this observation was made for an continuous process in which the pyruvate concentration is stable over time, this cannot be directly compared to the experimental data performed in batch mode.

For low pyruvate concentrations (less than 0.5 mM), a global maximum was discovered during activity assay progress. In case of the usual hPDC ratio, this maximum could be assigned to pyruvate depletion. For the E1 dilutions, the theoretical calculated pyruvate depletion and experimentally observed maximum don't match. This might have several reasons. Firstly, for lower E1 amount and pyruvate, more periods of transient oscillation shifted to later time points are present. This might elongate status quo, and also the initial reaction rate slows down quickly for the low E1 content activity curves. Furthermore, pyruvate is not the limiting factor in the activity rate of hPDC. This is why the calculation for the pyruvate depletion is not valid here.
Another point to note is that the experimental feasibility of a phenomena is a question of the assay resolution. As Higgins (1967) stated "If assay of the concentrations are taken at time intervals greater than the period of the oscillation, the resultant data will appear to form a curve with some interesting humps but not period oscillation." To be more concrete, sampling rate with a frequency would have to be at least twice as large as the frequency of the signal (Nyquist frequency Kammeyer, 2004). This would allow a clear reconstruction of the oscillation, otherwise non-linear distortions would be obtained. Theoretically, Selivanov et al.

(1994) reported an oscillation period of 10 sec, Zeng et al. (2002) up to 30 sec at appropriate conditions. The intervals of a measurement should theoretically thus be below, whereas these values do not represent the experimental sampling rates. Screening for multiple conditions simultaneously was only possible at a minimal interval time of 10 sec for the Multiskan Spectrum spectrophotometer used. This indicates that some information might not be captured, sampling frequency and assay resolution were not high enough. Probably better information about period and amplitude of the transient oscillations could be collected for narrowing the time intervals.

5.3.3 Parameters E2-E3BP amount and pyruvate concentration

The data presented in this, as well as the next subsection (section 5.3.4) were collected at a liquid handling station. After the experiments were conducted, a miss-calibration of the liquid handling station was revealed. In general, the volumes pipetted were slightly lower than adjusted, whereas this phenomena was pronounced stronger for low volumes than higher volumes. A set-off of 3.76 µl and a scaling factor of 1.04 were missing. Therefore, this data was handled qualitatively to obtain a possible trend for E2-E3BP dilutions in PDC.

For a ¼ E2-E3BP amount in the complex, specific activity was higher than for usual E2-E3BP ratio (see Figure 5.6). Further E2-E3BP dilution led to a decrease of specific activity again.
The maximum of the specific activity for all E2-E3BP dilutions were around 2 mM pyruvate, more or less pyruvate led to a reduced activity. Thereby, E2-E3BP in ⅛ displayed around half activity of E2-E3BP in ¼ amount over a broad pyruvate range. This disclosed E2-E3BP as limiting factor here. For 0.075 mM pyruvate and lower, the low substrate concentration dominated. The signal height as well as the specific activity of usual complex formation overtook the ¼ E2-E3BP diluted one.

Occasionally transient oscillation was visible for pyruvate concentrations lower than 0.2 mM pyruvate. However, there were no systematic transient oscillations as seen for the E1 dilutions (see Figure 5.7 A and B). A pyruvate concentration of 0.5 mM or more showed saturation like behavior of hPDC activity. Here, no transient oscillations were visible, but instead a small hump over the time course. For lower pyruvate concentrations, the decay of NADH was sifted to later time points. This turning point also marked the transition from one kinetic behavior

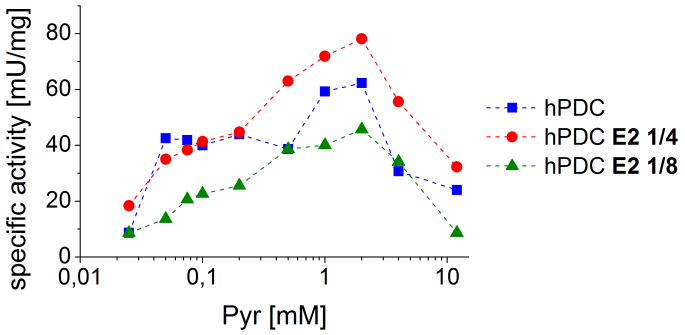


Figure 5.6: Specific hPDC activity over a broad range of pyruvate concentration and three different E2-E3BP diluted hPDC assemblies. To shorten the legend E2-E3BP is depicted as E2 only.

to another. The turning point was very sharp for the different E2-E3BP dilutions over different pyruvate concentrations, and accompanied by a small hump. For the usual E2-E3BP and ⅛ of E2-E3BP amount, the oscillations started around the same time point (20 to 26 minutes) over different pyruvate dilutions, and followed no obvious trend (see Figure 5.8 A). For the ¼ E2-E3BP amount this was different, for lower pyruvate concentrations the transient oscillation shifted slowly to later time points. The humps were present up to a pyruvate concentration of 0.4 to 0.5 mM (see Figure 5.8 B).
The highest pyruvate concentration of 12 mM showed a small hump in the first minutes, this matched observations made for the case of pyruvate excess presented in section 5.4.1.
Because of the miss-calibrated liquid handling station used in this experiment (slightly underestimating volumes), the data collected here must be regarded carefully. The given concentrations in Figure 5.7 might not be fully accurate, but the general trend might be valid. When assuming E3 is not causing non-linear effects for the protein complex activity (see section 5.3.4 and Zeng et al., 2002), a decrease of E2-E3BP amount can be set equal to an increase of E1 amount in the protein complex. The curve shape of the E2-E3BP dilutions actually show a very pointed maximum at the transition state (see Figure 5.7). This is similar for the usual hPDC ratio activity assay. If E1 is diluted, this shape changes to a flat and smooth maximum (see Figure 5.4). In this respect the indirectly increased E1 ratio (by diluted E2-E3BP) matches the trend of the curve shape, the higher the E1 content in the complex, the sharper the maximum is. Further, for the usual hPDC ratio, occasionally one full period of transient oscillation was visible which increased with decreasing E1 ratio. For the indirect increased E1 ratio only a hump is visible which would

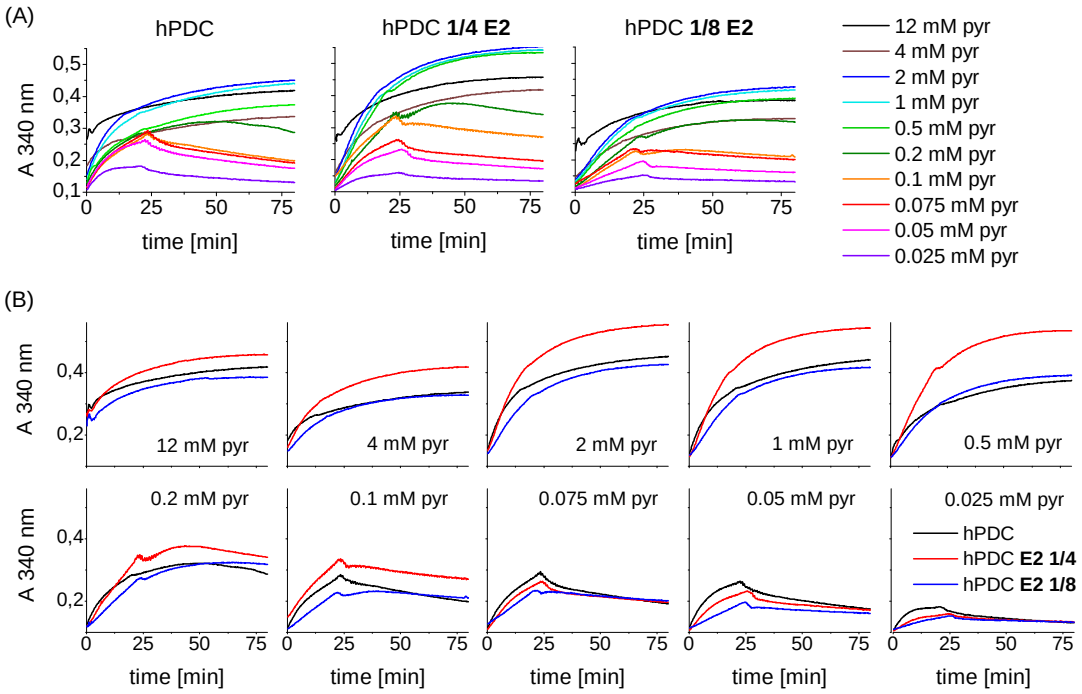


Figure 5.7: Over activity assay of hPDC E2-E3BP ratio and initial pyruvate concentration dilutions whereas absorption at 340 nm is plotted over time to track NADH formation. (A): Data sorted by the three different hPDC assemblies as depicted above the graphs. (B): Same data as (A) but sorted by initial pyruvate concentrations.

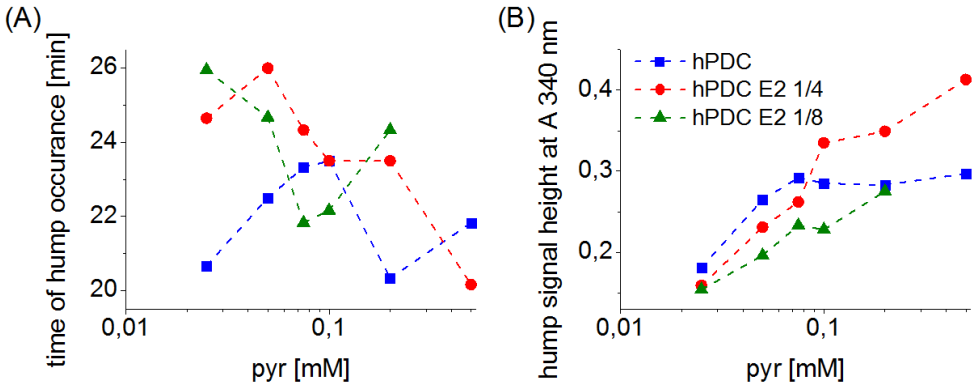


Figure 5.8: Analysis of hump (possibly overswing or a short transient oscillation) over a range of pyruvate concentration. The legend depicted in (B) is valid for both graphs. The local or global maximum occurring at the hump was analyzed for its time to appear (A) and the signal height at this time point (B).

also complete the consistent trend. This hump could represent an overshoot reaction, which is an oscillatory response, but without periodicity (Higgins, 1967).

No clear evidence for transient oscillatory behavior can be seen for E2-E3BP. The time points to which the possible nonlinearities occurred were randomly distributed over the pyruvate concentration and just sporadically visible. The recorded hump might be due to appropriate pyruvate conditions (as seen for E1 and pyruvate dilution experiments, lower than 0.5 mM pyruvate), not necessarily to E2-E3BP dilutions.

5.3.4 Parameters E3 amount and pyruvate concentration

The data presented in this section was collected at a liquid handling station slightly miss-calibrated as shown in the section above (section 5.3.3). In general, volumes pipetted were slightly lower than adjusted. Therefore, the collected data was handled qualitatively to obtain a possible trend for E3 dilutions in hPDC.

When performing E3 dilutions in hPDC, the protein complex activity decreased drastically. Over a broad range of pyruvate concentration a maximum for specific activity could be detected, usually around 2 mM pyruvate. For lower and higher pyruvate concentrations specific activity dropped. In this particular case the maximal specific activity for usual hPDC stoichiometry was at 0.5 mM (see Figure 5.9 B). Here, when having a closer look at the initial slope, it can be seen that the slope showed one elongated oscillation for the first few minutes, starting in the fourth quadrant and rising steep upwards. This caused the initial slope to be steeper than it would be without the distortion at this time point.

The hPDC assembly with 1xE3 and high pyruvate concentrations (more than 0.5 mM) showed typical, saturation like activity assay curve (see Figure 5.9A). Furthermore, there was no NADH decay within 80 min assay duration for the higher pyruvate concentrations (0.5 mM to 12 mM). For the pyruvate concentrations below 0.2 mM a NADH decay occurred, like described for E1 dilutions. For E3 in ¼ amount a consistent decrease in activity was detected (see Figure 5.9 B). For this condition a uniform specific activity over a broad range of pyruvate concentration (0.2 – 12 mM) was calculated to be between 21.8 to 19.1 mU/mg, which indicates that the rate limiting step here must be E3. Only for pyruvate concentration lower than 0.1 mM a slight decrease in activity was detected. A 1/16 E3 dilution led to a complete loss of activity. The kinetics showed a linear relationship, there was no indication for systematic transient oscillation. Nevertheless, because of the miss-calibrated liquid handling

station used in this experiment (slightly underestimating volumes), the given concentrations in Figure 5.9 might not be fully accurate, but the general trend might be valid.

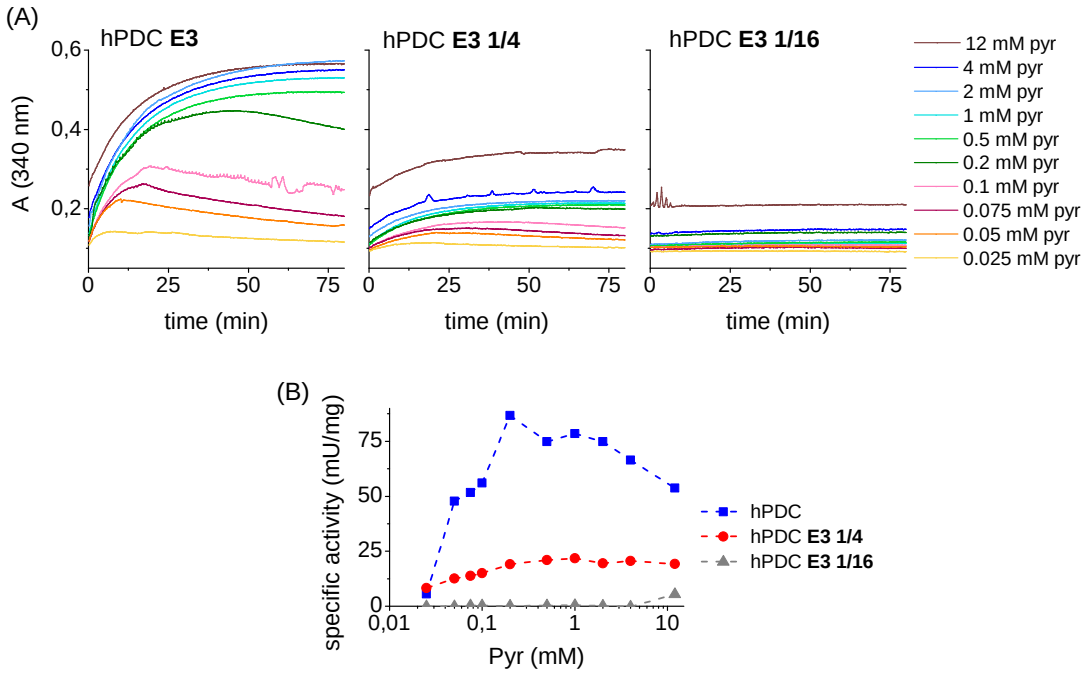


Figure 5.9: E3 and pyruvate dilutions in hPDC activity assay. (A): Activity assay of E3 and pyruvate dilutions whereas absorbance at 340 nm is plotted over time to track NADH formation. The E3 dilutions are depicted in the graphs. (B): specific activity of different amount of E3 in hPDC in the standard assay over a logarithmic range of initial pyruvate concentration.

The experiments for E3 dilutions revealed that stoichiometry variations with regard to this component do not result in nonlinearities or oscillations in the enzyme reaction. This matches the study of Modak et al. (2002). They already suggested in the *in vitro* model that only E1 subcomponent of PDC is sensitive to low pyruvate concentrations. E3 is not directly involved in the catalytically decarboxylation of pyruvate and the subsequent acetyl-CoA formation, which might be an explanation. Instead, E3 is required for the regeneration of the lipoyl moieties at E2 by a consequent reduction of NAD^+ to NADH (Patel et al., 2014). Because of this crucial regeneration property of E3, its dilution impacts the complex activity strongly. The hPDC overall activity assay detects the formed NADH at 340 nm. If the content of E3 is too low to catalyze the regeneration of E2, no further NADH can be formed and detected.

5.4 Investigation on total hPDC amount

At standard assay conditions described in the method part (3.5.3), the amount of hPDC used for an assay was 22.6 µg. It can be seen that the amount of 5.66 µg total protein, by keeping hPDC subcomponent ratio stable, displayed the highest specific activity under these conditions (see Figure 5.10 A). Nevertheless, the usual amount off 22.6 µg protein exhibited the largest signal in the assay (Figure 5.10 B), more protein was not enhancing the signal further and specific enzyme activity dropped. Because a standard pyruvate condition of 2 mM was used, the curve shapes displayed a conventional saturation like behavior without indications for oscillation or NADH decay.

To perform a multi-parameter screening of proteins over a broad range of concentrations, a large amount of protein is needed. Therefore, efforts were made to scale down total enzyme amount. When using only 5.66 µg hPDC with the standard assay conditions, no transient oscillation was observed within 100 minutes of assay duration. Here, pyruvate as well as E1 amount were varied without any sign of nonlinear dynamic behavior (data not shown). Instead, for this low total protein content, and in addition diluting one enzyme subcomponent, the overall activity was heavily influenced. Only low or no activity was detected at mentioned conditions. For the usual hPDC content of 22.62 µg and additional dilution of enzyme subcomponents, still well resolvable activity signals were detected. A reduction of hPDC by ¼ revealed highest specific enzyme activity, but decreased volumetric activity. This might be one reason why no oscillatory behavior could be detected for the low protein amount. The sensitivity of the assay might be insufficient to resolve the low amplitude transient oscillation (Higgins, 1967). The sensitivity of an assay is in general a limiting factor of experimental validation.

5.4.1 Pyruvate in excess conditions

The lower total protein content of 5.66 µg was used to screen a broad range of parameters. This protein content was selected to save proteins and still obtain good measuring signals. Not only low pyruvate, but also high pyruvate concentrations up to 40 mM were tested. A delay of enzyme activity was visible for high pyruvate concentrations, starting from above 12 mM (Figure 5.10 C). Here, a small hump over few minutes can be seen. Increasing the pyruvate concentration up to 40 mM elongates the lag phase to a plateau. This plateau lasted for some conditions more than ten minutes before an actual NADH formation was detected. Thus, no initial enzyme activity could be determined. After the delay, the following activity curve was flat and passed over into a

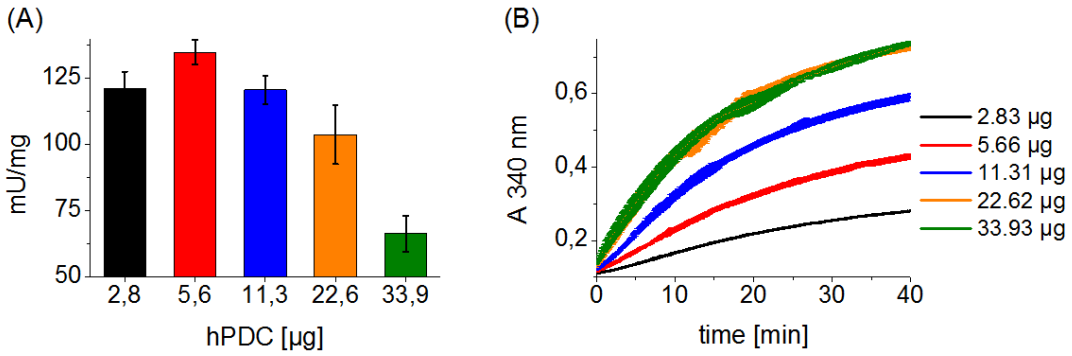


Figure 5.10: (A): Specific activity of different amount of hPDC in the standard activity assay with 200 µl volume. Error bars indicate triplicate measurement. (B): Activity assay of (A), whereas absorbance at 340 nm is plotted over time to track NADH formation. Error bars indicate triplicate measurement.

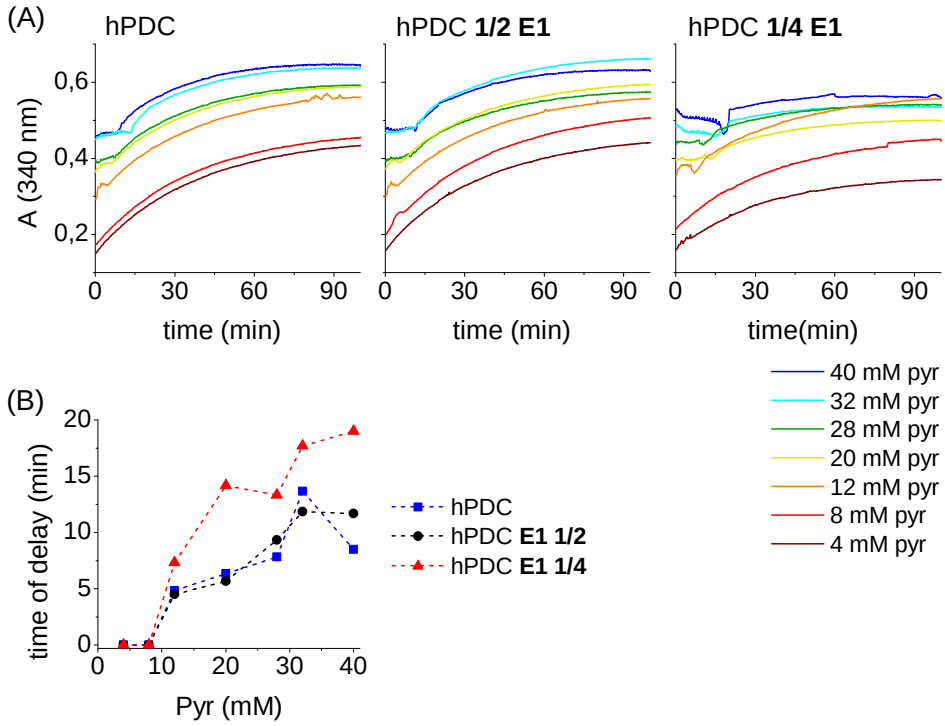


Figure 5.11: (A): Activity assay of 5.66 µg total hPDC amount in case of 1xE1. Different E1 and initial pyruvate concentration dilutions were tested as indicated next to the graph. (B): Duration of time delay until enzyme activity starts (last local minimum before positive enzyme slope) over pyruvate concentrations.

saturation like shape within 100 minutes of assay duration.
For ¼ E1 amount in the complex, the lag phase was further elongated as for usual E1 or ½ E1 amount. As shown in Figure 5.11 B, for the highest pyruvate concentration of 40 mM nearly 20 minutes of lag phase were present. Moreover, independent measurements with excess pyruvate conditions were conducted whereas the assay duration and pyruvate concentrations are not identical for all cases. Nevertheless, the total hPDC content was kept stable at 5.66 µg (for the case of stoichiometrical 1x E1 content). As can be seen in the appendix Figure 1.9, the lag phases varied from experiment to experiment, and were not visible in each case.

Pyruvate increases the baseline level of the NADH based activity assay measurement. Thus, an experiment was launched in which pyruvate was added to the reaction buffer, and the baseline was recorded. For 4 mM pyruvate an absorption level at 340 nm of 0.12 was measured, for 20 mM pyruvate the absorption level increased to 0.23 (see appendix Figure 1.10). However, the baseline levels were far below the plateau level of the activity measurements with pyruvate in excess. Thus, an interference of baseline shift with activity signal could be excluded.

In this study a pyruvate range was tested which is in great excess (more than 20 times) compared to the pyruvate conditions for the nonlinear kinetic studies. At excess pyruvate conditions, a delay in enzyme activity was recorded. Thereby, activity curves for usual hPDC stoichiometry and ½ E1 amount show very similar behavior. This might be due to the excess conditions of E1 in the usual stoichiometry set up as discussed in section 5.2. For E1 dilution of ¼ an elongation in enzyme delay time is observed. This concludes to E1 affecting and being at least one reason why the temporary enzyme inhibition occurs. A delay in hPDC activity is already reported for the E1 cofactor thiamine pyrophosphate in low concentrations (Strumiło, 2005; Sümegi and Alkonyi, 1983). This cofactor acts as an allosteric modulator which enhances enzyme to substrate affinity. Therefore, a possible substrate overflow inhibition in an allosteric way might be present. To prevent or shorten the enzyme lag phase a higher thiamine pyrophosphate concentration might be sufficient.
Xiu et al., 1998 report multiplicity for metabolic overflow which appears at substrate in excess conditions. Although Xiu et al., 1998 investigate a continuous culture, he describes unusual dynamic behavior for an enhanced product formation rate coupled with product inhibition. According to Xiu et al., 1998, prolonged time period for reaching a steady state as well as oscillatory transitions and sustained oscillation were phenomena that are present at named conditions.

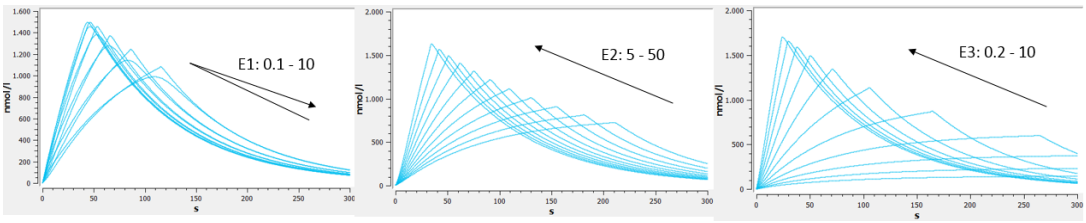
5.5 Model for hPDC reaction simulation

A modified model according to Zeng et al. (2002) was used to simulate the experimental behavior of the *in vitro* activity assay. To represent the batch like mode of the experimental activity assay, two modifications were introduced. The substrate pyruvate was added to the reactions and the detected product NADH was included to the model as a species with a low decay rate (see Appendix Table 1.5. The initial concentration of pyruvate was investigated in a range of 0.1 to 4000 nmol/L. The absolute concentrations of the model don't match the experimental concentrations. This is probably due to the theoretical values designed for an continuous process with optimal conditions. Therefore, the implemented new parameters of the model had to be estimated and adjusted qualitatively.
All parameters and their nominal values of the model as well as the reactions can be found in the appendix, Tables 1.5 and 1.6.

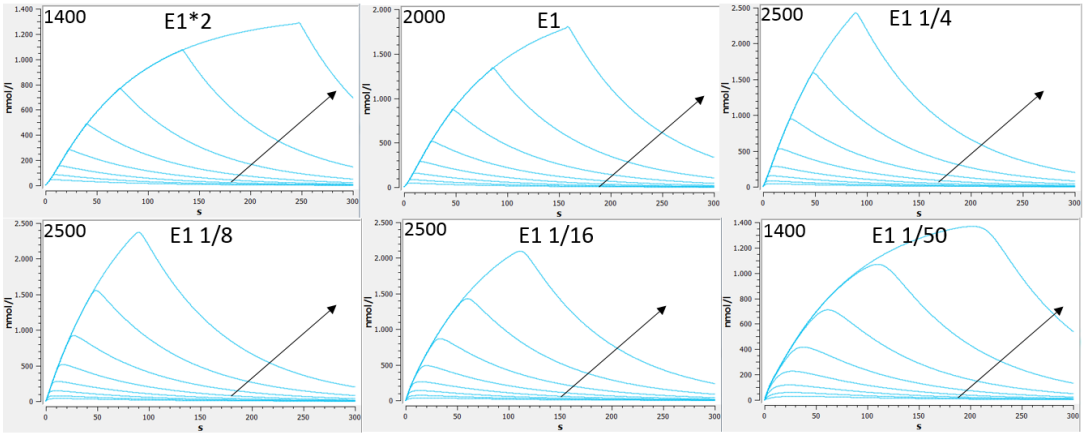
A dilution series of initial pyruvate concentrations at fixed E1 ratios in the full complex were simulated (Figure 5.12b). The model showed the lower the E1 level in the complex, the flatter the maximum appeared. For all cases an increase of pyruvate concentration led to higher NADH signal shifted to later time points. This reflected the experimental behavior of the full complex with E1 in excess amount. When zooming in into the graphs with a sharp maximum, an overshoot, which might represent a transient or non-periodically oscillation at this condition, was visible (see Figure 5.12c). This was more pronounced for decreasing total hPDC amount without varying the subcomponent ratio from its nominal value (see Table 1.5).

In Figure 5.12a the influence of the hPDC subcomponent variation is presented. The higher the E1 amount, the earlier the inflection point occurred and the sharper the maximum was. This corresponded qualitatively to the experimental E1 dilutions (see Figure 5.4). Additionally, the model delivers information for further increasing the E1 amount in the complex. The maximum started shifting to later time points again and was very sharp before NADH decrease followed.

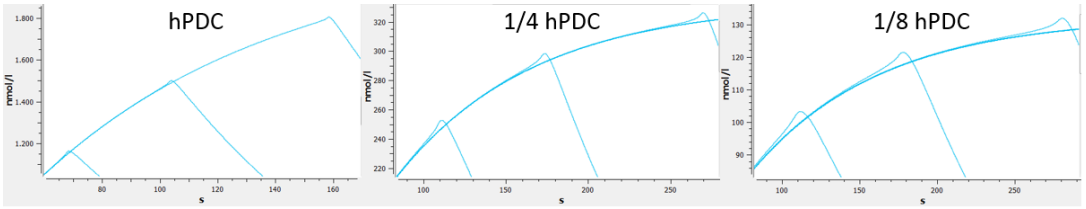
E2-E3BP and E3 subcomponents variations were also investigated over a broad pyruvate range (E2-E3BP variations: 5 - 50 nmol/L, as ratio around 1/3 to 3x and E3 variations: 0.2 - 10 nmol/L, as ratio around 1/10 to 5x). Both exhibited the same trend: for increasing subunit amount the initial reaction rate was also increased. Consequently, pyruvate was consumed faster and the maximum was reached to earlier time points. Afterwards, a decay of NADH took place. Hence,



(a) Simulation graphs of hPDC subcomponent variations for hPDC activity. Concentration of formed NADH is plotted against time. Initial pyruvate concentration is 2000 nmol/L. In each graph one hPDC subcomponent is varied by keeping other parameters fixed. For E1 dilutions (leftmost graph), nonlinear reaction behavior for increasing E1 amount can be seen. For lower E1 values of 0.1 - 1, the maximum is smooth and flat and migrates to earlier time points for increasing E1 amount. If E1 amount is increased further (above 1), a turning point is reached at which the shape of the maximum is sharp, and the maximum starts migrating back to later time points. E2-E3BP (middle graph) and E3 (right graph) were also tested as depicted in the graphs.



(b) Simulation graphs of E1 and pyruvate dilutions. Concentration of NADH is plotted against time (each window equals 300 s). In the top left corner of each graph the largest value of the y-axis is indicated, at the top center the amount of E1 in the complex is indicated. The arrow marks the direction of increasing pyruvate concentrations from 50 to 4000 nmol/L in seven intervals.



(c) Zoom in of (b) whereas pyruvate range was tested from 200 - 4000 nmol/L. In addition total hPDC amount was varied as depicted. The simulation data show a hump as maximum.

Figure 5.12: Investigation of hPDC subcomponents and pyruvate variation for hPDC activity via an *in vitro* model.

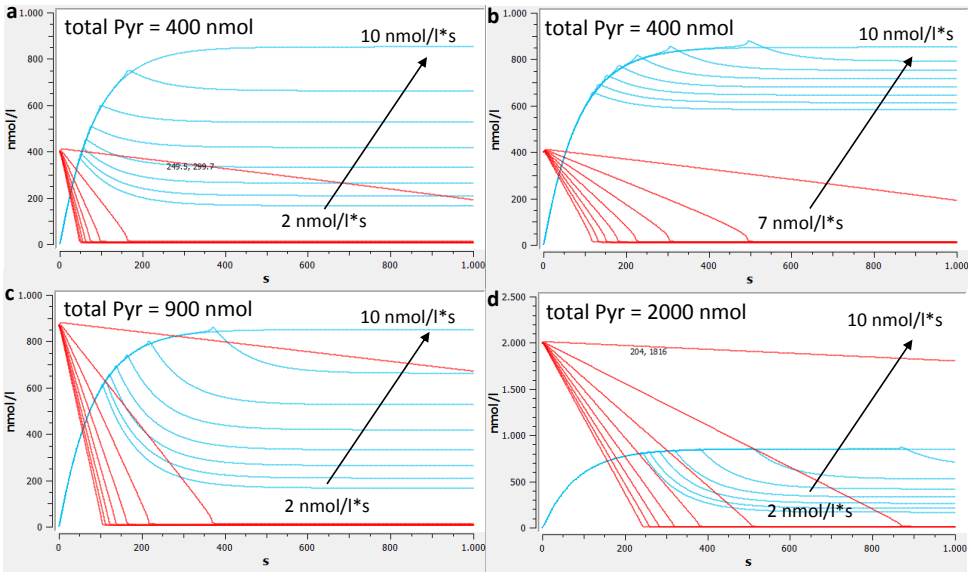


Figure 5.13: hPDC activity is simulated at different pyruvate feeding rates and initial pyruvate conditions. Red = pyruvate concentration, blue = NADH concentration. The initial start concentration of pyruvate is indicted in the left upper corner. The increase of pyruvate feeding rates (from 2 to 10 nmol L^{-1} s^{-1}) are indicated in the graphs.

the low pyruvate curves intersected with already dropping high pyruvate NADH curves (see Figure 5.12c).

In addition, a pyruvate feed rate was added to the batch model to allow for a theoretical investigation of possible sustained oscillation. The nominal value of the pyruvate feed rate so far was 0 s^{-1}. To investigate the continuous supply of pyruvate, a broad range of initial pyruvate concentration and range of the feed reaction was tested. If the feed rate was high in comparison to the initial pyruvate concentration, an ongoing NADH formation was seen, and no NADH decay occurred for the tested time span of 1000 seconds (see Figure 5.13). However it can be seen clearly, that as soon as pyruvate is consumed, the NADH formation is stopped and the decay of NADH takes place. Subsequent simulation of different feed rates at different initial pyruvate concentration showed no indication for sustained oscillation in the NADH formation, but for some conditions a hump before complete pyruvate depletion was seen (see Figure 5.13 b). However, no conditions to enable sustained oscillatory behaviour for the *in vitro* model could be identified.

5.5.1 Comparison of simulated vs. experimental data

The model of Zeng et al. (2002) represents simulations in a continuous reaction mode. This mode was adjusted to a batch-wise mode by the implementation of an initial amount of pyruvate provided for the reaction, which is consumed over time. The batch-wise *in vitro* model shows that the more E1 is diluted in the complex, the smother the transition from NADH formation to NADH decay is. Whereas the NADH decay is seen experimentally and it is not clear if this behaviour corresponds to instable NAHD decaying over time, or if this behaviour is caused by another effect. The model was adjusted to represent the experimental data, including a low NADH decay rate, which is visible as soon as pyruvate is depleted completely. However, the higher E1 amount in the complex causes a kind of transient oscillation, a single hump at the maximal NADH concentration. This represents the usual E1 amount in the complex well. Here, transient oscillation and global maximum fall together experimentally. For E1 dilutions the experimental data is not matching the model. The simulation shows that the transient oscillation appears with the global maximum (pyruvate depletion). Experimentally, this cannot be confirmed. This might be due to the simple model set up. No additional cofactor or accumulating product like acetyl-CoA or accumulating CO2 is simulated. The model in its current form might be not sufficient to display the experimental behavior with overlapping effects and limiting factors.
Furthermore, the model suggest to enhance the E1 amount and thus reverse its kinetic behavior. Here obviously a constraint is missing regarding the hPDC assembly. More E1 cannot interact with the complex because all E1 binding domains are occupied after exceeding the E2-E3BP capacity. Thus theoretically a reversed kinetic behavior would follow, but cannot happen experimentally. The usual hPDC assembly already displays E1 in excess and matches the simulation of E1 before the reversed kinetics, here an upper limit for E1 must be set *in silico*.

For the experimentally tested E1 dilutions, an intersection of the activity assay curves at fixed pyruvate concentrations are shown. The higher the E1 amount, the steeper the initial rate and the earlier the NADH decay starts, and thus intersects with the low E1 activity curves. This nonlinear behavior correlates to *in vitro* modeling by Zeng et al. (2002), the lower E1 amount activity intersects with the higher E1 amount pyruvate decarboxylation curves. Whereas also a steeper slope is visible for higher E1 amount. Further, for E2-E3BP and E3 this phenomenon was not detected, neither experimentally nor for the theoretical investigation by Zeng et al. (2002). It is known that E1 is the rate limiting step in the hPDC reaction (Patel et al., 2014), as well as the complex subcomponent at which enzyme regulation is performed by kinases. Thus, E1 represents a

regulation hub in hPDC, in which a change in reaction mechanism happens by causing transition states.

A difference concerning model and experiments are different theoretical assumptions about the complex composition. The theoretical ratio for the modeling according to Modak et al. (2002); Zeng et al. (2002) assumes 18 x E1 for the full complex whereas the *in vitro* stoichiomery is assumed with 30 x E1. The model was based on Modak et al. (2002) who assume the lower E1 stoichiometry. The actual stoichiometry of the hPDC is not resolved yet, different models propose different ratios. Patel et al. (2014) describe that 20-30 E1 heterotetramers and 6-12 E3 dimers are bound to the core. The core component ratio is also under debate, either the substitution model ($E2_{48}$ + $E3BP_{12}$ or $E2_{40}$ + $E3BP_{20}$, Prajapati et al., 2019) or the addition model ($E2_{60}$ + $E3BP_{12}$) are possible for core formation.

To enable sustained oscillation for the hPDC *in vitro* model a parameter was added to model the gradual consumption of pyruvate. Next to a starting pool concentration of pyruvate the pyruvate feed rate was investigated (see Figure 5.13). No appropriate rate could be identified at which oscillations occur, neither transient nor sustained. Regardless of the initial pyruvate concentration a continual trend is visible for the consumption rate. The higher the consumption rate the later and smoother the NADH decay (pyruvate total consumption) occurs. Whereas the decay state for NADH gets less and less. After reaching a high threshold feed rate no decay is visible anymore.

The dynamic behavior of the hPDC system showed differences for an assembly with different E1 ratio, not for different E2-E3BP dilutions. The transition of NADH formation to NADH decay is smooth for E1 dilutions, but peaky for E2 dilutions. This is also described by the modified batch-wise *in vitro* model. This model is built on a continuous model by Zeng et al. (2002). In this original model, there is no indication of nonlinear enzyme reaction caused by subcomponent E2-E3BP. However, for E2-E3BP and E3, the batch-wise simulation model and experimental data do not match. The experimental data show linear relationship and decreased signal if pyruvate is reduced, whereas the model suggests nonlinearity because there are intersections of pyruvate dilution curves over the decarboxylation progress. This is due to pyruvate depletion taking place later for increasing pyruvate concentration. The model cannot represent experimental data here, thus it is not valid for the simulation of E2-E3BP and E3. The continuous model of Zeng et al. (2002) is showing a linear relation of the enzyme kinetics when E2-E3BP or E3 are varied and thus match

the experimental behavior. This indicates that the modified batch *in vitro* model is missing essential reactions like limitations of cofactors or accumulation of by-products. Moreover, in the current *in vitro* model, NADH is directly released when pyruvate enters the reaction at E1, whereas actually NADH is formed and released at the end of the reaction cycle after all subcomponents performed their catalytic function (E1, E2 and E3, Patel et al., 2014), which could be taken into account.

Also, the total amount of hPDC is crucial for the development of nonlinearities. Decreasing the total amount of the complex experimentally about four times showed consistent linear kinetic behavior over pyruvate dilutions. Here the experimental signal resolution might be limiting. For decreased pyruvate concentration and lower protein amounts, the signal intensities of the activity assay becomes smaller. These two conditions combined might hinder an experimental detection of the already narrow amplitude of the transient oscillation seen for higher protein amount. Furthermore, one full period of the transient oscillation might also be shorter under named conditions, which would provoke a higher sampling rate for the detection of oscillations. The intervals of the measuring events may be too large to capture (transient) oscillation. For the model a scaling down of subcomponents and pyruvate still showed the nonlinear NADH kinetics because resolution of the curves is not limiting.

5.6 Conclusion and outlook

In this study, nonlinearities of the *in vitro* as well as *in silico* behaviour of hPDC activity in batch mode were investigated. It has been demonstrated that hPDC reaction can cause oscillatory behavior in metabolic flux depending on the external stimuli of pyruvate concentration. This is the first time oscillatory behaviour was detected experimentally during hPDC activity. A robotic system allowed for fast and high throughput processing of different reaction conditions, whereas the control over the modular assembly of hPDC out of its subcomponents *in vitro* enabled the protein complex composition as a parameter to be investigated.
The initial concentration of around 0.4 mM pyruvate is identified as the upper threshold limit to which transient oscillation occurs for the usual hPDC stoichiometry, this corresponds to physiological pyruvate concentrations. Additionally, this transient oscillation can be further modulated and the threshold pyruvate concentration increased by lowering the E1 level in the complex. Therefore, the absolute critical pyruvate concentration to which nonlinearities occur must be a function of E1, but other parameters as well, since there is

no linear relationship in dependence of E1. The parameter E1 is the main player in hPDC regulation as well as main factor for medical implications related to hPDC, therefore a holistic understanding and characterization of these nonlinearities could benefit to both points. The findings might contribute to practical considerations concerning hPDC deficiency, whereas E1 (more precisely E1α) is deficient in most cases (Brown et al., 1989). In particular the brain, where the energy metabolism is depended solely on the aerobic glucose oxidation, is affected. This results in abnormalities concerning the central nervous system with metabolic acidosis and neurological dysfunction. Further, implications are expected for bio-processes. A tight regulation of substrate supply should be present to be able to control reaction rates and yields in continuous cultures and avoid transition and lower steady state of hPDC.

The nonlinear kinetic hPDC behaviour dependent on parameter E1 was further confirmed by an *in vitro* model of hPDC. A model for a continuous hPDC process was available (Zeng et al., 2002) and adopted for the batch-wise *in vitro* reaction mode. When hPDC was assembled with the subcomponent E1 in stoichiometrically insufficient amount, the transient oscillations occurring during enzyme activity were elongated. No clear evidence on nonlinearity arising from subcomponent levels of E2-E3BP or E3 were found experimentally. Furthermore, the oscillation is sensitive to total protein content.
Moreover, a delay in enzyme activity was revealed for pyruvate concentrations in great excess which might be due to allosteric inhibition by the substrate. This phenomenon cannot be described by the ODE model yet.

In this study a minimal set of parameters were investigated batch-wise to obtain first experimental data on oscillatory hPDC behavior. This set of parameters might be expanded to cofactors like TPP, CoA and inhibitors like the kinase. To involve the whole regulation mechanism of hPDC also phosphatases might be employed to reactivate the enzyme. For investigating excess substrate concentration of pyruvate which lead to a delay of enzyme activity, TPP is an interesting parameter to screen. For the more physiological concentration of pyruvate, CoA or its product acetyl-CoA are worthwhile investigating *in vitro* since latter one acts inhibitory on E3 and thus hPDC (Selivanov et al., 1994; Zeng et al., 2002).
Each subcomponent of hPDC in the complex exhibits its influence on the reaction rates, in case it is the limiting factor. But only subcomponent E1 is causing transient oscillation, thus this subcomponent can be exploited in more detail by testing more dilutions. Here also complex independent data could help to understand E1 better, by performing E1 activity assays. The more critical

parameter, the substrate pyruvate, can be assessed in a context of continuous cultivation by also involving the metabolism up and downstream of hPDC. As described in the introduction, oscillations in the glycolysis are known already and therefore should be investigated together. Also as suggested by Menzel et al., 1998, in the pyruvate metabolism oscillations occur which are of interest for process optimization and control.

To have a closer look on the experimental challenges, following considerations might be involved:
To solve problems with unknown subcomponent ratio, it could be advisable to perform size exclusion chromatography of E2-E3BP and collect fractions with only full trimers to deplete non interacting E3BP units. To generally meet the theoretical trimer stoichiometry of 2:1 (E2:E3BP), enhance protein yield and activity, two copies of E2 gene could be cloned to one E3BP gene on the vector. To investigate the effect of NADH decaying after substrate depletion, kinetics for NADH behavior at reaction conditions could be recorded. Also, it could then be judged if the decay is caused by only the unstable NADH or if there are other phenomena underlying as well as a possible correction of the initial rate by referencing the NADH decay side reaction.

Finally, the digital twin, the batch-wise *in silico* model, needs improvement to reflect the whole hPDC reaction *in vitro* and deliver better information for an experimental screening. Here, a more realistic scenario can be followed in which all cofactors and substrates, also employed experimentally, could be included in the simulation setup as well.

Chapter 6

Structurally defined interrelations in hPDC

This final chapter handles the cross influence of hPDC core domains on the overall complex formation. Therefore, truncation mutants of the core were created by step-wise N-terminal truncation of the lipoyl and hinge domains of both core components. These were examined in regard to the kinetic protein to protein interaction of the sub-systems E1 to E2 and E3BP to E3 using BLI.

This chapter is mainly based on two Master Theses. Khansahib (2018) performed the molecular biology part and created all core variants and expressed them. In addition, he executed the DLS measurements and functionality testing of the core variants in an overall hPDC measurement. Finally he investigated binding kinetics of immobilized E1 or E3 subunit in interaction with core variants in solution. Based on this work, Steinbrecher (2019) helped to gain a more detailed understanding of the core variants by performing binding studies of immobilized core variants with E1 or E3 subunit in solution. He optimized BLI measurements and evaluated the obtained signals by manual fitting. Further, he set the experimental findings in relation to theoretical simulations.

6.1 Introduction

The architecture of hPDC is complex and leads to astonishing function. It is one of the largest protein assemblies known, to which it also owes its most important property, the efficient intermediate channeling (Guo et al., 2017). Firstly, this is efficient because of locally high concentration of multiple copies of each subcomponent. Secondly, they are connected to each other by protein to protein interaction and finally, they channel their intermediates - covalently

bound to flexible lipoic acid linker arms - to the catalytically active centers of the hPDC subunits (Patel et al., 2014).

A structure to function relationship needs to be established to resolve the connections between mechanisms and structural properties. Nevertheless, this might be a challenging task for multienzymatic complexes. Therefore, the interacting sub-systems in hPDC will be regarded individually.
Generally, a DNA template is translated to a 3D protein structure. Depending on its right fold, this structure then determines the functionality of the protein (Lee et al., 2007). As already seen in the chapter above (chapter 5 Nonlinear dynamics in hPDC), hPDC function is complex and its dynamic exploration showed nonlinearities occurring during hPDC activity. A structural investigation of interrelations in hPDC will support understanding hPDC in more detail.

hPDC core has a well-defined domain structure. The C-terminus contains the inner catalytic domain which forms the smallest functional unit of a trimer (Yu et al., 2008). This is followed by the BDs for E1 at E2 subunit and E3 at E3BP subunit (see scheme 6.1). Afterwards, different hinge regions follow which connect the N-terminal LDs to the core. These LDs bind the hPDC intermediates covalently and channel them to the distinct active centers of hPDC (Patel and Korotchkina, 2006).

The investigation of a possible cross-influence of the trimeric E2-E3BP core N-terminal domains on the distinct subcomponent binding is addressed in this chapter, to understand structurally how functional units are built up. Different parts of the flexible N-terminal lipoyl and hinge regions were truncated step-wise and their influence on hPDC subunit binding was analyzed. Truncation sites were selected to cover different, structurally clear distinguishable domains of the flexible linker arms and based on experimental data of Guo et al. (2017). The latter constructed a C-terminal truncated core variant (E2ΔC8-E3BPΔC7), highly active although the catalytic core was slightly minimized. Also, this variant is capable of forming agglomerates which are probably not the strict 60meric core assembly.
The affinity as well as interaction kinetics between N-terminal truncated core variants and surrounding hPDC subunits are determined through BLI measurements. This data will help to validate simulation based model of the fully assembled hPDC.

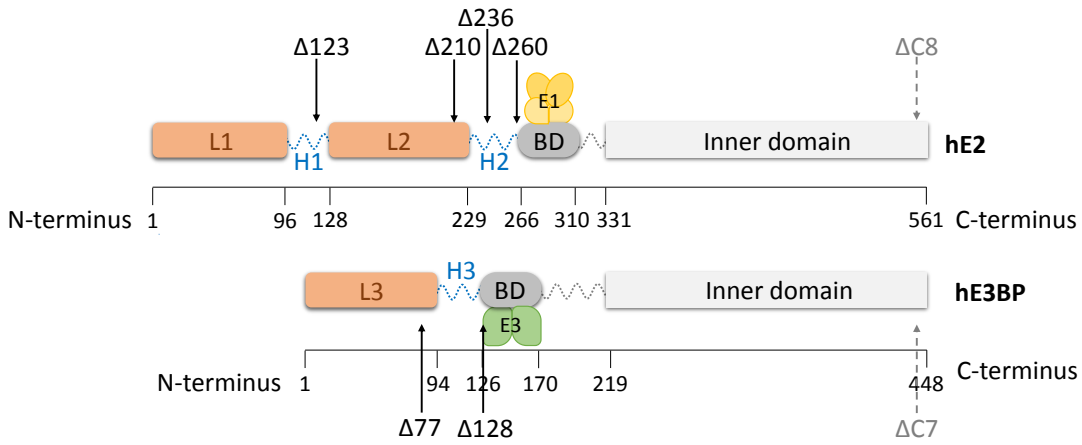


Figure 6.1: Schematic representation of hE2 and hE3BP structure with lipoyl- (orange), hinge- (blue), binding (grey)- and inner domains according to Hiromasa et al., 2004; Patel et al., 2009. Truncation sites are N-terminal and indicated by black arrow, additional C-terminal truncations are indicated by dashed arrows. Subunit E1 is represented in yellow, subunit E3 is represented in green.

Firstly, the core variant creation and characterization will be presented and discussed. Later, the establishment of a protocol for hPDC subcomplex evaluation via BLI will succeed in preliminary BLI studies, followed by the actual protein to protein interaction studies.

6.2 Domain deficient core variants generation and characterization

In this study N-terminal core truncations were targeted. Therefore, as a first step, the N-terminal His-tag of the starting core variant E2ΔC8/N6His-E3BPΔC7 was shifted successfully to the C-terminus of E3BPΔC7 (as can be seen in Appendix Figure 1.11). Then, N-terminal minimized variants of the core were created by In-Fusion® HD Cloning (Takara Bio USA, Inc.). Primers used for this purpose can be found in the Appendix (Table A.3). The truncation sites and their structural impact are shown in Figure 6.1 and Table 6.1. The truncation sites were roughly proposed by Uwe Jandt (unpublished), to investigate distinct N-terminal core domains until the BD. Dependent on the length of each domain, one or two step-wise truncations were performed.

Table 6.1: Overview of truncation sites on E2 and E3BP and indication on their structural impact. The deletion sites were selected to enable a detailed investigation of different N-terminal domains.

Enzyme	N-Terminal truncation	Structural impact
E2ΔC8	ΔN123	Deletion of L1 and part of H1
	ΔN210	Deletion of L1, H1 and L2
	ΔN236	Deletion of L1, L2 and part of H2
	ΔN260	Deletion of L1, L2 and H2
E3BPΔC7	ΔN77	Deletion of L3
	ΔN128	Deletion of L3 and H3

6.2.1 Overexpression of minimized core variants [1]

After cloning the above mentioned mutants, following 15 combinations of hPDC core proteins were expressed and purified successfully (see Table 6.2). A detailed cloning procedure can be found in the Thesis of Khansahib, 2018. The untreated WT E2-E3BP was also involved in following investigations. Thus, in total 16 E2-E3BP variants were analyzed. For some configurations a WT hPDC core deficient of E2 was employed. Here the gene of E2 was removed from the coexpression plasmid, and the E3BP subcomponent could still be purified via its His-tag ($E3BP_{only}$).

The overexpression of all variants shown in Table 6.2 could be verified after IMAC purification of the His-tagged proteins. Some variants are shown exemplarily in SDS-PAGE (Figure 6.2). Thereby it was seen, the more truncated the flexible lipoyl arms are, the more precise the theoretical molecular weight (see Table 6.2) of the protein matched the protein size at the SDS-PAGE. This discrepancy of theoretical size to size at the SDS-PAGE is assumed to be due to interference of the long and flexible interdomain linkers. These are rich in proline and alanine residues which might swell and hinder movement through the mesh of the SDS-PAGE (Vijayakrishnan et al., 2010). They impede the movement of the protein in the gel and thus, its apparent size on the gel is larger than the theoretical value. In case of E2 a molecular weight of 58.8 kDa is expected, whereas a molecular weight close to 70 kDa was observed on the gel. Similarly, E3BP exhibits a molecular weight of 48.3 kDa, but the band at the SDS-PAGE appears close to 55 kDa.

[1] The experimental work was mainly performed by Khansahib, 2018 and Steinbrecher, 2019.

Table 6.2: Overview matrix of all E2 and E3BP combinations. For E2 five variants were examined, for E3BP three variants were prepared. The first row and column indicate the starting variant, the other rows and columns show the additional N-terminal truncation to the template protein. In total 15 different core variants were obtained, in addition WT hPDC core was also investigated. Shown molecular weight in the matrix refers to theoretical size of a trimer consisting of 2xE2 and 1xE3BP.

	E2ΔC8 (58.8 kDa)	ΔN123 (46.2 kDa)	ΔN210 (36.6 kDa)	ΔN236 (33.8 kDa)	ΔN260 (31.6 kDa)
E3BPΔC7 (48.3 kDa)	C-terminally deleted starting variant	140.7 kDa	121.5 kDa	115.9 kDa	111.5 kDa
ΔN77 (40.3 kDa)	157.9 kDa	132.7 kDa	113.5 kDa	107.9 kDa	103.5 kDa
ΔN128 (36.6 kDa)	154.2 kDa	129.0 kDa	109.8 kDa	104.2 kDa	99.8 kDa

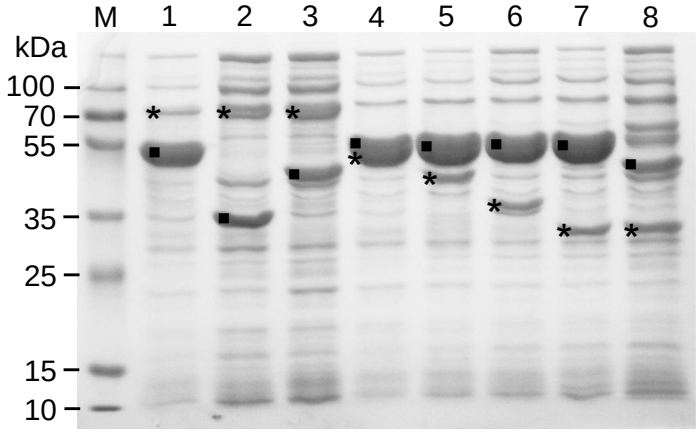
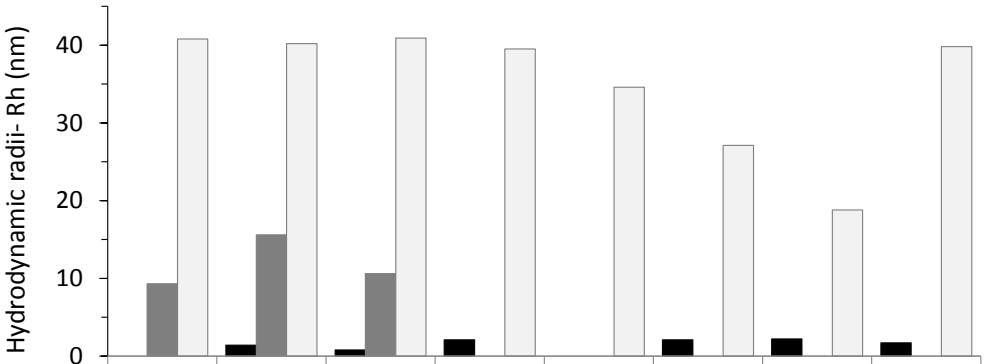


Figure 6.2: SDS-PAGE of some IMAC purified hPDC core variants. The E2 variants are indicated via an asterix (∗), the E3BP variants via a square (■). All shown proteins contain C-terminal truncations ($E2_{\Delta C8}$-$E3BP_{\Delta C7}$) and in addition: 1, $_{\Delta N0}$E2-$_{\Delta N0}$E3BP; 2, $_{\Delta N0}$E2-$_{\Delta N128}$E3BP; 3, $_{\Delta N0}$E2-$_{\Delta N77}$E3BP; 4, $_{\Delta N123}$E2-$_{\Delta N0}$E3BP; 5 $_{\Delta N210}$E2-$_{\Delta N0}$E3BP; 6 $_{\Delta N236}$E2-$_{\Delta N0}$E3BP; 7, $_{\Delta N260}$E2-$_{\Delta N0}$E3BP; $_{\Delta N260}$E2-$_{\Delta N77}$E3BP. A step wise reduction is size can be seen for the variants. Notably for E3BP truncated variants, the impure protein fraction is higher and the E2:E3BP ratio is more balanced. Figure adopted according to Ilhan *et.al*, submitted to PLOS ONE (2020).

For each truncation a decrease in size is clearly visible and corresponds better to the theoretical size. However, another effect is visible. After IMAC purification the proteins were not highly pure, but still contained impurities (see Figure 6.2). These impurities are common among all variants. Nevertheless, the E3BP truncated variants indicate a higher degree of contaminants than the other core variants. Wang et al. (2014) describes host cell protein (HCP) contamination as a matter of processing or raw materials, which then impact the stability and quality of the recombinant proteins. HCP are generated during expression of the protein of interest in the cells and differ strongly in type and quantity dependent on the expression host and purification process (Wang et al., 2014). A possible impact on protein stability is reported, like protein to protein interaction induced instability or aggregation formation. In this context, the higher level of impurities of the E3BP truncants may hint on a destabilizing effect on the hPDC core for modifications on the E3BP subunit. This might also explain why the ratio of E2 to E3BP is more balanced when E3BP is modified (ass can be seen in Figure 6.2 or Appendix Table 1.7. The truncated E3BP variants are likely to be unstable and thus probably insoluble, thus this fraction is not co-purified, especially lanes with truncations on E3BP show a more balanced E2:E3BP ratio. Vijayakrishnan et al. (2010) report that E3BP is already destabilizing the E2-E3BP complex since the interactions with E3BP are weaker than with only E2. The purification process needs to be optimized in this case. More gentle processing steps or additional polishing steps might ameliorate the incomplete purification procedure. Steinbrecher (2019) also reports that the balance of E2 to E3BP ratio is dependent strongly on the protein batch which support the assumption of processing influence on the proteins.

The protein yields and purities of the variants are also remarkable (see Appendix Table 1.7). If the WT is set as standard, the single (only E2 or E3BP) truncated variants exhibit around half the yield of protein of interests from the same culture volume, and purity is decreasing. For the double truncated variants (E2 and E3BP) target protein yields are likewise or higher than WT yields, whereas purity decreased further. For some of these variants, after IMAC purification, the impurities accounted to around 50 % of the total protein content. On one hand the impurities are increasing which hints to more unstable proteins in need of accompanying proteins. On the other hand yields increase which might be due to better solubility and better expression efficiency for the smaller proteins. For the single truncated variants this is in reverse. If there is an impact of the remaining impurities on the specific protein interactions within hPDC at all and to which extend was not studied in this work. However, the obtained interaction



	WT	E3BPΔC7 _E2ΔC8	E3BPΔC7 _ΔN77	E3BPΔC7 _ΔN128	E2ΔC8 _ΔN123	E2ΔC8 _ΔN210	E2ΔC8 _ΔN236	E2ΔC8 _ΔN260
Population 1		1,4	0,8	2,1		2,1	2,2	1,7
Population 2	9,3	15,6	10,6					
Population 3	40,8	40,2	40,9	39,5	34,6	27,1	18,8	39,8

Figure 6.3: Evaluated DLS data of WT E2-E3BP and different E2-E3BP truncation variants, according to Khansahib (2018). The variants entitled with E3BP truncations contain additionally E2ΔC8, the variants entitled with E2 truncations contain additionally E3BPΔC7. The impact of E3BP truncations is minor, whereas E2 truncations lead to smaller hydrodynamic radius of core agglomerates. Furthermore, population 2 is missing for some variants.

constants are all similar to each other (as presented in the following sections), and no difference can be seen between core variants with higher purity compared to variants with lower purity.

6.2.2 Examination of size distributions via DLS [2]

To investigate the size distribution of the core variants in solution, Dynamic Light Scattering (DLS) measurements were performed. This tool enables the study of macromolecular diffusion behavior in solution, and allows for the calculation of the hydrodynamic radius. The size and shape of the particle influences the latter (Stetefeld et al., 2016).

The measurements were performed at the core facility of sample preparation and characterization center at the European Molecular Biology Laboratory in Hamburg. A Wyatt Nanostar system was employed and for each sample an

[2]The experimental work was mainly performed by Khansahib, 2018.

average of 30 measurements were acquired. The core samples were used in a concentration of 1 mg/mL and filtered short before the measurement by 0.22 µm Ultrafree® MC GV 0.5mL Centrifugal Filter Unit (Millipore). The results for the WT, the minimized core (used as base, E2ΔC8/E3BPΔC7) and additional N-terminal truncation on either E2 or E3BP are presented in Figure 6.3.

In general, several populations were detected, which were not consistent along the core variants, see Figure 6.3. The smallest population (Population 1) with hydrodynamic radius (R_h) around 0.5 to 2.5 nm might indicate monomers or trimers (radius of gyration R_g equals 0.39 to 1.94 nm, calculated by using the equation shown in the next paragraph). The second population, only appearing for few samples, might be an agglomerate or the 60mer core. When calculating the R_h (by using R_g from 60mer core simulation of Hezaveh et al. (2018)), the resulting population should be around $R_h \sim 7.5$ nm. This is indeed roughly the case for the WT E2-E3BP measurement. The biggest population (population 3) might represent a big agglomerate, larger than the 60meric core. Guo et al. (2017) also detected a population with $R_h < 30$ nm which they could assign to big agglomerates not able to be assessed through size exclusion chromatography. They concluded that this big, not strictly ordered agglomerate is capable of hPDC core function like the ordered 60meric core assembly, although the mutant is C-terminal truncated.
In this work, the R_h of single truncated E2-E3BP variants were collected as can be seen in Figure 6.3. According to Guo et al. (2017) this can be related to radius of gyration of a spherical protein model by $R_g \sim 0.775 * R_h$. Therefore the collected data might be of interest for hPDC modelling. Hezaveh et al. (2017) reported values for radius of gyration for monomer simulations of WT E2 or E3BP between 1.63 to 1.79 nm. In case of trimer simulations ($E2_2$-$E3BP_1$) $R_g = 2.50 \pm 0.015$ nm was obtained. Also, for the 60meric core and in dependence of different stoichiometric models, R_g values between 9.65 to 9.67 nm were reported. Furthermore, they showed an effect of decreasing R_g over time for C-terminal double truncated variant.
The clear identification of the second population was problematic. Single acquisitions of one sample showed different populations and distributions. Although some samples were measured again after resting for two hours on ice, the resulting number of populations and distributions varied again. The results shown in Figure 6.3 represent only a part of the collected data. The problem arising because of the employed polydisperse solution (different subcomponents interacting in different kind of populations) hinders the detection of distinct populations if their size range is close to each other. Although the reducing R_h shows that the proteins were reduced in size, and approach theoretically

calculated values as reported by Hezaveh et al. (2017), (with exception of ΔN260-E2-ΔC8/E3BP-ΔC7), this can be only understood as a qualitative, but not quantitative measurement.
The E3BP truncations did not show a decrease in size, this is probably due to the relatively small parts deleted. As the simulations of Hezaveh et al. (2018) show, BD on E3BP is located near the N-terminal end. Whereas E2 has its BD close to the inner domain, thus larger parts of the protein were removed here which are detectable via DLS.

DLS suffers some drawbacks considering its resolution, for instance it cannot distinguish particles closely related, like monomer and dimer (Stetefeld et al., 2016) and will give a single wide distribution. This is clearly seen in the raw data of single acquisitions of the DLS measurements. Often, a single broad peak, sometimes biased or with a shoulder, from around 10 to 100 nm is seen. In some few acquisitions this is broken down to two single, very narrow peaks. Also, no monomer could be resolved from a trimer, instead a range covering both is seen. DLS is also not well suited to determine molecular weights. High molecular weight aggregates thereby disturb measurements because they scatter smaller particles and are a rate liming factor for biophysical studies. Furthermore, accurate distributions can be only obtained for monodisperse solutions. In case of polydisperse solutions (coexpression of E2 and E3BP and different forms of aggregates), the larger aggregates bias the results since they scatter more light. The detection of the smaller particles, even if they are present in a larger quantity, is challenging (Szatanek et al., 2017).
Therefore, with this DLS measurement it could be shown that the agglomerate size of the truncated variants is decreasing, thus the overall agglomerate size is getting smaller. Each truncation variant minimized by more than E2-ΔC8/ΔN77-E3BP-ΔC7 does not show a reliable second population any more which is related to a 60mer core. This could hint at the ordered core structure not being assembled any more. Furthermore, the decreasing size of the large unordered agglomerate could shield the smaller 60mer core, thus only a wide peak size distribution appear which cannot resolve the two single populations.

6.2.3 Activity testing of minimized core variants[3]

The modified hPDC cores were assembled with their interaction partners E1 and E3 to test their catalytic activity in an overall hPDC assay. As Guo et al. (2017) hypothesized, a large agglomeration formed which might not necessarily

[3]The experimental work was performed by Khansahib, 2018.

be the structured 60 meric core, leads to functionality. For the WT core an overall activity was detected and calculated to be 46 mU/mg. For all produced 15 core variants with His-tag shifted to the C-terminus compared to the WT, no activity was detected, their performance equaled the NCs (data can be found in Khansahib (2018)). An impairment of function was expected, since the flexible linker arms in charge for metabolic channeling were truncated. In this study a total loss of activity resulted for all N-terminal truncated lipoyl domains of the hPDC core components. As results of the interaction studies show in the next section, the complex formation is still taking place, thus the loss of function must be addressed to the loss of the lipoyl domain, maybe in combination with C-terminal His-tag.

Lawson et al. (1991) investigated E3BP and lipoyl moieties in the eukaryotic *Saccharomyces cerevisiae*. Thereby, they claimed that a truncation of 80 residues of the lipoyl domain at E3BP is impairing the overall PDC activity only little (still 60-70% of WT activity). E3BP is not having any catalytic, but structural function, it anchors E3 to the complex. Only the truncation of the E3-BD impaired PDC function. Therefore, in this study the core variants containing N-terminal truncations only at subunit E3BP were expected to exhibit activity, since the BDs were not harmed. E2 should still be able to perform its catalytic activity. An effect of C- and N-terminal truncation plus the His-tag shifted to the C-terminus of the core might lead to a loss of function.
Gopalakrishnan et al. (1989) investigated the role of E3BP in the mammalian PDC. They claim that a loss of E3BP-LD leads to a loss of function, which is in accordance with the results collected in this work. Thereby, it was shown that a progressing cleavage of E3BP-LD is not in a linear relationship to its loss of function, the cleavage proceeds faster Gopalakrishnan et al., 1989. This hints on the involvement of E3BP as a key partner for complex functionality.

For the only E2 modified mutants it was of interest to see whether there is a difference in the loss of only outer, or both, outer and inner lipoyl domain. Since for E2 the lipoyl arms are essential and bind the intermediates covalently (Patel and Korotchkina, 2006), it is clear why the CoA and NAD^+ linked oxidation of pyruvate is not possible anymore. Same accounts for the combined, E2 and E3BP modified variants.
A study of Miles et al., 1988 investigated the importance of the flexible hinge regions between LD and BD in *E.coli*. A short truncation of the linkers did not show any grave effects. For a linker with only 13 or less residues left the active site coupling was impaired and the reductive acetylation of the lipoyl domains hindered. This very short variants could still assemble to the complex,

but showed only around 50 % of WT activity. This shows the necessity of the appropriate linker lengths for the full complex activity. Nevertheless, since the LDs were still present, a residual activity was detected.

As suggested by DLS measurements, no formation of the 60meric core, but instead a larger agglomerate formation is happening for the N-terminal truncated core variants. This could be a reason to the loss of hPDC function, although it was claimed that activity was also exhibited by a large and unordered core assembly Guo et al., 2017.

Noteworthy is the complete loss of function for the C-terminal His-tagged E2-ΔC8/E3BP-ΔC7. Guo et al. (2017) employed this variant with an N-terminal His-tag successfully in activity assays and could reach 69 % of the WT activity. Thus, the loss of function here could be reasoned to the shift of the His-tag to the inner catalytic C-terminus. An impairment of core formation, which is mediated at the C-terminus, would be thinkable. This can be refuted by the results seen on the SDS-PAGE (see Figure 6.2). If there would be no interaction between the core components any more, also the co-purification of E2 subunit with the His-tagged E3BP subunit would fail. Obviously, the C-terminal His-tag nevertheless is influencing the catalytic activity of the core in a so far unknown fashion.
A His-tag in general is assumed to have no or little influence on the native protein structure, and so on the function of proteins (Carson et al., 2007). Even the number of crystal structures elucidated with His-tag is increasing. Nevertheless, there could be drawbacks like altered solubility or increase in aggregation of the purified protein. Thus, for every case a possible impact of the N- or C-terminal His-tag need do be evaluated, and in case of interference this tag might be cleaved to regain the full structure and activity. Perron-Savard et al. (2005) report for PhoP response regulator in *Salmonella enterica* an effect of a C-terminal His-tag on protein function, because the conformation and oligomerization state of the protein is influenced. To my knowledge, in case of PDC purification via His-tag in different studies, latter was always located at the N-terminus, so the influence of C-Terminal His-tag was not studied so far.

6.3 Establishment of BLI method for protein binding studies[4]

Protein to protein interaction studies were performed by BLI measurements. As this was the first time hPDC was studied via BLI, some preliminary experiments were performed to establish a protocol for screening of the mutants in binding experiments. The preliminary experiments were established with the WT subcomponents, and this protocol was applied for all core mutants to ensure comparability.

6.3.1 Preliminary BLI tests regarding hPDC core as analyte

Initially, the core was treated as analyte, therefore surrounding subunit E3 was used as ligand and its binding to the biosensor fixed in regard of concentration and incubation time. Subsequently, the parameters for the core association were figured out.

The hPDC subcomponents were immobilized on Anti-Penta-His antibody biosensors (HIS1K) if not stated otherwise via the His-tag of the proteins. Firstly, different concentrations and time steps were tested to find optimal assay conditions. Since the analyte interacting with the ligand also contains a His-tag, loading of the biosensor was performed longer than usually recommended (Tobias and Kumaraswamy, 2013) to saturate the biosensor tip. So, no interaction of the analyte with the biosensor but only the ligand should be enabled. In addition, reference sensors were employed in parallel by utilizing a protein not interacting to PDC (GFPuv), to reference baseline drift and unspecific protein binding.
Khansahib (2018) investigated E3 as ligand at concentrations of 25 and 50 µg/mL, and with varying binding time over 100, 200 and 400 seconds, respectively, to figure out the best immobilization conditions (data not shown). For the loading concentration 50 µg/mL was set as standard, although also 25 µg/mL showed good signal intensity. The response of the different time spans were similar, thus already 100 seconds are sufficient as binding time. However, longer binding times resulted in higher signal intensities which allowed to detect a dissociation step. Therefore, longer binding times were aimed at.
Next, the analyte (hPDC core) was investigated. As can be seen in Table 6.3, the quality of the fit, no matter if local or global, increases with lower analyte concentration. The analyte was diluted in a 1:2 series. Except of the highest

[4]The experimental work and evaluation was mainly performed by Khansahib, 2018 and Steinbrecher, 2019.

Table 6.3: Investigation of analyte concentrations suitable for BLI measurements. E3 was immobilized to the biosensor at a concentration of 50 mg/mL. Subsequently, different concentrations of WT E2-E3BP were associated to the ligand. To achieve a robust fit, a set of independent runs utilized and either fitted singly (local) or globally. The accuracy of the curve fit can be judged by R^2. Here the global fit shows better curve fitting accuracy. Table adopted according to Khansahib, 2018.

Assoc. sample conc. (nM)	R^2-Global fit	R^2-Local fit	k_{obs} (s^{-1})
294	0.418	0.9356	0.144
147	0.5872	0.9856	0.089
73.5	0.7747	0.9934	0.044
36.8	0.8103	0.9944	0.022
18.4	0.9599	0.9920	0.005

analyte concentration of 294 nM E2-E3BP, all R^2-Local fits are in an acceptable range (according to vendor R^2 larger 0.95), and the resulting k_{obs} showed a nearly perfect linear relationship to analyte concentration. The global analysis of all biosensors (thus concentrations) is thought to provide a more robust value, but is also prone to larger error because of individual R_{max} value of each biosensor. Therefore, the local measurement usually showed better R^2 values.
Finally, to achieve a good trade off between quality and intensity of the signal, parameters were set as shown in method section 3.7. As reference sample GFPuv was selected, because it contains a His-tag. Additionally, there is no known interaction of GFPuv with hPDC subcomponents and because of its small size which does not exceed hPDC subcomponent sizes. The relative long duration time for dissociation was used because of the low dissociation rate of the core variants from their ligands. As positive control to which the mutants should be set into relation, C-His-E3BPΔC7-E2ΔC8 was used.

For the reference sample of GFPuv it was seen that after a baseline like association step of GFPuv to bound E3, a negative slope for the dissociation phase followed. Since no analyte was associated, this dissociation must result from the immobilized ligand which aggravates the evaluation. If this reference sample is now subtracted from the actual samples, a positive curve results for the dissociation of E2-E3BP from E3. This problem could not be solved for E2-E3BP as analyte. Therefore, only local fit was applied and k_{obs} was used for evaluation purposes, since latter is a kinetic value not dependent of a dissociation step.
Furthermore, in the preliminary testing for the core as analyte, the system of E3

to E2-E3BP seems to be stable and does not dissociate, whereas E1 to E2-E3BP shows dissociation behavior over time.

In general, for this orientation (core as analyte) a strong dissociation during each baseline step was visible. The only buffer sample, which basically represents a prolonged baseline step, showed a steady dissociation of the ligand over time. This dissociation was stronger than the analyte dissociation. For E3 as ligand, the association with the core led to a stable complex which hardly dissociated, this was also the case for the sample controls. For E1 as ligand, during the dissociation phase the first few hundred seconds showed a plateau like behavior, afterwards a dissociation took place. The negative control samples showed a dissociation behavior throughout. This described kinetic behavior indicates interference and artefacts of the binding signals. One reason might be the multimeric nature of E1 and E3. Therefore, each ligand contains two His-taggs in their native states. Thus, one ligand could theoretically bind to two positions on the biosensor. This might cause problems like very tight binding as it is seen for E3, or unstable binding as it is seen for E1, dependent of the distances of the His-taggs from each other and the ligand size. A HIS1K biosensor provides approximately 10^9 capture sites, distributed over a biosensor tip diameter of 0.6 mm (information obtained from a request to Mol-FortéBio Support). This calculates theoretical to 5.96 nm spacing between each binding site. Dr. Uwe Jandt calculated theoretical spacing from one His-tagged end of the dimeric protein to the other His-tagged end (data not published). This is 10.86 nm for E3 dimer (from Ala-Ala) and 10.27 nm for E1 heterotetramer (from Phe to Phe). The sequence of E3 thereby contains a thrombin site which acts as a small flexible spacer and allows E3 more degrees of freedom. This might lead to a bridging of E3 dimer over two binding sites and indeed, double binding to the sensor, which is less probable for E1.
Furthermore, as can be seen nicely in the trimer simulations of Hezaveh et al., 2018, E1 and E3 subunits are quite small in comparison to the core with long flexible linker arms, and are covered by them. Maybe the accessibility of the large core component to the immobilized E1 or E3 subunit is unfavorable.

6.3.2 Preliminary BLI tests regarding hPDC core as ligand

To be able to test the specific binding of E1 or E3 to the core, overlaying effects caused by His-tag interaction of the subcomponents to the biosensors should be avoided.
A preliminary experiment showed higher binding rate of E1 to a low E2-E3BP loaded biosensor than a high loaded E2-E3BP biosensor (data can be found in

Steinbrecher, 2019). Since there are less E2-E3BP binding sites on the lower loaded biosensor, the resulting higher association rate must result from E1 interaction with the HIS1k biosensor. Therefore, this effect cannot be neglected. Consequently, a TEV-protease His-tag cleavage site was introduced to E1 and E3 genes to specifically cut away the His-tag after protein purification. Subsequently, the His-tag cleavage success was determined by binding experiments of the treated E1 and E3 to the empty biosensor. Thereby, only a partial elimination of the His-tag was revealed. For treated E3 the binding signal was reduced by 77 % in comparison to untreated E3. This shows that the removal of the His-tag is not complete. For E1, efficiency is even worse, the signal was reduced by 31 % (data can be found in Steinbrecher, 2019). This means, although the His-tag should have been removed, there is still a relative large portion of E1 and E3 proteins containing a His-tag and therefore interacting with the biosensor. This will cause a mixed association signal of the surrounding subunits to the immobilized core and anti-His-tag biosensor. The need of a proper biosensor binding strategy is important to minimize the analyte interaction with the biosensor. In case of E1, in which the His-tag removal was quite inefficient, the $E3BP_{only}$ core was employed as reference senor to filter out possible unspecific interaction with the biosensor.

The binding of $E3BP_{only}$ core to the empty biosensor was investigated over a concentration range from 25 to 600 nM. The lower concentrations showed a linear correlation to the signal height, but also for the highest concentrations no saturation in signal height was detected (data can be found in Steinbrecher, 2019). Since low binding concentration is advised by the vendor, as long as the following association signal is strong enough, 25 nM was set as core standard to which the surrounding subcomponents of hPDC are associated to. Tobias and Kumaraswamy (2013) report binding artefacts of high ligand loading density like steric hindrance, crowding and aggregation on the surface. Also weak unspecific interactions with the analyte, walking of analyte or rebinding of analyte may occur because of over-saturated biosensor.
Accordingly, the analyte was employed in a linear concentration from two to 50 nM, usually divided in three steps via a titration series to test the more robust association concentrations.

In this preliminary experiment another effect was revealed. WT and $E3BP_{only}$ were the variants which were not C-terminal truncated. After they were bound to the biosensor, a rather strong "dissociation" was detected for the subsequent

baseline step. The C-terminal truncated variants hardly showed dissociation (data can be found in Steinbrecher, 2019). This hints on a modified aggregation behavior of the C-terminal modified variants.

6.4 E1 to E2 binding system

In the following, the E1 to E2 binding system is subject of consideration. Thereby both orientations, core immobilized as ligand as well as the core in free solution employed as analyte will be investigated by BLI measurements. The significance values shown in this section were calculated by using a two-sided t-test for independent samples, whereas the small number of samples (E2 = 5 and E3BP = 3) must be taken into account.

6.4.1 E1 to E2 subcomplex, hPDC core variants as analyte

Evaluating the kinetic parameters for this system was troublesome. After referencing the GFPuv sample reference from hPDC interaction curves, positive slopes were obtained for the dissociation rates. This makes an evaluation for the equilibrium constant impossible. Despite a long ligand loading time to cover the biosensor binding sites completely, the subsequent association step might be an overlay of protein association to the ligand, but additionally also His-tagged analyte binding to the biosensor. This mixed binding obviously stabilizes the formed association, so hardly any dissociation is occurring. Therefore, the results in this parts will be presented by utilizing the kinetic apparent binding rate (k_{obs}). This is used because k_a and k_d cannot be determined independently, so k_{obs} is used which combines them, as explained in equation **??**.

Firstly, subcomponent E1 was immobilized at the biosensor tip and core variants in free solution were able to interact with it. For this interaction system of E1 to the E1-BD at E2 a truncation on E3BP is expected to show a minor effect. However, as shown in Figure 6.4, E3BP truncations are indeed effecting the E1 to E2 binding in a cooperative manner with E2 truncations.
The variants containing ΔN0-E3BP-ΔC7 truncations exhibit a significantly higher app. association rate than the variants containing ΔN77-E3BP-ΔC7 truncations ($p = 0.036$). So if the LD on E3BP is missing, the E1 to E2 binding system seems to be slightly destabilized. The ΔN128-E3BP-ΔC7 truncation series shows no significant difference to the other E3BP truncations, but a trend towards increased app. association rates in combination with LD deficient E2s. Along the E2 truncation series, no significant differences were measured

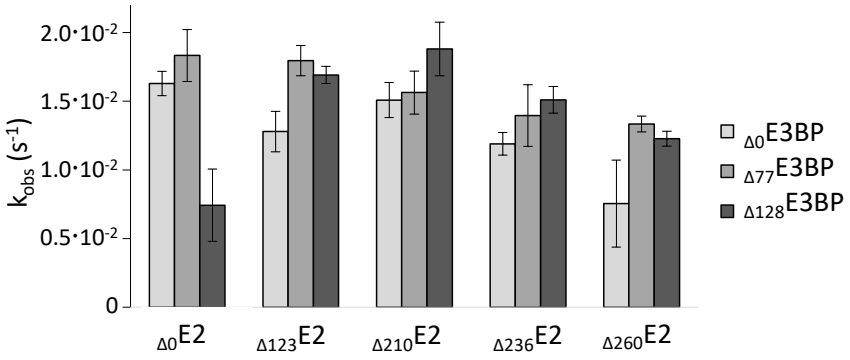


Figure 6.4: BLI investigation of immobilized E1 (ligand) and different core variants (analyte) as interaction partner in free solution. Because of low dissociation behavior of the formed complex in this set up, the apparent association rate is shown over all combinations of the E2-E3BP variants. In addition to the shown N-terminal truncations, all E2 subcomponents miss 8 residues C-terminally, all E3BP subcomponents miss 7 residues C-terminally. Figure adopted according to Khansahib (2018).

concerning the app. association rate. Nevertheless, a trend is visible in which the app. association rate is decreasing the closer the truncations approach the actual E1-BD. The E1 to E2 binding system is influenced marginally by E2 missing both LDs (app. association rates increase) and to the reverse for missing H2 (app. association rate decrease again). The maximum divergence within all app. association rates in this set up is a factor of 2.5.

The subcomponent E2 missing all components just before the E1-BD (L1, H1, L2 and complete H2) seems to be destabilized. For this most truncated E2 variant, a stabilizing effect of also truncated E3BP is visible. This might be due the structure of E2-E3BP. The E1-BD of E2 is in close proximity to the inner domain (Hezaveh et al., 2018), a large portion of the 3D structure of the protein is already missing for the ΔN260-E2-ΔC8 variant. The structure of E3BP is different, the hinge region between inner domain and E3-BD is longer and so the E3-BD is relatively far away from the inner domain. Aforementioned observation lead to the conclusion that most of the 3D structure remains. This might play a role for the stabilizing influence of the truncated, nevertheless long flexible arm of E3BP accommodating the strongly truncated and uncovered E2 subunit.

The improper curve fitting and low replicate number (n=3) does not allow to gain full information from the data. Therefore, a manual curve fitting which respects the multilayered binding effect is aimed to and addressed in the next section.

6.4.2 E1 to E2 subcomplex, hPDC core var. as ligand

For this BLI set up the elimination of the His-tag at the analyte preceded. The His-tag removal for E1 was only successful for a small portion of the E1 protein, thus overlaying effects for the interaction of E1 with the core variants and ligand binding site was expected. In some cases the E1 binding to the biosensor exhibited a higher signal than the E1 interaction to the immobilized core subcomponents. Steinbrecher, 2019 could show that the predominant interaction of E1 with the biosensor is correlated to the ratio of E2 to E2BP. The E2:E3BP ratio was calculated from the SDS-PAGE of the purified E2-E3BP proteins. When E2:E3BP ratio was around 0.5 or lower, the interaction of E1 to the unloaded biosensor was higher than E1 association to the E2-E3BP loaded biosensor. Only if the E2:E3BP ratio was more balanced, the E1 to core association was dominant over E1 interacting with the biosensor. This trend can be seen in Figure 6.5. For the core variants in which E3BP was not truncated N-terminal, the E2:E3BP ratio was the smallest (see Figure 6.2 and Appendix Table 1.7) which also accounts for the association rate calculated via 1:1 model. Replicating this experiment with fresh core preparations with more balanced E2:E3BP ratio showed slightly higher association rates (data can be found in Steinbrecher, 2019).

Obviously, the 1:1 model cannot reflect the real binding behavior of E1 to the core because of the additional interaction of E1 to the biosensor. Therefore, a 2:1 model, which represent two heterogeneous ligand binding sites for one analyte is more appropriate. In this case one ligand reflects the sensor surface, the other ligand reflects the binding domain at the immobilized core. Since it is unknown to which extend which binding mode enters the overall interaction, the two different association and dissociation rates must be estimated. The known association and dissociation rates of the E1 interaction to the empty biosensor were involved as initial condition for the parameter estimation.
For some conditions the experimental data could be fitted nicely and the two overlapping interactions were resolved, but not for all variants (data is shown in Steinbrecher, 2019). Because of the low dissociation behavior for both, hPDC interaction as well as E1 His-tag to biosenor ($k_d \sim 10^{-4}$ s^{-1}), the calculated equilibrium binding constant was quite erroneous and not all variants could be evaluated. The calculated equilibrium binding constants ranged from $1.1 \cdot 10^{-10}$ M (ΔN260-E2-ΔC8/Δ0-E3BP-ΔC7) to $5.3 \cdot 10^{-10}$ M (ΔN123-E2-ΔC8/Δ128-E3BP-

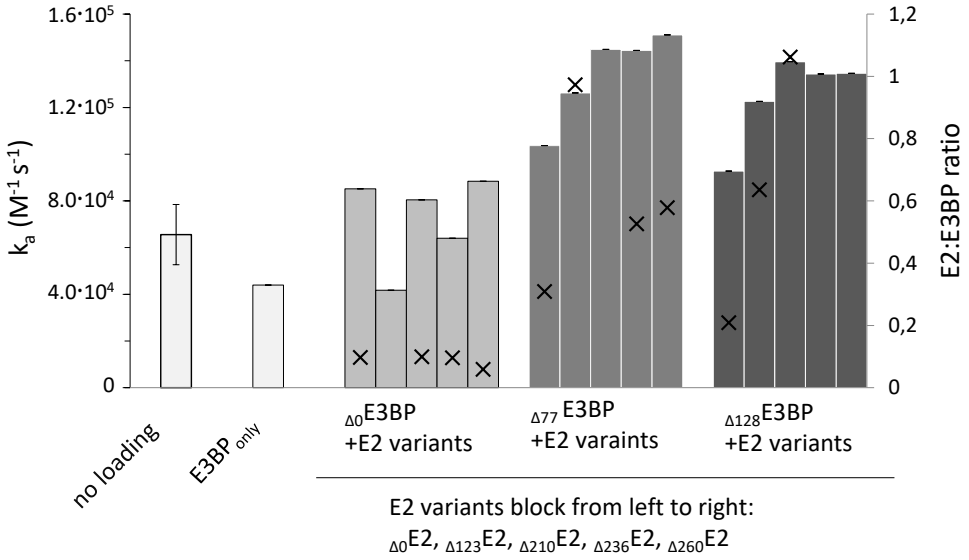


Figure 6.5: Association rate of E1 to the immobilized core variants. The calculations were based on a 1:1 binding model. The first (light grey) bar indicates E1 binding to the empty biosensor, thus "association" resulting from His-tag interaction to the anti-His biosensor. This was measured in a triplicate, which represents the error bar, all other cases represent a single measurement. The second (light grey) bar represents E1 association to only E3BP core. Further, all core variants follow as depicted in the graph. On the second y-axis to the right, the E2:E3BP ratio is indicated via a X, this ratio could not be determined for each variant because of overlapping bands on the SDS-PAGE. Graph adopted according to Steinbrecher (2019).

ΔC7). Consequently, a method was established to reference out the interfering signal experimentally.

The WT $E3BP_{only}$ protein was used as reference sensor, since the analyte E1 is theoretically not interacting to E3BP. The association rate of E1 to $E3BP_{only}$ is lower than E1 to the biosensor as can be seen in Figure 6.5, hence binding sites are occupied by $E3BP_{only}$ and no interaction is taking place. Subsequently this control was referenced from the collected signals, and a simple 1:1 binding model can be applied. Here, also not all variants could be evaluated, some others only in a single determination.
The rate and binding constants for the E1 to E2 sub-binding system, whereas the core is immobilized, are shown in Figure 6.6. The association rate constant is not differing over the E2 truncation series. However, a cooperativity of E3BP to the E1 to E2 sub-binding system is revealed. In general the ΔN77-E3BP-ΔC7 variants exhibit a slight decrease, whereas ΔN128-E3BP-ΔC7 variants a slight increase in k_a ($p = 0.002$). The dissociation rate constants were decreasing over E2 truncation series, until the biggest E2 truncation close to the E2-BD. The E2 variants missing the lipoyl domains but still owing a short portion of H2 show significantly reduced k_d over N-terminal fully intact E2 ($p < 0.02$), with maximum divergence by a factor of 4.6. Furthermore, for E3BP truncations no clear trend can be seen on k_d (here not all data was available). Finally, the k_d values exhibited the biggest influence on the K_D. Therefore, the K_D values were significantly reduced for the three largest E2 truncations ($p < 0.05$) in comparison to N-terminal intact E2. For E3BP no significant trend was revealed here. All variants show a high affinity in the same magnitude, in an area of less than 1 nM (0.22 to 0.71 nM).

6.4.3 Comparison of E1 to E2 binding system

The investigation of the E1 to E2 system binding is problematic with the set ups used for the BLI measurements. Firstly, the E2 ratio of the different E2-E3BP variants is low and fluctuating, which results in low and fluctuating BLI signals. Especially the dissociation was very low and could not be evaluated for all variants, thus all obtained values must be treated with caution. Furthermore, the remaining His-tag on TEV-protease treated E1 leads to overlapping binding effects. This challenge was addressed by appropriate reference sensors.
No matter which ligand and analyte orientation on the biosensor was chosen, these aforementioned interference effects occurred for both cases. However, the set ups were different. In case of E2-E3BP as analyte higher ligand (50 nM vs. 25 nM) and analyte concentrations (80 nM vs. two to 50 nM) in comparison to E2-E3BP as ligand were used. Furthermore, the experiments in which E2-E3BP

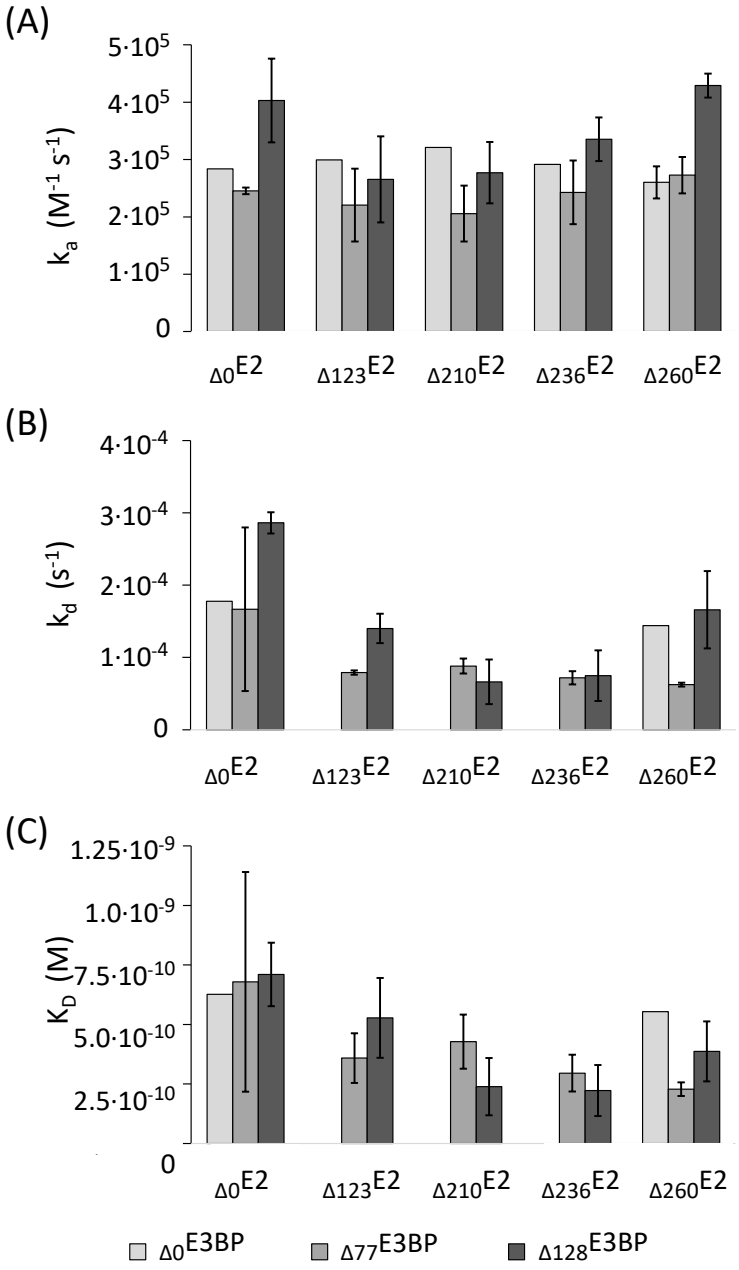


Figure 6.6: (A) association rate, (B) dissociation rate as well as (C) equilibrium binding constant of E1 to immobilized E2-E3BP variants. For the fitting of interaction curves a 1:1 model was used after subtraction of the $E3BP_{only}$ reference sensor. In addition to the shown N-terminal truncations, all E2 subcomponents miss 8 residues C-terminally, all E3BP subcomponents miss 7 residues C-terminally. Error bars indicate duplicate measurements. Graph adopted according to Steinbrecher (2019).

was employed as ligand contained BSA and Tween-20 in their buffer composition to reduce unspecific protein binding, the reverse orientation not. Although the complementary experiments do not show the same trend for each variant, the general conclusion is the same: no severe influence of the lipoyl and hinge domain truncations were detected for the protein ti protein interaction via binding domains of hPDC sub-system E1 to E2. Furthermore, both orientations exhibited an cooperativity of E3BP on the association of E1 to E2-BD. The binding constants in each case differed in a range by a factor of less than three. For the assay orientation, Abdiche et al. (2008); Tobias and Kumaraswamy (2013) suggest to use the larger interaction partner as ligand. This is because of sensitivity of the BLI method, the larger the ligand, the larger the following interaction signal. In this work both interaction partners are large enough (more than 50 kDa) to result in good interaction signals. Thus the sensitivity should not be a problem. Furthermore Tobias and Kumaraswamy (2013) advise to take care of the underlying binding mechanism, for a 2:1 binding the interaction partner with two binding sites should be immobilized, whereas the analyte should posses one binding site. In case of the E1 to E2 system in this study a 1:1 model is assumed, thus the orientation should not play a role. Assay orientation influences the resulting constants gravely and comparison to surface plasmon resonance (SPR) method results can differ (Abdiche et al., 2008). Additionally, rebinding effects may occur on the biosensor caused by mass transport effects. Therefore, for each case results must be evaluated carefully.

Korotchkina and Patel (2008) investigated the hPDC E1 to E2 sub-system in comparison to prokaryotic PDC where the core BD is shared by E1 and E3. In contrary, mono-specificity of E1 to the E1-BD at E2 is reported for eukaryotic hPDC. In that study, point mutations were inserted to the C-terminal part of hE1β. The SPR method was used to immobilize the hL2S via thiol coupling through its lipoyl moiety. Thereby, hL2S is a variant missing the outer lipoyl and hinge domain (N-terminal) and in addition the inner domain (C-terminal). Kinetic parameters of WT hE1 interacting to WT hL2S are given with $k_a = 1.3{\cdot}10^6\,M^{-1}\,s^{-1}$, $k_d = 1.23{\cdot}10^{-2}\,s^{-1}$ and $K_D = 9.47\,nM$. This binding constants were collected via SPR with His-tagged proteins. For a point mutation at hE1β or E1-BD a 119 or 86 fold decrease in protein affinity was detected. The D289A substitution at C-terminal hE1β led to a total impairment of binding to E1-BD. In this present work the binding rates between mutants differed at most by factor of around three.
To have better comparability to literature studies, experiments with the orientation of the core being ligand (as this is the case for Korotchkina and Patel, 2008) are used. In this work for the only C-terminal truncated variant E2-ΔC8/E3BP-

ΔC7 following parameters were obtained: $k_a = 3{\cdot}10^5\,M^{-1}\,s^{-1}$, $k_d = 1.83{\cdot}10^{-4}\,s^{-1}$ and $K_D = 0.63\,nM$. The equilibrium binding constant in this work is smaller compared to Korotchkina and Patel (2008), thus it indicates higher affinity of the interaction partners. This is due to the very small dissociation constant measured in this work. The k_ds differ by an order of magnitude. Steinbrecher (2019) calculated the thermodynamic potential of the free enthalpy for both above mentioned binding constants. For Korotchkina and Patel (2008) this is ΔG = -46.4 kJ Mol^{-1}, in this work a similar ΔG = -53.2 kJ Mol^{-1} is obtained. The set up for the characterization of the kinetic parameters of PDC are different for SPR and BLI measurements. Korotchkina and Patel (2008) used hLS2 as binding partner to E1, which was immobilized N-terminal to a surface via thiol coupling method. In this way the BD is freely accessible and thus possible binding events to E1 can occur. I addition, in this way no interference of the His-tag to the immobilization surface can occur. In this work full E2 containing L1 an H1 as well as the inner domain was employed, in assembly with E3BD to represent the trimeric core. These core variants were immobilized at C-terminally. In this work no significant contribution of L1 or H1 could be seen to the E1 to E2 binding system. Therefore, the difference in the reported binding rates might be due to the missing inner domain. As Guo et al., 2017 already reported C-terminal truncation on E2-E3BP has an effect on the complex assembly behavior. The results collected in this work were obtained with mutants that contain a largely intact C-terminal inner domain, forming the smallest core units of trimers. Literature results published so far are based on monomeric molecules deficient of the C-terminal inner domain. Therefore it can be concluded that either the C-terminal inner domain, or the actual E2-E3BP trimer stabilize the E1 to E2 sub-binding system.

Lastly, BLI measurements can also be performed by employing streptavidin biosensors, onto which the ligand of interest can be immobilized. This would be advantageous to exclude a possible interference of the His-tagged proteins to the biosensor. On the other hand, the ligand would have to be biotinylated to bind onto the biosensor surface. This is done via coupling with reactive NHS-biotin. The NHS-activated biotin thereby reacts with primary amine groups from the lysine side chain of the protein (Fisher, 2012). Since the flexible arms of hPDC core contain many lysine residues, the coupling would be random and no clear orientation of the ligand could be achieved. Furthermore, if the flexible arms would be immobilized, the flexible arms of hPDC core would get rigid which might impede the binding.

6.5 E3 to E3BP binding system

In the following the E3 to E3BP binding system is the subject of investigation. Thereby both orientations, core immobilized as ligand as well as the core in free solution employed as analyte will be investigated by BLI measurements. The significance values shown in this section were calculated by using a two-sided t-test for independent samples, whereas the small number of samples (E2 = 5 and E3BP = 3) must be taken into account.

6.5.1 E3 to E3BP subcomplex, hPDC core variants as analyte

The orientation of E2-E3BP variants being the analyte is again evaluated via the kinetic k_{obs} constant. The results can be seen in Figure 6.7. Only slight differences in apparent association rates are present for non-truncated E3BP-ΔC7 along the progressing E2-ΔC8 truncation series. Solely, the most E2-ΔC8 truncated variant (ΔN260-E2-ΔC8) shows decreased apparent association rate. Although in that case E3BP was not modified, the most truncated E2 variant is influencing the E3-E3BP binding system. A cross-influence of the core partner without appropriate binding domain on the sub-binding system was already seen for the E1 to E2 system. The large E2 truncation just before its E1-BD already destabilizes this subcomponent, which exhibits its influences on the E3-E3BP system. In general, although some cooperative effects of E2 on the E3 to E3BP binding system might be visible, among all E2 truncation series there is no significant influence on the E3 to E3BP binding system.
Interestingly, an increase of apparent binding rate is visible along the E2 truncation series for the most truncated E3BP (ΔN128-E3BP-ΔC7) variant. When two or all lipoyl domains are missing at the core components, the apparent association rate for E3 to E2-E3BP is increasing. Loss of lipoyl domain at E2 while E3BP is truncated just before the E3-BD enhances binding rate for the E3 to E2-E3BP system.
An opposing effect occurs for the ΔN77-E3BP-ΔC7 variants, at which the LD is missing but H3 is intact. The variants containing ΔN77-E3BP-ΔC7 truncations have a significantly lower apparent association rate than the ΔN128-E3BP-ΔC7 truncation series ($p = 0.015$). As soon as H3 is present in this system the apparent association rates are the lowest (except both core components are truncated just before their binding domains, which seems to be the most unstable association partner for E3). The maximum difference in the apparent association rate among all tested variants is by a factor of 2.9.

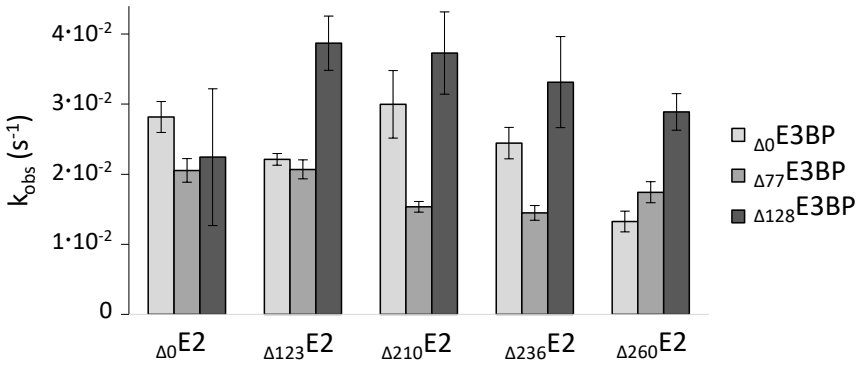


Figure 6.7: BLI investigation of immobilized E3 and different core variants in solution associated to it. In this setup there was almost no dissociation rate measured between the interaction partners. Since k_a is dependent of k_d, these rate constants could not be evaluated. Only an apparent association constant can be calculated for this data set, which is presented for all combinations of core variants. In addition to the shown N-terminal truncations, all E2 subcomponents miss 8 residues C-terminally, all E3BP subcomponents miss 7 residues C-terminally. This Figure is adopted according to Khansahib (2018).

6.5.2 E3 to E3BP subcomplex, hPDC core variants as ligand

When immobilizing E2-E3BP and subsequently interacting E3 in free solution, a concentration dependent artefact during the dissociation phase was obtained. For a binding signal which exceeded 1 nm, two dissociation curves overlaid, one fast dissociation curve within the first 100 seconds of the dissociation phase, and a following second and slow dissociation rate. Since the fast dissociation signal appeared occasionally, this is assumed to be an interference signal either resulting by loose E3 aggregated or any other kind of rearrangement of the free protein in buffer. This problem was tackled by adjusting the binding concentration to a signal intensity resulting between 0.6 to 0.8 nm. However, even tough to lower extend, the overlaying dissociation rates still occurred occasionally. Therefore the first 150 seconds of the dissociation phase were dismissed and only the last 350 seconds used for a 1:1 binding evaluation. More detailed information and data to this issue can be found in Steinbrecher (2019). Another reason of this mixed dissociation signal might be a loose interaction of E3 to the other core component E2. As already described in section 6.2.1, the E2:E3BP ratio is varying from mutant to mutant, mostly mutants with high E2 content showed this fast dissociation. Yang et al. 1997 showed 4 % remaining activity of hPDC without E3BP subcomponent, but E3 in high concentration. This might, but

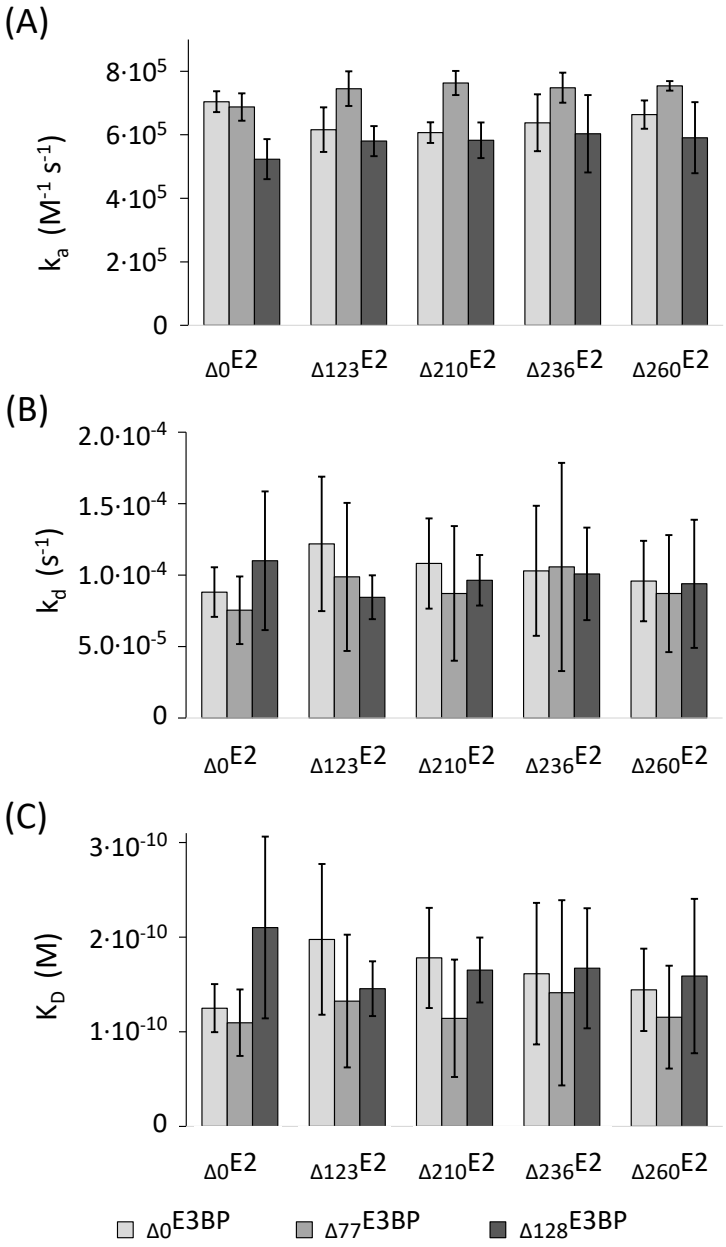


Figure 6.8: (A) Association rate, (B) dissociation rate and (C) equilibrium binding constant for core variants as ligand and E3 as analyte. The rate constants were measured via BLI, and a 1:1 model was applied to evaluate the data, whereas the first 150 seconds of the dissociation steps were dismissed. In addition to the shown N-terminal truncations, all E2 subcomponents miss 8 residues C-terminally, all E3BP subcomponents miss 7 residues C-terminally. Error bars indicate triplicate measurements. Graph adopted according to Steinbrecher (2019).

not necessarily, hint on a loose interaction of E3 to E2 to perform the overall hPDC reaction.

The association rates shown in Figure 6.8 are all close together. They vary mostly by a factor of 1.35. Thereby the truncations on E2 exhibit no influence on this binding system. Among the E3BP variants, each of the three truncation series is significantly distinguishable from the others ($p < 0.02$). The highest k_a is exhibited by the ΔN77-E3BP-ΔC7 variants, followed by ΔN0-E3BP-ΔC7 variants. The lowest k_a is exhibited by the ΔN128-E3BP-ΔC7 truncation series. For the k_d no significant differences were detected for E2 and E3BP variations, but the standard deviations were large. This finally influenced K_D, also here no significant differences among all tested variants are present. However, influenced by the high k_a values, the ΔN77-E3BP-ΔC7 variants show slightly decreased K_D values compared to the other variants. The maximal difference of the equilibrium constant from the smallest to the largest is less than a factor of two. Therefore, this results suggest marginal raise of the association rate for missing LD on E3BP, but no further influence of all the other hPDC core domains on the E3 to E3BP sub-binding system. As can be seen in Table 6.4 the parameters obtained for the WT E2-E3BP to E3 interaction is highly similar to the values obtained for the 15 core variants averaged. The $E3BP_{only}$ to E3 system exhibits the highest affinity (smalles K_D), however also these values are similar to the WT.

Table 6.4: The E2-E3BP WT and variants as well as $E3BP_{only}$ proteins were immobilized on the biosensor and interaction to E3 in free solution was measured via BLI. The location of the His-tag (N or C-terminal) is indicated next to the variant. Table adopted according to Ilhan *et al.* submitted to PLOS ONE (2020).

Core component	k_a (M^{-1} s^{-1})	k_d (s^{-1})	K_D (M)	fold -	ΔG (kJ Mol^{-1})
WT E2-E3BP (N6His)	$6.23 \cdot 10^5$	$5.10 \cdot 10^{-5}$	$8.20 \cdot 10^{-11}$	1	-58.5
$E3BP_{only}$ (N6His)	$5.57 \cdot 10^5$	$1.77 \cdot 10^{-5}$	$3.18 \cdot 10^{-11}$	0.39	-60.9
aver. core var. (C6His)	$6.54 \cdot 10^5$	$9.71 \cdot 10^{-5}$	$15.12 \cdot 10^{-11}$	1.8	-57.0

6.5.3 Comparison of E3 to E3BP binding system

Because the complement BLI experiments were performed with different buffer background, the additives (BSA and Tween-20) added to the orientation of core as ligand were let out for one passage. In this way the influence on the kinetic constants obtained from both buffer compositions can be compared. To note, all experiments regarding core as ligand were performed with the addition of

BSA and Tween-20, the contrariwise experiment does not contain these buffer components. So in case of core as ligand a repetition without BSA and tween was performed and compared. For the addition of BSA and Tween-20 an influence on the absolute rates, but not on the trend is revealed. The averaged values for the E3 association on the core without BSA and Tween are: $k_a = 3.01 \cdot 10^5\,M^{-1}\,s^{-1}$, $k_d = 1.76 \cdot 10^{-4}\,s^{-1}$ and $K_D = 0.63\,nM$. For the addition of BSA and Tween: $k_a = 6.54 \cdot 10^5\,M^{-1}\,s^{-1}$, $k_d = 9.71 \cdot 10^{-5}\,s^{-1}$ and $K_D = 0.15\,nM$. Averaged, the dissociation rates were smaller for the case with BSA and Tween-20 and thus, the resulting equilibrium constant was also slightly smaller. This indicates a higher affinity of the binding system with BSA and Tween-20 addition, and could represent a more specific hPDC binding. It is assumed that BSA binds unspecific to proteins and thus only the actual high affinity protein to protein interaction occurs.

Since the equilibrium binding constants for all variants are similar, the averaged value of 0.15 nM shall be used for literature comparison. Furthermore, the assay orientation used in this case matches the assay orientation of literature data. A K_D of 1.6 nM for WT hE3 to L3S1, calculated via isothermal titration calorimetry (ITC) was reported Park and Patel, 2010. L3S1 thereby means hE3BP missing the inner domain, but containing all hinge-, lipoyl- and BD-domains. The same binding partner was investigated via SPR in another study, and a K_D of 5.26 nM was obtained Patel et al., 2009. Although the same binding partners, there is a difference in the equilibrium constant of more than a factor three (ITC vs SPR). The differences of the investigated variants in this present work were mostly lower than a factor two. As a further method, the binding of E3-BD to E3 was investigated by ITC Brautigam et al., 2006. They used a fusion construct of E3-BD containing the lipoyl domain and obtained a value of 0.75 nM. All these values show a dependence of the measurement technique (BLI vs. SPR vs. ITC) as well as the modified proteins. It is difficult to compare all these values since in each study not the actual full length WT protein, but just a part of it was investigated. Nevertheless, they are all in a highly affine range in the low nM. Steinbrecher (2019) approached an comparison between the two most diverging equilibrium constants by calculating the thermodynamic free enthalpy. These are very close together (factor of 1.2). For Patel et al., 2009 the $\Delta G = -47.9\,KJ\,Mol^{-1}$ and in this work $\Delta G = -56.8\ KJ\ Mol^{-1}$.

For the comparison of kinetic rates of E3 to E3BP, literature values of Patel et al. (2009) (determined by SPR) will be used. There, the core component L3S1 was immobilized through thiol coupling at the lipoyl domain. In this work experiments with both interaction partners each immobilized once was performed. To have a better comparability, the orientation of the core used as ligand will be

employed here. However, in this work the inner domain of the core component was immobilized. Patel et al. (2009) obtained $k_a = 2.74{\cdot}10^6\,M^{-1}\,s^{-1}$, $k_d = 1.44{\cdot}10^{-2}\,s^{-1}$. In this work the averaged values are $k_a = 6.5{\cdot}10^5\,M^{-1}\,s^{-1}$ and $k_d = 9.7{\cdot}10^{-5}\,s^{-1}$. Evidently, the rates in this work are lower, particularly the dissociation rate is more than 100 x smaller. This big discrepancy might be explained to some extend by the missing inner domain in the literature study Patel et al., 2009. This would hint on the inner domain stabilizing the complex and keeping it together. Since the core of the complex is assembled via the inner domain, which then assembles the whole complex this is thinkable. In this work the trimer core component was investigated, in literature so far monomeric molecules were examined. Therefore not only the C-terminal inner domain, but maybe formed covalent trimer assembly stabilizes the formed sub-system interaction. Furthermore, the orientation of the immobilization (N- or C-terminal) might also play a role for the accessibility of the interaction partner. Since for both experiments there were hinge regions between the immobilized component and the BD, the accessibility of the interaction partner should be given.

6.6 Additional influences on the binding

Two further factors (temperature and CoA) for the interaction of hPDC subcomponents were tested in single experiments.

6.6.1 Thermostability of the core

A first and simple temperature stability test was performed with subcomponents E2-E3BP and E3 to investigate the influence of a heat treatment on hPDC subcomponents. Both subcomponents were heated for 10 minutes at 95 °C and subsequently cooled down to RT again. After, each subcomponent was investigated singly for its capability to still interact with its non-heat treated binding partners. As can be seen in Table 6.5, treated E3 loses its ability to bind to E3BP. Whereas treated E2-E3BP nearly fully retains its binding capacity compared to untreated E2-E3BP. It should be noted that this represents a single measurement and needs verification by replicates.
Another protein system, namely the glycine decarboxylase complex or glycine cleavage system (GCS), shows similarity to hPDC in some points. It is also composed of several different proteins (T-, P-, L- and H-protein), located in the mitochondrial matrix and catalyzes multi-step reactions. The lipoic acid-containing H-protein interacts with all other subcomponents to perform a full reaction cycle (Douce et al., 2001), which is similar to E2-E3BP. Thermostability of H-protein is already reported (Fujiwara et al., 1979). During the purification

procedure of H-protein it was boiled for 90 seconds and afterwards showed the same activity as untreated H-protein.
For prokaryotic hPDC, thermostability until 54 °C of the BD was shown, which is astonishing for such a small protein (Spector et al., 1998). It owns its stability, structure formation and activity to a buried aspartic acid residue. If this is transferable to human PDC, this would mean that the BD is denatured reversibly, and so is able to maintain its function.

Table 6.5: Influence of heat treatment of 95 °C for 10 minutes to hPDC core (E2-E3BP) and subunit E3. Interaction studies were performed with one heat-treated subcomponent interacting with an non-treated interaction partner. These BLI measurements were conducted with His2 biosensors. The max. response of the signal (in nm) is shown next to the kinetic constants to enable a relative comparison of the forming bio-layer. Table adopted according to Steinbrecher (2019).

treated subcomponent	non	E2-E3BP	E3
Max. response (nm)	0.79	0.68	ND
k_a (M^{-1} s^{-1})	$4.76 \cdot 10^5$	$3.44 \cdot 10^5$	ND
k_d (s^{-1})	$1.00 \cdot 10^{-4}$	$5.19 \cdot 10^{-5}$	ND
K_D (nM)	0.21	0.15	ND

6.6.2 CoA addition

The addition of 0.5 mM CoA on the E3BP to E3 system was investigated in a single measurement. Only E3BP was immobilized on a His2 biosensor and E3 in solution interacted in three different concentrations with it. Hereby, no significant difference could be detected for k_a or k_d, which thus result in similar K_D of 0.25 nM without CoA addition and 0.21 nM with CoA addition (detailed data can be found in Steinbrecher, 2019). This finding indicates that cofactor CoA has no influence on the binding of the sub-system E3BP to E3. Prajapati et al. (2019) reported that E3BP is not capable of binding CoA, only E2 is binding CoA. Therefore this result is consistent with literature.

6.7 Conclusion and outlook

hPDC core variants deficient of N-terminal lipoyl and hinge domains were constructed. Thereby, 15 different variants with single or double truncations on the core components E2 and E3BP could be successfully cloned and expressed. An investigation of the single truncated variants via DLS exhibited different populations, whereas the largest population represents a protein agglomerate larger than the 60meric core. With progressive deletion of the N-terminal LDs on E2, the hydrodynamic radius of this large agglomeration decreased step wise. Most probably the truncations on E3BP were too small to exhibit an impact on the size. However, because of limitations regarding resolution and overlaying effects of larger agglomerates shielding smaller agglomerates, an investigation of all aggregate populations via DLS was hindered.
Furthermore, each N-terminal core truncation led to a complete loss of hPDC overall activity. Although the complex is still formed, as the results for the interaction studies show, the hPDC activity is impaired. This can be attributed to the absence of the LDs that are essential for complex activity.

In preliminary studies - to develop a protocol for hPDC interaction studies by employing the new method of BLI - first challenges were revealed. Concentration ranges and time spans were identified in which the interaction of hPDC subcomponents could be evaluated. Furthermore, by addition of reference samples (GFPuv), reference sensors ($E3BP_{only}$), and later the addition of BSA and Tween-20 to the protein buffer, the influence of interfering factors could be minimized. The question how to orientate the interaction partners on the biosensor to enable good interaction signals was addressed by experiments with different setups. Best results could be obtained for the core components employed as ligand (immobilized) and E1 or E3 as interaction partner in solution (analyte). The low E2 content of some purified core variants resulted in low signals for E1 interactions, which was not sufficient for data evaluation in some cases.

The examination of the non-covalent interaction of E1 or E3 to the trimeric core components revealed differences in binding affinity and cross-influence. Although the core contains mono-specific BDs, for the E1 to E2 binding system an influence of both core components, E2 and E3BP is detected. As soon as the LDs on E2 and E3BP were missing, the binding affinities for the E1 to E2 system increased slightly. Therefore, a passive, probably steric role of the N-terminal domains on the highly specific protein interaction via binding domains can be concluded. The K_D value among all variants vary mostly by a factor of 3.2, whereas the elimination of all LDs led to a decreased K_D. The investigation of the E3BP to E3 binding system shows no influence of the core component E2

on the protein interaction. Here, a significant increase in the association rate was detected for the ΔN77-E3BP-ΔC7 variants, although among all variants the association rates varied by mostly 1.5. A missing LD on E3BP but intact hinge region exhibited highest K_D values. However, the N-terminal truncations have only minor effect on the overall very strong binding of hPDC subcomponents: E1 to E2 averaged $K_D = 4.4 \cdot 10^{-10}$ M and E3BP to E3 averaged $K_D = 1.5 \cdot 10^{-10}$ M. A marginal influence of the LDs could be demonstrated on the binding affinities of hPDC subcomplexes. However, the E1 to E2 subcomplex is cross-influenced by E3BP. Furthermore, the LDs are essential for hPDC function, but not for the complex formation.
Despite that, another effect was revealed in this protein interaction study. So far in literature hPDC sub-system binding was investigated with monomeric and C-terminal inner domain deficient core variants. In this work the C-terminus of the core was mostly intact and formed the hPDC core trimers, which were used for the immobilization on the biosensor. Comparing literature values to this study, a decrease in the dissociation rates by 70 to 100 times was seen. Therefore, although the N-terminal end of the core components are not having a severe influence on the complex formation, the presence of the inner domain of the core might exhibit a stabilizing effect on the formed complex, since the hPDC core is assembled C-terminally. This effect should be studied in more detail in future. These new insights to interrelations within hPDC might facilitate the approach of targeted multienzyme-complex engineering in future.

The parameter of CoA exhibited no influence on the E2-E3BP to E3 sub-system, which is consistent with literature. However, Prajapati et al. (2019) showed that E2 is binding CoA and therefore, the influence of CoA on the E1 to E2-E3BP binding could be tested. Thereby the LDs on E2 should stay intact since the CoA binding occurs there. Because the LDs exhibit a minor influence on the sub-system binding (as shown in this study), only a little effect of CoA can be expected.
Furthermore, first indications on thermostability of E2-E3BP was seen, but not for the interacting subcomponent E3. This phenomenon of heat stability for E2-E3BP is not reported so far and worthwhile further attention. A more systematic investigation by testing different temperatures and time spans as well as different "regeneration" strategies can be aimed at. In this way it could be clarified which subcomponent developed thermostability and for which purpose. The mentioned property might be of interest for reactions under harsh temperature conditions.

The partially purified proteins may be polished further by an additional purification step like size exclusion chromatography (SEC). This was not pursued in this work since BLI measurements are performed with highly specific biosensors that can perform experiments with crude background. Nevertheless, as seen in this study, SEC could help to filter out the excess monomeric E3BP core components and thus balance the ratio of E2:E3BP. In addition, the gene for E2 on the expression vector could be duplicated and might increase the E2:E3BP ratio towards E2. Furthermore, an immobilization strategy independent of a His-tag would improve the BLI data quality when employing Anti-His biosensors and His-tagged analyte.

Up to now crystallography data of hPDC in good resolution is missing, probably due to the long and flexible linker arms. In this study it could be shown that the influence on the binding properties of the sub-systems is minor if both core components are truncated shortly before the binding domain. Therefore, the constructed mutants could be employed to facilitate the crystallization process and deliver native results of inner domain and BD of hPDC core.

Chapter 7

Summary and future perspectives

Each bioprocess is defined by the employed biocatalyst, which specifically converts raw materials to desired products. As a representative, hPDC performs chemical reactions under mild conditions, being an environmental friendly alternative to chemical synthesis. The favourable properties of hPDC - such as self-assembly, modularity and efficiency - were investigated and explored from different perspectives in this work. The new insights gained and the established automated protein engineering strategy will help to exploit the characteristics of hPDC for improved bioprocesses, and might even serve for medical purposes.

At first, the feasibility of producing hPDC subcomponents in a cell free manner for rapid protein manipulation, production and screening was tested. Thereby hPDC subcomponents E1α-E1β, E2-E3BP and E3 were expressed via a commercial kit of reconstituted, so called PURE cell free lysate. The expression yields were to low for activity testing, furthermore the expression background interfered with the activity assay. The purification of the His-tagged proteins via magnetic beads could be performed in small scale, allowing for buffer exchange and concentration determination, eventually leading to reliable functionality testing in future attempts.
In parallel, homemade *E.coli* based cell lysates were produced for low cost protein expression, offering the freedom to adjust and screen for different expression conditions. An endogenous, polymerase based coupled transcription and translation lysate (TX-TL) after Sun et al., 2013 was established by utilizing variants of the role model protein GFP. Under optimized conditions, a yield of up to 90 µg/mL of fluorescent proteins was detected on plasmid based TX-TL lysate reaction. This not only validates protein production, but also the functionality relying on the correct folding of the protein to develop its fluorescence. Parameters

influencing the TX-TL lysate such as size, temperature, tube geometry, DNA concentration and incubation time were investigated to finally enable a transfer of plasmid-based to crude LET-based CF protein production. After some parameters were optimized, the LET based protein yield reached a third of the yield achieved by plasmids based CFPS. The optimizations included the addition of non-coding overhangs on the LET sequence and the employment of the exonuclease inhibitor gamS. The LETs were produced by a two-step assembly PCR method enabling fast and flexible dsDNA generation with high concentration and good sequence fidelity. In this manner DNA production from scratch and mutant generation without tedious cellular cloning procedure were established, saving a great deal of experimental time. Finally, the detection and quantification of the produced fluorescent molecules were accomplished by the development of a novel flow cytometer based method. the new method allows for fast, robust and highly sensitive measurements, even of samples in crude background (Wurm et al. 2019).
While reasonable yields could be achieved in this work, there is still future potential to increase CF produced protein yields. The hPDC subcomponent production could not be detected via the homemade cell lysate. Therefore, an optimization of the cell lysate and lysate system could be realized through an eukaryotic cell lysate and the addition of chaperones, mild detergents or even vesicles for the correct folding and solubility of the target protein. Also depending on the nature of the protein, a strong T7 based cell lysate might lead to better yields (Chong 2014; Gregorio et al. 2019).

The established process chain of LET generation and CFPS was embedded in an automated protein optimization workflow. Thereby, different point mutations in the sequence of GFP were chosen to shift its wavelength spectrum. Accordingly, LETs were produced, subsequently employed in a CF reaction and screened via flow cytometry. This modular and flexible workflow was completely automatized at a liquid handling station and processed six mutants in less than six hours, leaving potential to shorten this time span even more through optimized incubation times and higher degree of parallelization. To evaluate the spectral shift of the produced mutants, a cost function (ratio of the fluorescence signal in different fluorescence channels) was set up. Proteins resulting from purified LETs meet the cost function, whereas results from crude LETs employed in CFPS were less obvious and more iteration rounds were needed to obtain robust results. However, at the same time synergistic effects of different mutation sites became apparent, impacting the final results in an unknown fashion. These synergistic effects should be taken into account for future set ups and the evaluation should be adopted to allow gathering of such complex relationships. The next steps

will comprise the establishment of an algorithm to evaluate the results and give suggestions for subsequent protein optimization loops. This proof of concept for a new, fast and automated protein optimization workflow with the role protein GFP can be applied to other proteins of interest in future via a fluorescence tag or an altered detection method. Moreover, protein to protein interaction could be detected via a set up using FRET events (Cubitt et al., 1995; Heim et al., 1994).

Beside this route, the exploration of hPDC was continued with conventionally expressed proteins. One aspect studied was based on a theoretical study of *in silico* nonlinear kinetic behaviour of PDC published by Zeng et al. (2002). For the experimental validation, a multi-parameter screening was performed with self-constituted hPDC to identify reaction conditions causing this nonlinearities in enzyme activity. Varying subcomponent ratios, substrate concentration and total protein amounts were parameters tested through an online, *in vitro* batch-wise activity assay with fully assembled hPDCs. Pyruvate was thereby identified as the biggest influence factor which indeed causes oscillations during hPDC activity. The upper threshold concentration at which transient oscillatory behavior was found is around 0.4 mM pyruvate. This indicates regulatory effects occurring at physiological pyruvate concentrations. A further modulation of the amplitude and period number of the transient oscillation during hPDC activity was found to be in charge of subcomponent E1, which elongates the transient oscillation when diluted. This component is the main actor of hPDC regulation (Patel and Korotchkina 2006; Strumiło 2005) and thus, a likely candidate for the transient oscillations. No clear evidence on nonlinearity arising from subcomponent levels of E2-E3BP or E3 were found experimentally. Apart from this, during the hPDC activity a decay of NADH after the transient oscillation was detected. Whether this can be fully assigned to unstable NADH decaying over time or another underlying effect needs to be clarified. A further phenomenon, probably caused by allosteric substrate inhibition, was found for pyruvate in excess conditions. Here, the start of pyruvate conversion was delayed for several minutes until the usual enzyme activity kinetics followed.
These mentioned effects certainly deserve more attention. A connection to hPDC regulation is evident which will have influence on process parameters like altered reaction rates and yields during the transition state, or even adoption of a lower steady state metabolism. Initially, a minimal set of parameters was tested within this study to figure out experimental conditions for oscillatory hPDC behavior. In future this set could be expanded to include TPP and kinase inhibitors which act directly on E1. The investigation of TPP as a cofactor of E1 could also exhibit an influence on the condition of pyruvate in excess since E1 is the rate limitng step in hPDC (Strumiło, 2005; Sümegi and Alkonyi, 1983). Next

to the overall hPDC assay, also a DCPIP assay according to Ke et al. (2014) could be tested, in which only E1 activity is determined. In this way the kinetic conversion behavior of E1 independent from E2-E3BP and E3 can be analyzed and might deliver additional insights. A further parameter considering the overall complex could be CoA, since acetyl-CoA acts as inhibitor on E3 (Selivanov et al., 1994; Zeng et al., 2002). Finally, an extension of the batch-wise mode to a continuous mode would be thinkable to possibly keep undamped oscillations ongoing.

As a last aspect of hPDC, a cross-influence on the dynamic complex formation of the N-terminal core domains was investigated in detail, split up in distinct sub-binding systems of E1 to E2 and E3 to E3BP. This topic was based on a C-terminal minimized hPDC core variant (Guo et al., 2017). Both core components, E2 and E3BP, were truncated N-terminally at different positions at the liopyl and hinge regions until the respective BDs. In this way 15 newly variants were obtained and evaluated. Some single truncated core variants were tested by DLS measurements which revealed different sized core populations. The largest population exhibited a step wise-reduction in size for the progressive degree of truncation. However, all these variants exhibited protein to protein interaction and were still able to form a complex. The kinetic interaction parameters were determined by BLI measurements. Thereby, the trimeric core component, or the surrounding subcomponents E1 or E3 were immobilized via their His-tag and the interaction partner was provided in solution. No matter which orientation or sub-system, an influence of the N-terminal core domains on the subcomplex assembly by a factor of two to three was found, which is significant, but small compared to the influence of the binding domain itself. A cross-influence of the E3BP core component on the E1 to E2 subcomplex was revealed. In turn, E2 is not influencing the E3BP to E3 system. For this system a slight but significant increase in the association rate was measured for LD deficient E3BP components. An average of the binding affinities of all variants were calculated to be K_D = 0.15 nM for E3 to E3BP and 0.44 nM for E1 to E2. These values are one magnitude smaller than literature values (Brautigam et al., 2006; Korotchkina and Patel, 2008; Park and Patel, 2010; Patel et al., 2009), which is mainly caused by the small dissociation rate constants recorded in this study. This means, once an association is formed, the interaction partners hardly dissociate from each other. The core interaction partners used in literature were monomeric and deficient of an inner domain, in this study the inner domain of the core was mainly intact. Therefore, the altered dissociation behavior of the complex might result from a stabilizing effect of the inner domain. Consequently, a cross-influence of C-terminal E2-E3BP domains on the complex stability could be investigated in future. All the N-terminal truncated hPDC core variants lost their hPDC

overall activity. Whether this is due to the C- and N-terminal deletions only, or a possible implication with the His-tag shifted to the C-terminus requires further investigation.
The mutants produced in this study might improve the crystallization process of hPDC, which is troublesome so far because of the long and flexible linker arms. The available mutants are able to deliver native results on inner domain and BD of trimeric hPDC core without influencing the subcomponent binding gravely.

Bibliography

Y. Abdiche, D. Malashock, A. Pinkerton, and J. Pons. Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. *Analytical Biochemistry*, 377(2):209–217, 2008. doi: 10.1016/j.ab.2008.03.035.

A. Adan, G. Alizada, Y. Kiraz, Y. Baran, and Nalbant A. Flow cytometry: basic principles and applications. *Critical Reviews in Biotechnology*, 37(2):163–176, 2017. doi: 10.3109/07388551.2015.1128876.

K. R. Albe, M. H. Butler, and B. Wright. Cellular Concentrations of Enzymes and Their Substrates. *Journal of Theor. Biology*, 143:163–195, 1990. doi: 10.1016/S0022-5193(05)80266-8.

S. K. Amundsen, T. Spicer, A. C. Karabulut, L. M. Londoño, C. Eberhart, F. Vega, F. Virneliz, T. D. Bannister, P. Hodder, and G. R. Smith. Small-molecule inhibitors of bacterial AddAB and RecBCD helicase-nuclease DNA repair enzymes. *ACS Chemical Biology*, 7(5):879–891, 2012. doi: 10.1021/cb300018x.

D. L. Andrews. *Encyclopedia of applied spectroscopy: Molecular Fluorescence*. Wiley-VCH, Weinheim Germany, 2009. ISBN 9783527407736.

M. Aoki, T. Matsuda, Y. Tomo, Y. Miyata, M. Inoue, T. Kigawa, and S. Yokoyama. Automated system for high-throughput protein production using the dialysis cell-free method. *Protein Expression and Purification*, 68(2):128–136, 2009. doi: 10.1016/j.pep.2009.07.017.

F. H. Arnold. Directed Evolution: Bringing New Chemistry to Life. *Angewandte Chemie (International ed. in English)*, 57(16):4143–4148, 2018. doi: 10.1002/anie.201708408.

J. A. J. Arpino, P. J. Rizkallah, and D. D. Jones. Crystal structure of enhanced green fluorescent protein to 1.35 Å resolution reveals alternative conformations for Glu222. *PLOS ONE*, 7(10):e47132, 2012. doi: 10.1371/journal.pone.0047132.

O. Ball, S. Robinson, K. Bure, D. A. Brindley, and D. Mccall. Bioprocessing automation in cell therapy manufacturing: Outcomes of special interest group automation workshop. *Cytotherapy*, 20(4):592–599, 2018. doi: 10.1016/j.jcyt.2018.01.005.

D. L. Bates, M. J. Danson, G. Hale, E. A. Hooper, and R. N. Perham. Self-assembly and catalytic activity of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. *Nature*, 268(5618):313–316, 1977. doi: 10.1038/268313a0.

R. Bhambure, K. Kumar, and A. S. Rathore. High-throughput process development for biopharmaceutical drug substances. *Trends in Biotechnology*, 29(3):127–135, 2011. doi: 10.1016/j.tibtech.2010.12.001.

C. A. Brautigam, R. M. Wynn, J. L. Chuang, M. Machius, D. R. Tomchick, and D. T. Chuang. Structural insight into interactions between dihydrolipoamide dehydrogenase (E3) and E3 binding protein of human pyruvate dehydrogenase complex. *Structure*, 14(3):611–621, 2006. doi: 10.1016/j.str.2006.01.001.

C. A. Brautigam, R. M. W., J. L. Chuang, and D. T. Chuang. Subunit and catalytic component stoichiometries of an in vitro reconstituted human pyruvate dehydrogenase complex. *The Journal of Biological Chemistry*, 284(19):13086–13098, 2009. doi: 10.1074/jbc.M806563200.

A. Broadbent. Aminoacyl-tRNA Synthetase Production for Unnatural Amino Acid Incorporation and Preservation of Linear Expression Templates in Cell-Free Protein Synthesis Reactions. *Birmingham Young University*, Master thesis, 2016.

G. K. Brown, R. M. Brown, R. D. Scholem, D. M. Kirby, and H.-H. Dahl. The Clinical and Biochemical Spectrum of Human Pyruvate Dehydrogenase Complex Deficiency. *Annals New York Academy of Sciences*, 573(1):360 – 368, 1989. doi: 10.1111/j.1749-6632.1989.tb15011.x.

E. D. Carlson, R. Gan, C. E. Hodgman, and M. C. Jewett. Cell-free protein synthesis: Applications come of age. *Biotechnology Advances*, 30(5):1185–1194, 2012. doi: 10.1016/j.biotechadv.2011.09.016.

M. Carson, D. H. Johnson, H. McDonald, C. Brouillette, and L. J. Delucas. His-tag impact on structure. *Acta crystallographica. Section D, Biological crystallography*, 63(Pt 3):295–301, 2007. doi: 10.1107/S0907444906052024.

F. Caschera and V. Noireaux. Synthesis of 2.3 mg/ml of protein with an all *Escherichia coli* cell-free transcription-translation system. *Biochimie*, 99: 162–168, 2014. doi: 10.1016/j.biochi.2013.11.025.

M. Castellana, M. Z. Wilson, Y. Xu, P. Joshi, I. M. Cristea, J. D. Rabinowitz, Z. Gitai, and N. S. Wingreen. Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nature Biotechnology*, 32(10): 1011–1018, 2014. doi: 10.1038/nbt.3018.

S. Chong. Overview of cell-free protein synthesis: historic landmarks, commercial systems, and expanding applications. *Current Protocols in Molecular Biology*, 108:16.30.1–11, 2014. doi: 10.1002/0471142727.mb1630s108.

E. M. Ciszak, L. G. Korotchkina, P. M. Dominiak, S. Sidhu, and M. S. Patel. Structural basis for flip-flop action of thiamin pyrophosphate-dependent enzymes revealed by human pyruvate dehydrogenase. *The Journal of Biological Chemistry*, 278(23):21240–21246, 2003. doi: 10.1074/jbc.M300339200.

E. M. Ciszak, A. Makal, Y. S. Hong, A. K. Vettaikkorumakankauv, L. G. Korotchkina, and M. S. Patel. How dihydrolipoamide dehydrogenase-binding protein binds dihydrolipoamide dehydrogenase in the human pyruvate dehydrogenase complex. *The Journal of Biological Chemistry*, 281(1):648–655, 2006. doi: 10.1074/jbc.M507850200.

D. P. Clark, N. Jean Pazdernik, A. Held, and B. Jarosch. *Molekulare Biotechnologie: Grundlagen und Anwendungen*. Spektrum Akademischer Verlag, Heidelberg, 2009. ISBN 9783827421289.

G. M. Cooper. *The cell: A molecular approach*. ASM Press, Washington, DC, 2. ed. edition, 2000. ISBN 9780878931064. URL `http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=cooper`.

B. P. Cormack, R. H. Valdivia, and S. Falkow. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, (173):33–38, 1996. doi: 10.1016/0378-1119(95)00685-0.

E. J. Crampin, S. Schnell, and P. E. McSharry. Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. *Progress in Biophysics and Molecular Biology*, 86(1):77–112, 2004. doi: 10.1016/j.pbiomolbio.2004.04.002.

F. Crick. Central Dogma of Molecular Biology. *Nature*, 227:561–563, 1970.

A. B. Cubitt, R. Heim, S. R. Adams, A. E. Boyd, L. A. Gross, and R. Y. Tsien. Understanding, improving and using green fluorescent protein. *Trends in Biochemical Science*, (20 (11)):448–455, 1995.

S. Dahndayuthapani, L. E. Via, C. A. Thomas, P. M. Horowitz, D. Deretic, and V. Deretic. Green fluorescent protein as a marker for gene expression and cell biology of mycobacterial interactions with macrophages. *Molecular Microbiology*, (17 (5)):901–912, 1995.

R. N. Day and M. W. Davidson. The fluorescent protein palette: tools for cellular imaging. *Chemical Society Reviews*, 38(10):2887, 2009. doi: 10.1039/b901966a.

C. A. Denard, H. Ren, and H. Zhao. Improving and repurposing biocatalysts via directed evolution. *Current Opinion in Chemical Biology*, 25:55–64, 2015. doi: 10.1016/j.cbpa.2014.12.036.

W. Deng, X.-K. He, L.-D. Zhang, A. Zeng, J.-L. Song, and J.-J. Zou. Target infrared detection in target spray. *Guang Pu Xue Yu Guang Pu Fen Xi*, 28 (10):2285–2289, 2008.

P. N. Depta, U. Jandt, M. Dosta, A.-P. Zeng, and S. Heinrich. Toward multiscale modeling of proteins and bioagglomerates: An orientation-sensitive diffusion model for the integration of molecular dynamics and the discrete element method. *Journal of Chemical Information and Modeling*, 59(1):386–398, 2019. doi: 10.1021/acs.jcim.8b00613.

R. Douce, J. Bourguignon, M. Neuburger, and F. Rébeillé. The glycine decarboxylase system: a fascinating complex. *Trends in Plant Science*, 6(4):167–176, 2001. doi: 10.1016/S1360-1385(01)01892-1.

G. Dupont, L. Combettes, G. S. Bird, and J. W. Putney. Calcium oscillations. *Cold Spring Harbor Perspectives in Biology*, 3(3), 2011. doi: 10.1101/cshperspect.a004226.

D. Eliezer. Structurally Disordered Proteins. In Gordon C. Roberts, editor, *Encyclopedia of Biophysics*, pages 2484–2488. Springer Berlin, Berlin, 2013. ISBN 978-3-642-16711-9. doi: 10.1007/978-3-642-16712-6{_}610.

D. T. Eriksen, J. Lian, and H. Zhao. Protein design for pathway engineering. *Journal of Structural Biology*, 185(2):234–242, 2014. doi: 10.1016/j.jsb.2013.03.011.

Thermo Fisher. Easy molecular bonding - crosslinking technology. *Crosslinkers Technical Handbook, Thermo Fisher Scientific Inc.*, 2012.

R. Fuchs, D. D. Y. Ryu, and A. E. Humphrey. Effect of Surface Aeration on Scale-Up Procedures for Fermentation Processes. *Industrial & Engineering Chemistry Process Design and Development*, 10(2):190–196, 1971. doi: 10.1021/i260038a009.

K. Fujiwara, K. Okamura, and Y. Motokawa. Hydrogen carrier protein from chicken liver: Purification, characterization, and role of its prosthetic group, lipoic acid, in the glycine cleavage reaction. *Archives of Biochemistry and Biophysics*, 197(2):454–462, 1979. doi: 10.1016/0003-9861(79)90267-4.

A. L. Givan. Chapter 2 principles of flow cytometry: An overview. In Zbigniew Darzynkiewicz, editor, *Cytometry*, volume 63 of *Methods in Cell Biology*, pages 19–50. Acad. Press, San Diego, 2001. ISBN 9780125441667. doi: 10.1016/S0091-679X(01)63006-1.

A. Goldbeter. Oscillatory enzyme reactions and Michaelis-Menten kinetics. *FEBS Letters*, 587(17):2778–2784, 2013. doi: 10.1016/j.febslet.2013.07.031.

B. N. Goldstein and V. A. Selivanov. Unusual kinetic behavior predicted for complexes a-keto acid dehydrogenase. *FEBS Letters*, 319(3):267–270, 1993. doi: 10.1016/0014-5793(93)80560-h.

B. C. Goodwin. Oscillatory behaviour in enzymatic control processes. *Advances in Enzyme Regulation*, 3:425 – 438, 1965. doi: 10.1016/0065-2571(65)90067-1.

S. Gopalakrishnan, M. Rahmatullah, G. A. Radke, S. Powers-Greenwood, and T. E. Roche. Role of protein x in the function of the mammalian pyruvate dehydrogenase complex. *Biochemical and Biophysical Research Communications*, 160(2):715–721, 1989. doi: 10.1016/0006-291X(89)92492-3.

N. E. Gregorio, M. Z. Levine, and J. P. Oza. A User's Guide to Cell-Free Protein Synthesis. *Methods and Protocols*, 2(1), 2019. doi: 10.3390/mps2010024.

J. Guo, S. Hezaveh, J. Tatur, A.-P. Zeng, and U. Jandt. Reengineering of the human pyruvate dehydrogenase complex: from disintegration to highly active agglomerates. *The Biochemical Journal*, 474(5):865–875, 2017. doi: 10.1042/BCJ20160916.

R. A. Harris, R. Paxton, S. M. Powell, G. W. Goodwin, M. J. Kuntz, and A. C. Han. Regulation of branched-chain a-ketoacid dehydrogenase complex by covalent modification. *Advances in Enzyme Regulation*, 25:219–237, 1986. doi: 10.1016/0065-2571(86)90016-6.

R. A. Harris, M. M. Bowker-Kinley, P. Wu, J. Jeng, and K. M. Popov. Dihydrolipoamide Dehydrogenase-binding Protein of the Human Pyruvate Dehydrogenase Complex: DNA-derived Amino Acid Sequence, Expression, and Reconstitution of the Pyruvate Dehydrogenase Complex. *Journal of Biological Chemistry*, 272(32):19746–19751, 1997. doi: 10.1074/jbc.272.32.19746.

R. Heim, D. C. Prasher, and R. Y. Tsien. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America*, 91(26):12501–12504, 1994. doi: 10.1073/pnas.91.26.12501.

S. Hezaveh, Zeng A.-P., and U. Jandt. Investigation of Core Structure and Stability of Human Pyruvate Dehydrogenase Complex: A Coarse-Grained Approach. *ACS Omega*, 2(3):1134–1145, 2017. doi: 10.1021/acsomega.6b00386.

S. Hezaveh, A.-P. Zeng, and Jandt U. Full Enzyme Complex Simulation: Interactions in Human Pyruvate Dehydrogenase Complex. *Journal of Chemical Information and Modeling*, 26;58(2):362–369, 2018. doi: 10.1021/acs.jcim.7b00557.

J. Higgins. The Theory of Oscillating Reactions - Kinetics Symposium. *Industrial & Engineering Chemistry*, 59(5):18–62, 1967. doi: 10.1021/ie50689a006.

Y. Hiromasa, T. Fujisawa, Y. Aso, and T. E. Roche. Organization of the cores of the mammalian pyruvate dehydrogenase complex formed by E2 and E2 plus the E3-binding protein and their capacities to bind the E1 and E3 components. *The Journal of Biological Chemistry*, 279(8):6921–6933, 2004. doi: 10.1074/jbc.M308172200.

F. Hucho. The pyruvate dehydrogenase multienzyme complex. *Angewandte Chemie (International ed. in English)*, 14(9):591–601, 1975. doi: 10.1002/anie.197505911.

T. R. Hvidsten, A. Laegreid, A. Kryshtafovych, G. Andersson, K. Fidelis, and J. Komorowski. A comprehensive analysis of the structure-function relationship in proteins based on local structure similarity. *PLOS ONE*, 4(7):e6266, 2009. doi: 10.1371/journal.pone.0006266.

Inkwood Research. Global bio-based chemicals market forecast 2017-2025, 2017. URL `http://news.bio-based.eu/global-bio-based-chemicals-market-forecast-2017-2025/`.

Invitrogen. Expressway cell-free *E.Coli* expression system: Cell-free protein synthesis system for expression of recombinant protein, 27. September 2011. URL `https://t1p.de/l5e1`.

T. Izard, A. Ævarsson, M. D. Allen, A. H. Westphal, R. N. Perham, A. de Kok, and W. G. Hol. Principles of quasi-equivalence and euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes. *PNAS*, (96):1240–1245, 1999. doi: 10.1073/pnas.96.4.1240.

M. C. Jewett, K. A. Calhoun, A. Voloshin, J. J. Wuu, and J. R. Swartz. An integrated cell-free metabolic platform for protein production and synthetic biology. *Molecular Systems Biology*, 4:220, 2008. doi: 10.1038/msb.2008.57.

D. Jullesson, F. David, B. Pfleger, and J. Nielsen. Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnology Advances*, 33(7):1395–1402, 2015. doi: 10.1016/j.biotechadv.2015.02.011.

K. Kamil. Coding Deep Learning for Beginners — Linear Regression (Part 2): Cost Function, 2018. URL `https://t1p.de/qko5`.

K.-D. Kammeyer. *Nachrichtenübertragung*. Informationstechnik. Vieweg+Teubner Verlag and Imprint, Wiesbaden, 3., neubearbearbeite und ergänzte auflage edition, 2004. ISBN 978-3-322-94062-9.

F. Katzen, G. Chang, and W. Kudlicki. The past, present and future of cell-free protein synthesis. *Trends in Biotechnology*, 23(3):150–156, 2005. doi: 10.1016/j.tibtech.2005.01.003.

C.-J. Ke, Y.-H. He, H.-W. He, X. Yang, R. Li, and J. Yuan. A new spectrophotometric assay for measuring pyruvate dehydrogenase complex activity: A comparative evaluation. *Analytical Methods*, 6(16):6381, 2014. doi: 10.1039/C4AY00804A.

I. Khansahib. *Structural and functional investigations of the modified core components of human Pyruvate Dehydrogenase Complex*. Master thesis, Technical University Hamburg, Hamburg, 2018.

T. Kigawa, T. Yabuki, N. Matsuda, T. Matsuda, R. Nakajima, A. Tanaka, and S. Yokoyama. Preparation of *Escherichia coli* cell extract for highly productive cell-free protein expression. *Journal of Structural and Functional Genomics*, 5 (1-2):63–68, 2004. doi: 10.1023/B:JSFG.0000029204.57846.7d.

D.-M. Kim, T. Kigawa, C.-Y. Choi, and S. Yokoyama. A Highly Efficient Cell-Free Protein Synthesis System from *Escherichia coli*. *Eur. J. Biochem.*, 239: 881–886, 1996. doi: 10.1111/j.1432-1033.1996.0881u.x.

E. Klipp, W. Liebermeister, C. Wierling, and A. Kowald. *Systems biology: A textbook*. Wiley-VCH, Weinheim, 2nd, completely revised and enlarged ed. edition, 2016. ISBN 9783527336364.

A. Knepper, M. Heiser, F. Glauche, and P. Neubauer. Robotic platform for parallelized cultivation and monitoring of microbial growth parameters in microwell plates. *Journal of Laboratory Automation*, 19(6):593–601, 2014. doi: 10.1177/2211068214547231.

R. R. Kopito. Aggresomes, inclusion bodies and protein aggregation. *Trends in Cell Biology*, 10(12):524–530, 2000. doi: 10.1016/S0962-8924(00)01852-3.

L. G. Korotchkina and M. S. Patel. Binding of pyruvate dehydrogenase to the core of the human pyruvate dehydrogenase complex. *FEBS Letters*, 582(3): 468–472, 2008. doi: 10.1016/j.febslet.2007.12.041.

L. G. Korotchkina, M. M. Tucker, T. J. Thekkumkara, K. T. M., and M. S. Patel. Overexpression and characterization of human tetrametic pyruvat dehydrogenase and its individual subunits. *Protein Expression and Purification*, 6(1):79–90, 1995. doi: 10.1006/prep.1995.1011.

S. Kosuri and G. M. Church. Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods*, 11:499 EP –, 2014. ISSN 1548-7091. doi: 10.1038/nmeth.2918. URL `https://doi.org/10.1038/nmeth.2918`.

B. U. Kozak, H. M. van Rossum, M. A. H. Luttik, M. Akeroyd, K. R. Benjamin, L. Wu, S. de Vries, J.-M. Daran, J. T. Pronk, and A. J. A. van Maris. Engineering acetyl coenzyme A supply: functional expression of a bacterial pyruvate dehydrogenase complex in the cytosol of *Saccharomyces cerevisiae*. *mBio*, 5(5):e01696–14, 2014. doi: 10.1128/mBio.01696-14.

G.-J. Kremers, J. Goedhart, E. B. van Munster, and T. W. J. Gadella. Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Förster radius. *Biochemistry*, 45(21):6570–6580, 2006. doi: 10.1021/bi0516273.

N. Krinsky, M. Kaduri, J. Shainsky-Roitman., M. Goldfeder, E. Ivanir, I. Benhar, Y. Shoham, and A. Schroeder. A Simple and Rapid Method for Preparing a Cell-Free Bacterial Lysate for Protein Synthesis. *PLOS ONE*, 11(10):e0165137, 2016. doi: 10.1371/journal.pone.0165137.

K. Kruse and F. Jülicher. Oscillations in cell biology. *Current Opinion in Cell Biology*, 17(1):20–26, 2005. doi: 10.1016/j.ceb.2004.12.007.

J. E. Lawson, R. H. Behal, and L. J. Reed. Disruption and mutagenesis of the *Saccharomyces cerevisiae* PDX1 gene encoding the protein X component of the pyruvate dehydrogenase complex. *Biochemistry*, 30(11):2834–2839, 1991. doi: 10.1021/bi00225a015.

D. Lee, O. Redfern, and C. Orengo. Predicting protein function from sequence and structure. *Nature Reviews. Molecular cell biology*, 8:995–1005, 2007. doi: 10.1038/nrm2281.

G. Lee, J. Yoo, B. J. Leslie, and T. Ha. Single-molecule analysis reveals three phases of DNA degradation by an exonuclease. *Nature Chemical Biology*, 7(6): 367–374, 2011. doi: 10.1038/nchembio.561.

M. Lelimousin, M. Noirclerc-Savoye, C. Lazareno-Saez, B. Paetzold, S. Le Vot, R. Chazal, P. Macheboeuf, M. J. Field, D. Bourgeois, and A. Royant. Intrinsic dynamics in ECFP and Cerulean control fluorescence quantum yield. *Biochemistry*, 48(42):10038–10046, 2009. doi: 10.1021/bi901093w.

W. Leong, B. Nankervis, and J. Beltzer. Automation: what will the cell therapy laboratory of the future look like? *Cell and Gene Therapy Insights*, pages 679–694, 2018. doi: 10.18609/cgti.2018.067.

D. V. Liu, James F. Z., and J. R. Swartz. Streamlining *Escherichia coli* S30 extract preparation for economical cell-free protein synthesis. *Biotechnology Progress*, 21(2):460–465, 2005. doi: 10.1021/bp049789y.

W.-Q. Liu, L. Zhang, M. Chen, and J. Li. Cell-free protein synthesis: Recent advances in bacterial extract sources and expanded applications. *Biochemical Engineering Journal*, 141:182–189, 2019. doi: 10.1016/j.bej.2018.10.023.

S. S. Mande, S. Sarfaty, M. D. Allen, R. N. Perham, and W. G. Hol. Protein–protein interactions in the pyruvate dehydrogenase multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase. *Structure*, 4(3):277–286, 1996. doi: 10.1016/S0969-2126(96)00032-9.

A. Marquart. SPR-pages, 30.05.2020. URL `https://www.sprpages.nl/data-fitting/models`.

C. Marsac, D. Stansbie, G. Bonne, J. Cousin, P. Jehenson, C. Benelli, J-P Leroux, and G. Lindsay. Defect in the lipoyl-bearing protein X subunit of pyruvate dehydrogenase complex in two patients with encephalomyelopathy. *The Journal of Pediatrics*, 123(6):915–920, 1993.

R. Marshall, C. S. Maxwell, S. P. Collins, T. Jacobsen, M. L. Luo, M. B. Begemann, B. N. Gray, E. January, A. Singer, Y. He, C. L. Beisel, and V. Noireaux. Rapid and Scalable Characterization of CRISPR Technologies Using an *E. coli* Cell-Free Transcription-Translation System. *Molecular Cell*, 69(1):146–157.e3, 2018. doi: 10.1016/j.molcel.2017.12.007.

T. Matsuura, H. Yanagida, J. Ushioda, I. Urabe, and T. Yomo. Nascent chain, mRNA, and ribosome complexes generated by a pure translation system. *Biochemical and Biophysical Research Communications*, 352(2):372–377, 2007. doi: 10.1016/j.bbrc.2006.11.050.

K. Menzel, K. Ahrens, A.-P. Zeng, and W. D. Deckwer. Kinetic, dynamic, and pathway studies of glycerol metabolism by *Klebsiella pneumoniae* in anaerobic continuous culture: IV. Enzymes and fluxes of pyruvate metabolism. *Biotechnology and Bioengineering*, 60(5):617–626, 1998. doi: 10.1002/(sici)1097-0290(19980905)59:5⟨544::aid-bit3⟩3.0.co;2-a.

J. S. Miles, J. R. Guest, S. E. Radford, and R. N. Perham. Investigation of the mechanism of active site coupling in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli* by protein engineering. *Journal of Molecular Biology*, 202(1):97–106, 1988. doi: 10.1016/0022-2836(88)90522-0.

C. L. Mills, P. J. Beuning, and M. J. Ondrechen. Biochemical functional predictions for protein structures of unknown or uncertain function. *Computational and Structural Biotechnology Journal*, 13:182–191, 2015. doi: 10.1016/j.csbj.2015.02.003.

J. Modak, W.-D. Deckwer, and A.-P. Zeng. Metabolic control analysis of eucaryotic pyruvate dehydrogenase multienzyme complex. *Biotechnology Progress*, 18(6):1157–1169, 2002. doi: 10.1021/bp025613p.

K. C. Murphy. The lambda Gam protein inhibits RecBCD binding to dsDNA ends. *Journal of Molecular Biology*, 371(1):19–24, 2007. doi: 10.1016/j.jmb.2007.05.085.

A. Nasiri, M. Sadeghi, A. Vaisi-Raygani, S. Kiani, Z. Aghelan, and R. Khodarahmi. Emerging regulatory roles of mitochondrial sirtuins on pyruvate dehydrogenase complex and the related metabolic diseases: Review. *Biomedical Research and Therapy*, 7(2):3645–3658, 2020. doi: 10.15419/bmrat.v7i2.591.

D. L. Nelson, M. M. Cox, and A. L. Lehninger. *Lehninger principles of biochemistry, Chapter 16: The citric acid cycle*. W.H. Freeman, New York, NY, 5. ed., 5. print edition, 2008. ISBN 1429208929.

D. L. Nelson, A. L. Lehninger, M. M. Cox, M. Osgood, and K. Ocorr. *Lehninger principles of biochemistry*. W.H. Freeman, New York, 5. ed. edition, 2009. ISBN 9780716771081.

H. H. Nguyen, J. Park, S. J. Park, C.-S. Lee, S. Hwang, Y.-B. Shin, T. H. Ha, and M. Kim. Long-Term Stability and Integrity of Plasmid-Based DNA Data Storage. *Polymers*, 10(1), 2018. doi: 10.3390/polym10010028.

J. Nielsen. Synthetic biology for engineering acetyl coenzyme a metabolism in yeast. *mBio*, 5(6):e02153, 2014. doi: 10.1128/mBio.02153-14.

M. W. Nirenberg and J. H. Matthaei. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, 47:1588–1602, 1961. doi: 10.1073/pnas.47.10.1588.

H. Ohashi, T. Kanamori, Y. Shimizu, and T. Ueda. A Highly Controllable Reconstituted Cell-Free System - a Breakthrough in Protein Synthesis Research. *Current Pharmaceutical Biotechnolgy*, (11):267–271, 2010. doi: 10.2174/138920110791111889.

M. G. Ormerod. *Flow cytometry: A basic introduction*. Lulu Press, Raleigh, NC, 2008. ISBN 978-0-9559812-0-3.

S. S. Pai, T. M. Przybycien, and R. D. Tilton. Protein PEGylation attenuates adsorption and aggregation on a negatively charged and moderately hydrophobic polymer surface. *Langmuir : the ACS Journal of Surfaces and Colloids*, 26(23):18231–18238, 2010. doi: 10.1021/la102709y.

Pall ForteBio LLC. *Octet® Data Analysis High Throughput (HT) User Guide*, volume Release 10.0. 2017.

G. J. Palm, A. Zdanov, G. A. Gaitanaris, R. Stauber, G. N. Pavlakis, and A. Wlodawer. The structural basis for spectral variations in green fluorescent protein. *Nature Structural Biology*, 4(5):361 – 365, 1997. doi: 10.1038/nsb0597-361.

S. Panke and M. G. Wubbolts. Enzyme technology and bioprocess engineering. *Current Opinion in Biotechnology*, 13(2):111–116, 2002. doi: 10.1016/S0958-1669(02)00302-6.

J. W. Park and Y. M. Rhee. Interpolated mechanics-molecular mechanics study of internal rotation dynamics of the chromophore unit in blue fluorescent protein and its variants. *The Journal of Physical Chemistry. B*, 116(36):11137–11147, 2012. doi: 10.1021/jp306257t.

S. Park, J. H. Jeon, B. K. Min, C. M. Ha, T. Thoudam, B. Y. Park, and K. L., In. Role of the Pyruvate Dehydrogenase Complex in Metabolic Remodeling: Differential Pyruvate Dehydrogenase Complex Functions in Metabolism. *Diabetes & Metabolism Journal*, 42(4):270–281, 2018. doi: 10.4093/dmj.2018.0101.

S. W. Park, S. Kang, and T. S. Yoon. Crystal structure of the cyan fluorescent protein Cerulean-S175G. *Acta Crystallographica. Section F, Structural Biology Communications*, 72(Pt 7):516–522, 2016. doi: 10.1107/S2053230X16008311.

Y.-H. Park and M. S. Patel. Characterization of interactions of dihydrolipoamide dehydrogenase with its binding protein in the human pyruvate dehydrogenase complex. *Biochemical and Biophysical Research Communications*, 395(3): 416–419, 2010. doi: 10.1016/j.bbrc.2010.04.038.

M. S. Patel and Y. S. Hong. Lipoic acid as an antioxidant: the role of dihydrolipoamide dehydrogenase. *Methods in Molecular Biology*, 108, 1998. doi: 10.1385/0-89603-472-0:337.

M. S. Patel and L. G. Korotchkina. Regulation of mammalian pyruvate dehydrogenase complex by phosphorylation: complexity of multiple phosphorylation sites and kinases. *Experimental & Molecular Medicine*, 33(4):191–197, 2001. doi: 10.1038/emm.2001.32.

M. S. Patel and L. G. Korotchkina. Regulation of the pyruvate dehydrogenase complex. *Biochemical Society Transactions*, 34(Pt 2):217–222, 2006. doi: 10.1042/BST20060217.

M. S. Patel, L. G. Korotchkina, and S. Sidhu. Interaction of E1 and E3 components with the core proteins of the human pyruvate dehydrogenase complex. *Journal of Molecular Catalysis. B, Enzymatic*, 61(1-2):2–6, 2009. doi: 10.1016/j.molcatb.2009.05.001.

M. S. Patel, N. S. Nemeria, W. Furey, and F. Jordan. The pyruvate dehydrogenase complexes: structure-based function and regulation. *The Journal of Biological Chemistry*, 289(24):16615–16623, 2014. doi: 10.1074/jbc.R114.563148.

R. N. Perham. Swinging Arms and Swinging Domains in Multifunctional Enzymes: Catalytic Machines for Multistep Reactions. *Annu. Rev. Biochem.*, (69):961–1004, 2000. doi: 10.1146/annurev.biochem.69.1.961.

P. Perron-Savard, G. de Crescenzo, and H. Le Moual. Dimerization and DNA binding of the Salmonella enterica PhoP response regulator are phosphorylation independent. *Microbiology (Reading, England)*, 151(Pt 12):3979–3987, 2005. doi: 10.1099/mic.0.28236-0.

M. Porcar. Beyond directed evolution: Darwinian selection as a tool for synthetic biology. *Systems and Synthetic Biology*, 4(1):1–6, 2010. doi: 10.1007/s11693-009-9045-4.

J. L. Porter, R. A. Rusli, and D. L. Ollis. Directed Evolution of Enzymes for Industrial Biocatalysis. *ChemBioChem*, 17(3):197–203, 2016. doi: 10.1002/cbic.201500280.

V. Potapov and J. L. Ong. Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLOS ONE*, 12(1):e0169774, 2017. doi: 10.1371/journal.pone.0169774.

S. Prajapati, D. Haselbach, S. Wittig, M. S. Patel, A. Chari, C. Schmidt, H. Stark, and K. Tittmann. Structural and Functional Analyses of the Human PDH Complex Suggest a Division-of-Labor Mechanism by Local E1 and E3 Clusters. *Structure*, 27(7):1124–1136.e4, 2019. doi: 10.1016/j.str.2019.04.009.

K. Radhakrishnan and A. C. Hindmarsh. Description and Use of LSODE, the Livermore solver for ordinary differential equations. *NSA Reference Publication*, 1993. URL `https://ntrs.nasa.gov/search.jsp?R=19940030753`.

L. J. Reed. A trail of research from lipoic acid to alpha-keto acid dehydrogenase complexes. *The Journal of Biological Chemistry*, 276(42):38329–38336, 2001. doi: 10.1074/jbc.R100026200.

A. Rekas, J.-R. Alattia, T. Nagai, A. Miyawaki, and M. Ikura. Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity. *Journal of Biological Chemistry*, 277(52):50573–50578, 2002. doi: 10.1074/jbc.M209524200.

R. L. Rich and D. G. Myszka. Higher-throughput, label-free, real-time molecular interaction analysis. *Analytical Biochemistry*, 361(1):1–6, 2007. doi: 10.1016/j.ab.2006.10.040.

J. Rolf, K. Rosenthal, and S. Lütz. Application of Cell-Free Protein Synthesis for Faster Biocatalyst Development. *Catalysts*, 9(2):190, 2019. doi: 10.3390/catal9020190.

J.-M. Rouillard, W. Lee, G. Truan, X. Gao, X. Zhou, and E. Gulari. Gene2Oligo: oligonucleotide design for in vitro gene synthesis. *Nucleic Acids Research*, 32 (Web Server issue):W176–80, 2004. doi: 10.1093/nar/gkh401.

L. Rover, J. C.B. Fernandes, G. de Oliveira Neto, L. T. Kubote, E. Katekawa, and S.H.P. Serrano. Study of NADH Stability Using Ultraviolet Visible Spectrophotometric Analysis and Factorial Design. *Analytical Biochemistry*, 260:50–55, 1998. doi: 10.1006/abio.1998.2656.

Andrew Rubin and Galina Riznichenko, editors. *Mathematical Biophysics*. Biological and Medical Physics, Biomedical Engineering. Springer US, Boston, MA, 2014. ISBN 978-1-4614-8701-2. doi: 10.1007/978-1-4614-8702-9.

M. Rustad, A. Eastlund, P. Jardine, and V. Noireaux. Cell-free TXTL synthesis of infectious bacteriophage T4 in a single test tube reaction. *Synthetic Biology*, 3(1):253, 2018. doi: 10.1093/synbio/ysy002.

S. J. Sanderson, C. Miller, and J. G. Lindsay. Stoichiometry, Organisation and Catalytic Function of Protein X of the Pyruvate Dehydrogenase Complex from Bovine Heart. *Eur. J. Biochem.*, (236):69–77, 1996. doi: 10.1111/j.1432-1033.1996.00068.x.

G. Scheele and P. Blackburn. Role of mammalian RNase inhibitor in cell-free protein synthesis. *Proc. Natl. Acad. Sci.*, 76(10), 1979. doi: 10.1073/pnas.76.10.4898.

S.-M. Schinn, A. Broadbent, W. T. Bradley, and B. C. Bundy. Protein synthesis directly from PCR: progress and applications of cell-free protein synthesis with linear DNA. *New Biotechnology*, 33(4):480–487, 2016. doi: 10.1016/j.nbt.2016.04.002.

A. Schmid, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, and B. Witholt. Industrial biocatalysis today and tomorrow. *Nature*, 409(6817):258–268, 2001. doi: 10.1038/35051736. URL `https://doi.org/10.1038/35051736`.

D. Schwarz, F. Junge, F. Durst, N. Frölich, B. Schneider, S. Reckel, S. Sobhanifar, V. Dötsch, and F. Bernhard. Preparative scale expression of membrane proteins in *Escherichia coli*-based continuous exchange cell-free systems. *Nature Protocols*, 2(11):2945–2957, 2007. doi: 10.1038/nprot.2007.426.

V. A. Selivanov, D. T. Zakrzhevskaya, and B. N. Goldstein. Activity oscillations predicted for pyruvate dehydrogenase complexes. *FEBS Letters*, 345(2-3):151–153, 1994. doi: 10.1016/0014-5793(94)00425-0.

A. F. Sequeira, J. L. A. Brás, C. I. P. D. Guerreiro, R. Vincentelli, and C. M. G. A. Fontes. Development of a gene synthesis platform for the efficient large scale production of small genes encoding animal toxins. *BMC Biotechnology*, 16(1):86, 2016. doi: 10.1186/s12896-016-0316-3.

Y. Shimizu, I. Akio, T. Yukihide, S. Tsutomu, Y. Takashi, N. Kazuya, and T. Ueda. Cell-free translation reconstituted with purified components. *Nature Biotechnology*, (19):751–755, 2001. doi: 10.1038/90802.

Y. Shimizu, T. Kanamori, and T. Ueda. Protein synthesis by pure translation systems. *Methods (San Diego, Calif.)*, 36(3):299–304, 2005. doi: 10.1016/j.ymeth.2005.04.006.

J. Shin and V. Noireaux. Efficient cell-free expression with the endogenous *E. Coli* RNA polymerase and sigma factor 70. *Journal of Biological Engineering*, 4(1):8, 2010. doi: 10.1186/1754-1611-4-8.

J. Shin and V. Noireaux. An *E. coli* cell-free expression toolbox: application to synthetic gene circuits and artificial cells. *ACS Synthetic Biology*, 1(1):29–41, 2012. doi: 10.1021/sb200016s.

J. Siltberg-Liberles, J. A. Grahnen, and D. A. Liberles. The evolution of protein structures and structural ensembles under functional constraint. *Genes*, 2(4): 748–762, 2011. doi: 10.3390/genes2040748.

A. D. Silverman, A. Y. S. Karim, and M. C. Jewett. Cell-free gene expression: an expanded repertoire of applications. *Nature reviews. Genetics*, 21(3):151–170, 2020. doi: 10.1038/s41576-019-0186-3.

R. D. Sleator and P. Walsh. An overview of in silico protein function prediction. *Archives of Microbiology*, 192(3):151–155, 2010. doi: 10.1007/s00203-010-0549-9.

S. Spector, B. Kuhlman, R. Fairman, E. Wong, J. A. Boice, and D. P. Raleigh. Cooperative folding of a protein mini domain: the peripheral subunit-binding domain of the pyruvate dehydrogenase multienzyme complex. *Journal of Molecular Biology*, 276(2):479–489, 1998. doi: 10.1006/jmbi.1997.1522.

P. S. Spencer and J. M. Barral. Genetic code redundancy and its influence on the encoded polypeptides. *Computational and Structural Biotechnology Journal*, 1:e201204006, 2012. doi: 10.5936/csbj.201204006.

E. Spiess, F. Bestvater, A. Heckel-Pompey, K. Toth, M. Hacker, G. Stobrawa, T. Feurer, C. Wotzlaw, U. Berchner-Pfannschmidt, T. Porwol, and H. Acker. Two-photon excitation and emission spectra of the green fluorescent protein variants ECFP, EGFP and EYFP. *Journal of Microscopy*, (217):200–204, 2005. doi: 10.1111/j.1365-2818.2005.01437.x.

A. Spirin, V. I. Baranov, L. A. Ryabova, S. Ovodov, and Y. Alakhov. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science*, 242(4882):1162–1164, 1988. doi: 10.1126/science.3055301.

A. S. Spirin and J. R. Swartz, editors. *Cell-free protein synthesis: Methods and protocols*. Wiley-VCH, Weinheim, 2008. ISBN 978-3-527-31649-6. doi: 10.1002/9783527622702. URL `http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10302762`.

M. Stech, R. B. Quast, R. Sachse, C. Schulze, D. A. Wüstenhagen, and S. Kubick. A continuous-exchange cell-free protein synthesis system based on extracts from cultured insect cells. *PLOS ONE*, 9(5):e96635, 2014. doi: 10.1371/journal.pone.0096635.

T. H. Steinbrecher. *Experimentelle und modellgestützte Untersuchung des Bindungsverhaltens modifizierter Subkomponenten des Pyruvatdehydrogenasekomplexes*. Master thesis, Technical University Hamburg, Hamburg, 2019.

K. Steiner and H. Schwab. Recent advances in rational approaches for enzyme engineering. *Computational and Structural Biotechnology Journal*, 2:e201209010, 2012. doi: 10.5936/csbj.201209010.

W. P. C. Stemmer, A. Crameri, K. D. Ha, T. M. Brennan, and H. L. Heyneker. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*, 164(1):49–53, 1995. doi: 10.1016/0378-1119(95)00511-4.

J. Stetefeld, S. A. McKenna, and T. R. Patel. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophysical Reviews*, 8 (4):409–427, 2016. doi: 10.1007/s12551-016-0218-6.

S. Strumiło. Short-term regulation of the mammalian pyruvate dehydrogenase complex. *Acta Biochimica Polonica*, 52(4):759–764, 2005.

B. Sümegi and I. Alkonyi. Elementary steps in the reaction of the pyruvate dehydrogenase complex from pig heart. *Eur. J. Biochem.*, (136):347–353, 1983. doi: 10.1111/j.1432-1033.1983.tb07748.x.

Z. Z. Sun, C. A. Hayes, J. Shin, F. Caschera, R. M. Murray, and V. Noireaux. Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *Journal of Visualized Experiments*, (79): e50762, 2013. doi: 10.3791/50762.

Z. Z. Sun, E. Yeung, C. A. Hayes, V. Noireaux, and R. M. Murray. Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synthetic Biology*, 3(6):387–397, 2014. doi: 10.1021/sb400131a.

R. Szatanek, M. Baj-Krzyworzeka, J. Zimoch, M. Lekka, M. Siedlar, and J. Baran. The Methods of Choice for Extracellular Vesicles (EVs) Characterization. *International Journal of Molecular Sciences*, 18(6), 2017. doi: 10.3390/ijms18061153.

P. Teerawanichpan, T. Hoffman, P. Ashe, R. Datla, and G. Selvaraj. Investigations of combinations of mutations in the jellyfish green fluorescent protein (GFP) that afford brighter fluorescence, and use of a version (VisGreen) in plant, bacterial, and animal cells. *Biochimica et Biophysica Acta*, 1770(9):1360–1368, 2007. doi: 10.1016/j.bbagen.2007.06.005.

J. R. TerMaat. *De novo gene synthesis by rapid polymerase chain reaction assemblycoupled with immunoaffinity purification: A novel process and workstation*. Chemical & Biomolecular Engineering Theses, Dissertations, & Student Research. 8, 2011.

R. Tobias and S. Kumaraswamy. Biomolecular Binding Kinetics on Octet: Application Note 14. *ForteBio*, 2013.

S. Topell, J. Hennecke, and R. Glockshuber. Circularly permuted variants of the green fluorescent protein. *FEBS Letters*, 457(2):283–289, 1999. doi: 10.1016/S0014-5793(99)01044-3.

R. Y. Tsien. The green fluorescent protein. *Annu. Rev. Biochem.*, (67):509–544, 1998. doi: 10.1146/annurev.biochem.67.1.509.

C. Tuckey, H. Asahara, Y. Zhou, and S. Chong. Protein synthesis using a reconstituted cell-free system. *Current Protocols in Molecular Biology*, 108: 16.31.1–22, 2014. doi: 10.1002/0471142727.mb1631s108.

T. C. Vessoni Penna, M. Ishii, Luciana Cambricoli de Souza, and O. Cholewa. Expression of green fluorescent protein (GFPuv) in *Escherichia coli* DH5-a, under different growth conditions. *African Journal of Biotechnology*, (3 (1)): 105–111, 2004. doi: 10.5897/AJB2004.000-2019.

S. Vijayakrishnan, S. M. Kelly, R. J. C. Gilbert, P. Callow, D. Bhella, T. Forsyth, J. G. Lindsay, and O. Byron. Solution structure and characterisation of the human pyruvate dehydrogenase complex core assembly. *Journal of Molecular Biology*, 399(1):71–93, 2010. doi: 10.1016/j.jmb.2010.03.043.

J. Wallner, G. Lhota, D. Jeschek, A. Mader, and K. Vorauer-Uhl. Application of Bio-Layer Interferometry for the analysis of protein/liposome interactions. *Journal of Pharmaceutical and Biomedical Analysis*, 72:150–154, 2013. doi: 10.1016/j.jpba.2012.10.008.

C. Walter. Oscillation in Enzyme Reactions. *Nature*, 209(5021):404–405, 1966. doi: 10.1038/209404b0.

W. Wang, A. A. Ignatius, and S. V. Thakkar. Impact of residual impurities and contaminants on protein stability. *Journal of Pharmaceutical Sciences*, 103(5): 1315–1330, 2014. doi: 10.1002/jps.23931.

Y. Wang, J. Chen, and Z. Kang. In Silico Protein Design Promotes the Rapid Evolution of Industrial Enzymes. *Biochemistry*, 58(11):1451–1453, 2019. doi: 10.1021/acs.biochem.8b00896.

R. M. Watcher, M.-A. Elsliger, K. Kallio, G. T. Hanson, and S. J. Remington. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure*, (6):1267–1277, 1998. doi: 10.1016/S0969-2126(98)00127-0.

W. Wei, H. Li, N. Nemeria, and F. Jordan. Expression and purification of the dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase subunits of the *Escherichia coli* pyruvate dehydrogenase multienzyme complex: A mass spectrometric assay for reductive acetylation of dihydrolipoamide acetyltransferase. *Protein Expression and Purification*, 28(1):140–150, 2003. doi: 10.1016/S1046-5928(02)00674-5.

O. H. Wieland. The mammalian pyruvate dehydrogenase complex: Structure and regulation. In R. H. Adrian, H. Zur Hausen, E. Helmreich, H. Holzer, R. Jung, O. Krayer, R. J. Linden, P. A. Miescher, J. Piiper, H. Rasmussen, U. Trendelenburg, K. Ullrich, W. Vogt, and A. Weber, editors, *Reviews of Physiology, Biochemistry and Pharmacology, Volume 94*, pages 123–170. Springer Berlin Heidelberg, Berlin, Heidelberg, 1982. ISBN 978-3-540-11701-8. doi: 10.1007/BFb0031008.

M. J. Wilkerson. Principles and applications of flow cytometry and cell sorting in companion animal medicine. *The Veterinary Clinics of North America. Small Animal Practice*, 42(1):53–71, 2012. doi: 10.1016/j.cvsm.2011.09.012.

F. H. C. Wong, D. S. Banks, A. Abu-Arish, and C. Fradin. A molecular thermometer based on fluorescent protein blinking. *Journal of the American Chemical Society*, 129(34):10302–10303, 2007. doi: 10.1021/ja0715905.

J. M. Woodley. Bioprocess intensification for the effective production of chemical products. *Computers Chemical Engineering*, (105):297–307, 2017. doi: 10.1016/j.compchemeng.2017.01.015.

M. Wurm, S. Ilhan, U. Jandt, and A.-P. Zeng. Direct and highly sensitive measurement of fluorescent molecules in bulk solutions using flow cytometry. *Analytical Biochemistry*, 2019. doi: 10.1016/j.ab.2019.01.006.

Z.-L. Xiu, A.-P. Zeng, and W.-D. Deckwer. Multiplicity and stability analysis of microorganisms in continuous culture: Effects of metabolic overflow and growth inhibition. *Biotechnology and Bioengineering*, 57(3):251–261, 1998. doi: 10.1002/(SICI)1097-0290(19980205)57:3⟨251::AID-BIT1⟩3.0.CO;2-G.

T. Yabuki, Y. Motoda, K. Hanada, E. Nunokawa, M. Saito, E. Seki, M. Inoue, T. Kigawa, and S. Yokoyama. A robust two-step PCR method of template DNA production for high-throughput cell-free protein synthesis. *Journal of Structural and Functional Genomics*, 8(4):173–191, 2007. doi: 10.1007/s10969-007-9038-z.

D. Yang, J. Song, T. Wagenknecht, and T. E. Roche. Assembly and Full Functionality of Recombinantly Expressed Dihydrolipoyl Acetyltransferase Component of the Human Pyruvate Dehydrogenase Complex. *Journal of Biological Chemistry*, 272(10):6361–6369, 1997. doi: 10.1074/jbc.272.10.6361.

X. Yu, Y. Hiromasa, H. Tsen, J. K. Stoops, T. E. Roche, and Z. H. Zhou. Structures of the Human Pyruvate Dehydrogenase Complex Cores: A Highly Conserved Catalytic Center with Flexible N-Terminal Domains. *Structure*, 16 (1):104–114, 2008. doi: 10.1016/j.str.2007.10.024.

A.-P. Zeng, J. Modak, and W.-D. Deckwer. Nonlinear dynamics of eucaryotic pyruvate dehydrogenase multienzyme complex: decarboxylation rate, oscillations, and multiplicity. *Biotechnology Progress*, 18(6):1265–1276, 2002. doi: 10.1021/bp020097.

G. Zhang, V. Gurtu, and S. R. Kain. An Enhanced Green Fluorscent Protein Allows Sensitive Detection of Gene Transfer in Mammalian Cells. *Biochemical and Biophysical Research Communicatons*, (227):707–711, 1996. doi: 10.1006/bbrc.1996.1573.

N. J. Zuidam. *Encapsulation Technologies for Active Food Ingredients and Food Processing. Chapter: Bioprocess Intensification of Beer Fermentation Using Immobilised Cells* . Springer Science+Business Media, LLC, New York, NY, 2010. ISBN 978-1-4419-1008-0.

List of Tables

List of Figures

Appendix A

A.1 Additional data to materials

In the following additional data is shown to complement the results presented in chapter 3 to chapter 6.

A.1.1 Sequences of proteins used in this work

In the following sequences used for this thesis will be listed. Thereby, the eGFP positive control sequence for TX-TL reactions as well as the regulatory parts of this positive control, a OR2-OR1-PR and UTR1 sequence are presented. Furthermore, the GFPuv sequence used for calibration curve of GFPuv in crude lysate and also used for LET generation via two-step PCR:

OR2-OR1-PR-UTR1 sequence used in this work:
TGAGCTAACACCGTGCGTGTTGACAATTTTACCTCTGGCGGTGATAATGGTTGC
AAATAATTTTGTTTAACTTTAAGAAGGAGATATA

eGFP sequence used in this work:
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGA
GCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGG
GCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGC
TGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCT
TCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGC
CCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACA
AGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT
ACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCA
TCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCG
CCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCA
TGGACGAGCTGTACAAGTAA

GFPuv sequence used in this work:
ATGGCTAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAA
TTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGT
GATGCTACATACGGAAAGCTTACCCTTAAATTTATTTGCACTACTGGAAAACTA
CCTGTTCCATGGCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTTTT
CCCGTTATCCGGATCATATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCG
AAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGAACTACAAGA
CGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAA
AAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACACAAACTCGAGTACA
ACTATAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCA
AAGCTAACTTCAAAATTCGCCACAACATTGAAGATGGATCCGTTCAACTAGCAG
ACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAA
CCATTACCTGTCGACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGTGA
CCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGA
TGAGCTCTACAAATAA

A.1.2 Primers and oligomers used in this work

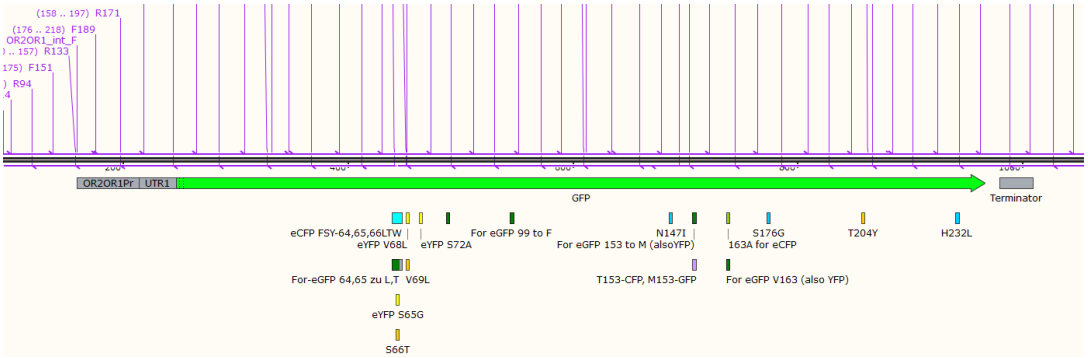


Figure A.1: Part of forward and reverse 56 oligomers (purple) for the GFPuv LET assembly, representation in SnapGene. To the left and right non-coding sequence overhangs are added, flanked by regulatory parts (grey) and the fluorescent protein (green). Indicated are some mutation sites for wavelength shift of the resulting fluorescent protein.

Table A.1: Oligomers designed for the LET production of GFPuv and variants via a two-step PCR as presented in section 4.4.1.

name	sequence	origin
F0	gggatcttgtggttCTGAATGAAGCCATACCAAACGACG	gene2oligo
R21	GCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTT	gene2oligo
F39	AGCGTGACACCACGATGCCTGCAGCAATGGCAACAA	gene2oligo
R57	GTTAATAGTTTGCGCAACGTTGTTGCCATTGCTGCAG	gene2oligo
F75	CGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAG	gene2oligo
R94	ATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCA	gene2oligo
F114	CTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGA	gene2oligo
R133	GATCGATCTCGATCCTCTACTCCGCCTCCATCCAGTCT	gene2oligo
F151	GTAGAGGATCGAGATCGATCTCTGAGCTAACACCGTGC	gene2oligo
R171	CCAGAGGTAAAATTGTCAACACGCACGGTGTTAGCTCAGA	gene2oligo
F189	GTGTTGACAATTTTACCTCTGGCGGTGATAATGGTTGCAAATA	gene2oligo
R211	ATCTCCTTCTTAAAGTTAAACAAAATTATTTGCAACCATTATCACCG	gene2oligo
F232	ATTTTGTTTAACTTTAAGAAGGAGATATAATGGCTAGCAAAGGAGAA	gene2oligo
R258	ACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTAT	gene2oligo
F279	GAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATG	gene2oligo
R299	TTTGTGCCCATTAACATCACCATCTAATTCAACAAGAATTGGG	gene2oligo
F322	GTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGG	gene2oligo
R342	GTATGTAGCATCACCTTCACCCTCTCCACTGACAGAAAA	gene2oligo
F361	GTGAAGGTGATGCTACATACGGAAAGCTTACCCTTAAATTTATTT	gene2oligo
R381	AGGTAGTTTTCCAGTAGTGCAAATAAATTTAAGGGTAAGCTTTCC	gene2oligo
F406	GCACTACTGGAAAACTACCTGTTCCATGGCCAACACTT	gene2oligo
R426	CACCATAAGAGAAAGTAGTGACAAGTGTTGGCCATGGAAC	gene2oligo
F444	GTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCCCGTTATC	gene2oligo
R466	GCCGTTTCATATGATCCGGATAACGGGAAAAGCATTGAA	gene2oligo
F487	CGGATCATATGAAACGGCATGACTTTTTCAAGAGTGCC	gene2oligo
R505	GTACATAACCTTCGGGCATGGCACTCTTGAAAAAGTCAT	gene2oligo
F525	ATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCA	gene2oligo
R544	CTTGTAGTTCCCGTCATCTTTGAAAGATATAGTGCGTTCCT	gene2oligo
F565	AAGATGACGGGAACTACAAGACGCGTGCTGAAGTCAAG	gene2oligo
R585	ACAAGGGTATCACCTTCAAACTTGACTTCAGCACGCGT	gene2oligo
F603	TTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATT	gene2oligo
R623	GAATGTTTCCATCTTCTTTAAAATCAATACCTTTTAACTCGATACGATTA	gene2oligo
F648	GATTTTAAAGAAGATGGAAACATTCTCGGACACAAACTCGAGTA	gene2oligo
R673	GTATACATTGTGTGAGTTATAGTTGTACTCGAGTTTGTGTCCGA	gene2oligo
F692	CAACTATAACTCACACAATGTATACATCACGGCAGACAAACAA	gene2oligo
R717	AAGTTAGCTTTGATTCCATTCTTTTGTTTGTCTGCCGTGAT	gene2oligo
F735	AAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACATTG	gene2oligo
R758	CTAGTTGAACGGATCCATCTTCAATGTTGTGGCGAATTTTG	gene2oligo
F778	AAGATGGATCCGTTCAACTAGCAGACCATTATCAACAAAATACTC	gene2oligo
R799	CAGGGCCATCGCCAATTGGAGTATTTTGTTGATAATGGTCTG	gene2oligo
F823	CAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCAT	gene2oligo
R841	CAGATTGTGTCGACAGGTAATGGTTGTCTGGTAAAAGGA	gene2oligo
F861	TACCTGTCGACACAATCTGCCCTTTCGAAAGATCCCA	gene2oligo
R880	TGTGGTCACGCTTTTCGTTGGGATCTTTCGAAAGGG	gene2oligo
F898	ACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAAC	gene2oligo
R916	ATGTGTAATCCCAGCAGCAGTTACAAACTCAAGAAGGACCA	gene2oligo
F938	TGCTGCTGGGATTACACATGGCATGGATGAGCTCTACA	gene2oligo
R957	CGGGCTTTGCTAGGTTAATTATTATTTGTAGAGCTCATCCATGCC	gene2oligo
F976	AATAATAATTAACCTAGCAAAGCCCGCCGAAAGGCGGGCTTTTC	gene2oligo
R1002	GCTAGTTCCTCCTTTCAGACAGAAAAGCCCGCCTTTCGG	gene2oligo
F1020	TGTCTGAAAGGAGGAACTAGCCGGGCAAGAGCAACTCGG	gene2oligo
R1041	GAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCG	gene2oligo
F1059	TCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGT	gene2oligo
R1077	GCTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGA	gene2oligo
F1097	ACTCACCAGTCACAGAAGCATCTTACGGATGGCATGAC	gene2oligo
R1116	GCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGAT	gene2oligo
F692-N147I	CAACTATATCTCACACAATGTATACATCACGGCAGACAAACAA	in this work
R673-N147I	GTATACATTGTGTGAGATATAGTTGTACTCGAGTTTGTGTCCGA	in this work
F778-S176G	AAGATGGAGGCGTTCAACTAGCAGACCATTATCAACAAAATACTC	in this work
R758-S176G	CTAGTTGAACGCCTCCATCTTCAATGTTGTGGCGAATTTTG	in this work
F938-H232L	TGCTGCTGGGATTACACTGGGCATGGATGAGCTCTACA	in this work
R916-H232L	CAGTGTAATCCCAGCAGCAGTTACAAACTCAAGAAGGACCA	in this work
F861-T204Y	TACCTGTCGTATCAATCTGCCCTTTCGAAAGATCCCA	in this work
R841-T204Y	CAGATTGATACGACAGGTAATGGTTGTCTGGTAAAAGGA	in this work
F444-S66T	GTCACTACTTTCACCTATGGTGTTCAATGCTTTTCCCGTTATC	in this work
R426-S66T	CACCATAGGTGAAAGTAGTGACAAGTGTTGGCCATGGAAC	in this work
F735-A163V	AAGAATGGAATCAAAGTTAACTTCAAAATTCGCCACAACATTG	in this work
F692-T153M	CAACTATAACTCACACAATGTATACATCATGGCAGACAAACAA	in this work
R717-A163V,T153M	AAGTTAACTTTGATTCCATTCTTTTGTTTGTCTGCCATGAT	in this work
F444-F64L,S65T	GTCACTACTTTGACCTATGGTGTTCAATGCTTTTCCCGTTATC	in this work
R426-F64L,S65T	CACCATAGGTCAAAGTAGTGACAAGTGTTGGCCATGGAAC	in this work
F487-R80Q	CGGATCATATGAAACAGCATGACTTTTTCAAGAGTGCC	in this work
R466-R80Q	GCTGTTTCATATGATCCGGATAACGGGAAAAGCATTGAA	in this work
F444-S65G,V68L,S72A	GTCACTACTTTCGGTTATGGTCTTCAATGCTTTGCCCGTTATC	in this work
R426-S65G	GACCATAACCGAAAGTAGTGACAAGTGTTGGCCATGGAAC	in this work
R466-S72A,V68L	GCCGTTTCATATGATCCGGATAACGGGCAAAGCATTGAA	in this work
F444-F64L,S65T,Y66W	GTCACTACTTTGACCTGGGGTGTTCAATGCTTTTCCCGTTATC	in this work
R426-F64L,S65T,Y66W	CACCCCAGGTCAAAGTAGTGACAAGTGTTGGCCATGGAAC	in this work

Table A.3: List of primers used for the construction of N-terminal minimized core variants and TEV protease recognition site. These primers were designed by Khansahib, 2018 and Steinbrecher, 2019.

Primer	Sequence (5‘ to 3‘)
E3BPΔN220-F	GGCACCTTTACCGAAATTCCGGCAA
E3BPΔN220-R	ATCACCGTGATGATGATGATGATGGG
E2ΔN312-F	TACCGGTCCGGGTATGGCA
E2ΔN312-R	ATGCGGAGGCAGGCTCAT
NcoI-E3BP	GATATACCATGGGTCCGATTAAAATCCTGATGCCG
E3BP-His-SalI	AAGCTTGTCGACCTACtagTGATGATGATGATGATGGCTGCTG CCCAGATTGGCTTTAAAGCTTTTCAGAAAACG
E3BPΔN217-R	GCCCATGGTATATCTCCTTCTTAAAGTTAAAC
E3BPΔN77Val-lipoyl-F	GTTGAAGAAGGTGAAGATTGGAAACATGTC
E3BPΔN128ARG-F	CGTCTGAGTCCGGCAGCAC
E3BPΔN173Gly-F	GGTAAAATTACCGAAAGCCGTCC
E3BPΔN192Ser-F	AGTCCGCTGCAGGCAACC
E2ΔN129Ser-F	AGCTATCCGCCTCACATGCAG
E2ΔN216Ser-F	AGCGCCTTTGCAGATTATCGTCC
E2ΔN242Ala-F	GCAGCAGTTCCGCCTACA
E2ΔN266Pro-F	CCGAAAGGTCGTGTTTTTGTtagTCCG
E1-TEV-F	GAAAACCTGTACTTCCAGGGCTTTGCAAATGATGCCACCTTCG
E1-TEV-R	GTGGTGATGATGATGATGCATATG
E3-TEV-F	GAAAACCTGTACTTCCAGGGCGCAGATCAGCCGATTGATGC
E3-TEV-R	GCCGCTGCTGTGATGATGA

A.2 Preliminary work and challenges concerning hPDC

A.2.1 Coexpression of E2-E3BP and E1α-E1β

When coexpressig and purifying hPDC subcomponents E1α-E1β or E2-E3BP, it is observed that one protein is usually more prominent than the other. The E1α subunit is prominent over the E1β subunit, for the second construct, E3BP is prominent over E2 subunit (as can be seen in 5.2). The His-tag in the used constructs are located at the more prominent proteins E1α and E3BP. The method of co-purification was used to ensure only interacting subunits are purified, which might hold true for the interaction partner without His-tag, but for the His-tagged partner probably also non-functional and interacting proteins are purified and thus shift the balance of the two interaction partners. Not only the unbalanced ratio appears as problem when it comes to assemble the hPDC in the right stoichiometry, but also it influences the general yield of the coexpressed proteins, which is low compared to the E3 subcomponent.
On gene level another problem arises. Although the Duet pET family is constructed for protein coexpression (Novagen 2017), over time the loss of one gene on the plasmid is observed. Namely this is for E1β and E3BP. This was verified

by sequencing and by plasmid digestion experiments. Sequencing with primers that should cover the missing part failed, sequencing for other "stable" parts worked out nicely. The sequencing company Seqlab GmbH that uses Sanger sequencing also explains this phenomena with mixed population that leads to an unclear sequence result. Different set ups for restriction enzymes also revealed a mixed population with two different constructs. Their sizes fit to the full construct and also the construct with the specific missing gene (see Figure A.2). Two restriction enzymes that cut the plasmid only once were used. The resulting pattern on the agarose gel was expected to be two fragments. Instead, a pattern with three fragments is obtained, whereas there is one prominent fragment and two more faint ones. The prominent fragment corresponded to the backbone part which is stable. The more faint bands correspond to the multiple cloning sites (MCS) parts with the gene insert, here two different sizes were obtained. This also explains why the intensity of this bands are lower. The obtained fragment sizes matched the full constructs and also the only pETDuet-1-N6HisE1α or pETDuet-1-E2. The backbone for both constructs is pETDuet-1, so the cloning work and analysis can be done with the same enzymes in both cases. Figure A.2 only shows mixed population for the E2-E3BP construct, for the E1α-E1β construct it is the same, the gene in MCS1 (E1β) is missing.
This results were in addition verified by colony PCR of an ON culture. The cells were plated on selection LB-agar plates and single colonies were picked. Some of them showed bands on the agarose gel which correspond to the full size construct, some bands correspond to the construct with one missing gene, see Figure A.2 B.

In order to regulate gene expression better and avoid loss of a gene, the cryo stock should be tested and refreshed regularly. For the loss of E3BP, which contains the His-tag, no proteins can be purified via IMAC anymore. This reduces the overall yield for a mixed population cultivation. For missing E1β still protein can be purified, but the protein is inactive because of the missing interaction partner. The loss of one gene is observed over time during storage at -80 °C. Although DNA storage on plasmid is reported to be highly stable over many years (Nguyen et al., 2018), in this case the loss of the first gene in the plasmid is observed. Another possibility to avoid the pressure of a coexpression is to separate the coexpression of two genes on two different plasmids. The idea is to co-transform the two plasmids with one gene on each into one expression host and perform protein expression, in this way a simultaneous folding of both proteins can be ensured by circumventing drawbacks of coexpression from one plasmid. A co-transformation need to be successful to determine whether yield

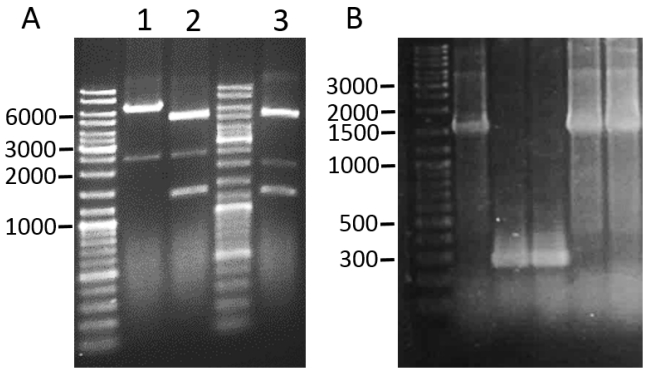


Figure A.2: A: agarose gel of E1α-E1β construct digested with different restriction enzymes. obtained fragment sizes digested by following restriction enzymes: 1= XbaI and SalI → 2.5 and 6.5 kb. 2 = NcoI and XhoI → 1.4, 2.5 and 5.5 kb. 3 = NotI and XhoI → 1.3, 2.0 and 5.5 kb. B: colony PCR of E2-E3BP construct. Single colonies obtained from an ON culture are investigated. In case of missing E3BP 270 bp band results, in case of present E3BP 1550 bp band results.

and performance of proteins are impaired. In case of E1α-E1β, coexpression and purification of both subunits is crucial for its functionality (Korotchkina et al., 1995).

A.2.2 hPDC subcomponents batch to batch variability

Protein purification of E1α-E1β and E2-E3BP was performed regularly because only low amount of proteins were obtained from one batch which was not sufficient for all experiments. Thereby differences between batches were revealed. Although the expression and purification was executed in the same way like described in material and method section 3.4.3, target protein purity of each batch varied. To test the reproducibility of the hPDC purification and activity of the hPDC subcomponents, an activity assay was set up. As internal positive control PDC purchased from Sigma Aldrich was used. This is roughly purified PDC from porcine heart (see first lane of Figure A.3). Because only total protein concentration is known, 5 µL are used for a hPDC activity measurement.
Specific activity was calculated for the reactions, therefore triplicates reaction were employed and the initial enzyme slope within the first six minutes was analyzed. Reaction 1 (internal standard) exhibited an enzyme activity of 1.53 nmol/min or specific activity of 20.1 mU/mg (by assuming 78 mg protein in 5 µL). Here the ratio of the subcomponents show a very prominent band for E2 (visible around 70 kDa, see Figure A.3). In contrary to the self-constituted hPDCs (lanes two to four) exhibit only a faint E2 band. Reaction 2 (reconstituted hPDC)

exhibits an enzyme activity of 1.32 nmol/min which translates to a specific activity of 60.2 mU/mg. The other self-constituted hPDC (reactions 3 purple and 4 green) show lower enzyme activity, although from the SDS-PAGE (Figure A.3) it can be seen that the protein purity and amount is very similar. This assay was performed multiple times, whereas a range for the for self-constituted hPDC from 20 to 78 mU/mg is measured. Some batches are more active than others, although they are treated the same. In addition it is challenging to identify the bad subcomponent since the complex is always assembled out of three different protein batches. For lane 1 some impurity in the lower kDa range is visible, this could represent host cell proteins which stick to the protein of interest and may have stabilizing effects. Furthermore, the aliquot size and the storage time at -80 °C also influence activity. This shows the variation range and importance of an internal control. In addition also the conditions for the activity assay may influence the protein performance, here all chemicals and cofactors should be stored accordingly to manufacturer's recommendations and treated carefully.

Table A.4: Three indepentdent hPDC activity assays as described in section 3.5.3. Two different batches of each hPDC subcomponent were used (E1α-E1β, E2-E3BP and E3) to reconsitute different hPDCs (rxn2, 3 and 4). The different batches are named after therit final protein cocentration in mg/mL, and for each rxn it is indicated via an x which batcg was employed. aAlos the resulting activity is indicated.

	E1α-E1β		E2-E3BP		E3		activity
mg/mL	5.7	1.0	15.0	1.2	4.4	8.2	mU/mg
rxn 2	x		x		x		60.2 ± 3.1
rxn 3		x		x	x		12.9 ± 0.7
rxn 4		x		x		x	18.1 ± 0.7

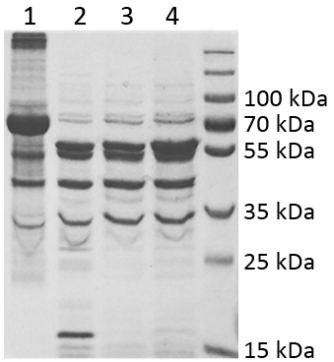


Figure A.3: SDS-PAGE of hPDCs derived from different purification batches and reconstituted for activity assay measurements. Indicated number over lanes match rxn number of Table. Rxn 2 equals lane 2 and rxn in the graph above so on. Lane 1 PDH purchased from sigma. Lane 2,3 and 4 different reconstituted hPDCs from different purified batches, lane 5 is the marker.

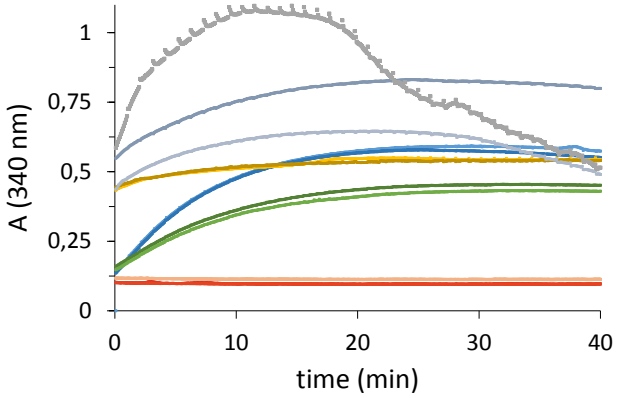


Figure 1.4: Overall PDC activity assay. Track of absorption of NADH formation at 340 nm at 37 °C. Blue: duplicate of PC (5 µL PDH), green: duplicate self constituted hPDC, light yellow: E1 and E3 pure, E2-E3BP from NEB kit, dark yellow: E1 and E3 pure, DHFR from NEB kit as NC, grey: triplicate TX-TL lysate, red: no substrate NC, pink: no enzyme NC.

1.2.3 hPDC activity assays with CF proteins and crude buffer background

As described in section 4.5.3, experiments were launched to test CF produced hPDC subcomponents in an overall activity assay. Furthermore, the crude reaction buffer background was tested in different compositions as shown in the following paragraphs.

1.2.3.1 Activity assay of CF produced hPDC subcomponents

The overall PDC activity assay was performed with only E2-E3BP produced by PURExpress® In Vitro Protein Synthesis Kit from NEB, the other subcomponents remained conventionally produced in cells. The results of this assay are presented in Figure 1.4. For comparison reason the internal standard reactions with 5 µL porcine heart PDC from Sigma Aldrich (blue) and the self-constituted PDC (green) are shown as well. The reaction indicated in light yellow shows the activity curve for a PDC that consist of conventionally produced E1α-E1β and E3, the E2-E3BP component was produced in a cell free manner by using the NEB kit. The dark yellow reaction is its negative control. Here, the E2-E3BP component is replaced by the DHFR positive control of the kit since it is assumed that this protein is not interacting with PDC subcomponents. Both curves start at a quite high zero absorbance of around 0.45 compared to the purified PDC or negative control curves (red, around 0.1). This is probably to side reactions of the

protein rich background of crude E2-E3BP. Also, both curves show a flat slope within the first minutes, which could indicate an enzyme activity. Because the negative control overlays with the mixed reconstituted PDC, the slope originates from unspecific protein reactions. A flat slope was expected because E2-E3BP amount is lower than used in a conventional assay. In this case the content of a 65 µL CFPS reaction was used after concentration with ultracentrifugal unit to 50 µL, the amount of E2-E3BP is roughly estimated to around 1.3 µg (usually 6.79 µg). The activity assay does not allow to add more than 10 % of reaction volume as protein content, so there is an upper limit of volume of subcomponents to add (which was exceeded in this assay by a factor of 2.5 already). In addition, it is not possible to measure the exact content of target protein in the crude background, so it is difficult to assign an accurate target protein amount.
In addition, E2-E3BP produced by TX-TL reactions were measured. Because no E2-E3BP genes were available with a constitutive promoter, T7 RNA polymerase was added to the CFPS reaction. The dark grey slope is supplemented with 8 U T7 RNA polymerase, light grey slope with 50 U of T7 RNA polymerase. The negative control reaction consists of an empty plasmid TX-TL reaction (grey). Although no target protein is visible on the SDS-PAGE (data not shown) this set up was performed. The signal of the TX-TL negative reaction is unusual compared to other curves from this assay. There seems to be an interference with the absorbance measurement. It is thus not possible to subtract probable background reactions from the TX-TL produced PDC subcomponent graphs. Furthermore, the used E2-E3BP amount was only a fraction (if present at all) of the standard PDC activity assay procedure, so activity resulting from this set up is unlikely. With this experiment, the functionality of PDC subcomponents produced in a cell free manner cannot be demonstrated. Firstly, the enzyme amount is not sufficient due to low protein yield of CFPS and the crude background produces noise which needs to be filtered out with proper negative controls. Hence, the crude buffer background interference was investigated in more detail in the next section.

1.2.3.2 Crude enzyme activity assays

PDC subcomponents were expressed and processed to different purity stages from crude protein lysate to pure target protein. With these proteins different set ups for activity assay measurements were performed which are presented in Figure 1.5. The internal standard (grey) exhibited an activity of 1.03 nmol/min or 13.5 mU/mg, the self-constituted PDC (purple) of 1.17 nmol/min or 53 mU/mg for a linear range within the first five minutes. Figure 1.5 B and C represent PDC assembled from crude protein lysate. In B, all subcomponents are mixed in crude form (green graphs), in C only the E2-E3BP subcomponent is in crude form, the

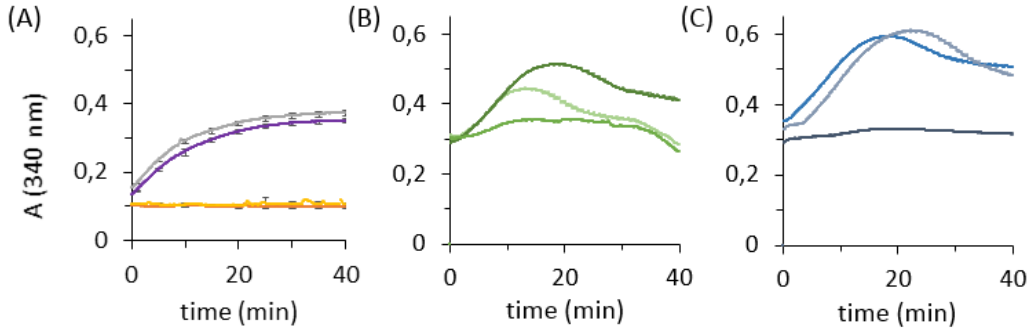


Figure 1.5: Overall PDC activity assay. Track of absorption of NADH formation at 340 nm at 37 °C. (A): activity assay measurement of PC (grey), self-constituted hPDC (purple) and NCs no substrate with PC (orange) and no enzyme (yellow) with pyruvate. Error bars show triplicates. (B): triplicate of hPDC complex: E1α-E1β and E3 subcomponents pure, E2-E3BP subcomponent crude. (C): triplicate of crude hPDC complex. (B) and (C) were plotted separately and are to be compared to the pure WT hPDC (purple). Also no error bars for (B) and (C) are shown, but instead the triplicate to visualize the variance of the replicates.

other subcomponents remain pure (grey/blue curves). For both set ups an initial lag phase is visible, moreover the triplicates differ strongly. When determining the linear slopes for set up B and C (from 5 to 10 min), the resulting slope equals the internal standard, although E2-E3BP amount in the crude mixtures is clearly lower. Also, there is no difference if the crude lysate background is diluted (B to C). For this reasons it can be assumed that the calculated activity results from background rather than PDC activity. Also the triplicates vary strongly, so reproducibility cannot be ensured. Therefore, in the following the influence of crude enzyme is investigated in more detail.

In this last experiment the influence of crude background in different dilutions is tested, see Figure 1.6. The left graph shows the influence of only one crude component (E2-E3BP) and other subcomponents being pure. In contrast, the right sided graph shows crude hPDC. The addition of only crude E2-E3BP subcomponent is investigated with different set ups. The orange graph consists of PDC that is reconstituted with the usual amounts of pure E1α-E1β and E3, the remaining volume for enzyme to add to the assay was filled with crude E2-E3BP. The setup of the green curve equals forgoing, but only half amount of crude E2-E3BP was used, the second half was replaced by crude cell lysate without target protein content. Finally a set up with complete replacement of the E2-E3BP subcomponent by crude lysate (yellow curve) was performed. The different E2-E3BP contents result in different curve shapes which is proportional with activity. The orange curve shows a typical active enzyme slope. With

a steep beginning and a saturation after some time. The triplicates overlay well so the error bars remain small. The same curve with only half amount of E2-E3BP shows a curve that runs faster into saturation and also does not reach the adsorption signal of foregoing. In addition the triplicates vary from minute 20 and so the reproducibility is a problem in that case. The yellow curve with no E2-E3BP shows a short slope at the beginning, but remains unstable until the end. Only crude background (dark blue slope) indicates a lag phase for 15 min, afterwards shows a steep slope with high error bars also. Depending on the protein content the initial adsorption signal is shifted, but the curve shapes can be distinguished for different set ups. Enzyme activities cannot be calculated, because exact target protein content is not known, a relative comparison can nevertheless be made. Qualitative differences are seen for different E2-E3BP amount, which must result from actual enzyme activity, but overlaid with an interference signal because green and orange curves exhibit the same initial slope. The fast saturation might be the result of fast energy and cofactors depletion since background reactions also utilize them. The right graph shows the total enzyme complex in crude background. Again, since target protein concentration is not determinable, equivolume of each subcomponent was used. Two different concentrations of crude constituted PDC are used: black slope with maximum possible crude content and purple slope with half PDC and half crude lysate background content. Both slopes show a lag phase and a subsequent slope with big variations in the triplicates. The half crude PDC slope (purple) in comparison to double amount of PDC shows a shorter lag phase but also a lower adsorption signal. The reproducibility and significance of crude PDC is not given, the signal is noisy. One big difficulty in keeping the set up comparable and reproducible is the unassigned target enzyme amount in the crude lysate. The actual amount of E2-E3BP is likely multiple times less than compared to a conventional overall PDC assay set up. However, an enzyme concentration dependent difference when only one hPDC subcomponent in crude lysate is employed is detectable, but not for the case of each hPDC subcomponent in crude background.

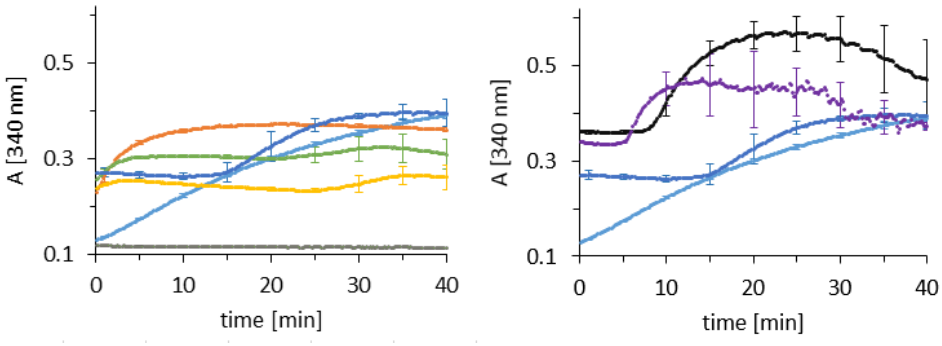


Figure 1.6: Overall hPDC activity assay. Crude enzyme is tested in different set ups. Left represents only E2-E3BP subcomponent in crude form, other subcomponents remain in pure form. Dark blue curves represent an enzyme lysate without target protein, so background reactions. Light blue is the conventional, self-constituted pure hPDC activity. Yellow curve is the negative control to orange (2 x E2-E3BP) and green (1 x E2-E3BP), because the E2-E3BP subcomponent is replaced by a crude lysate background. Right shows all hPDC subcomponents in crude form, with double (black) and half (purple) amount of target protein content.

1.3 Additional data to nonlinearities in hPDC kinetics

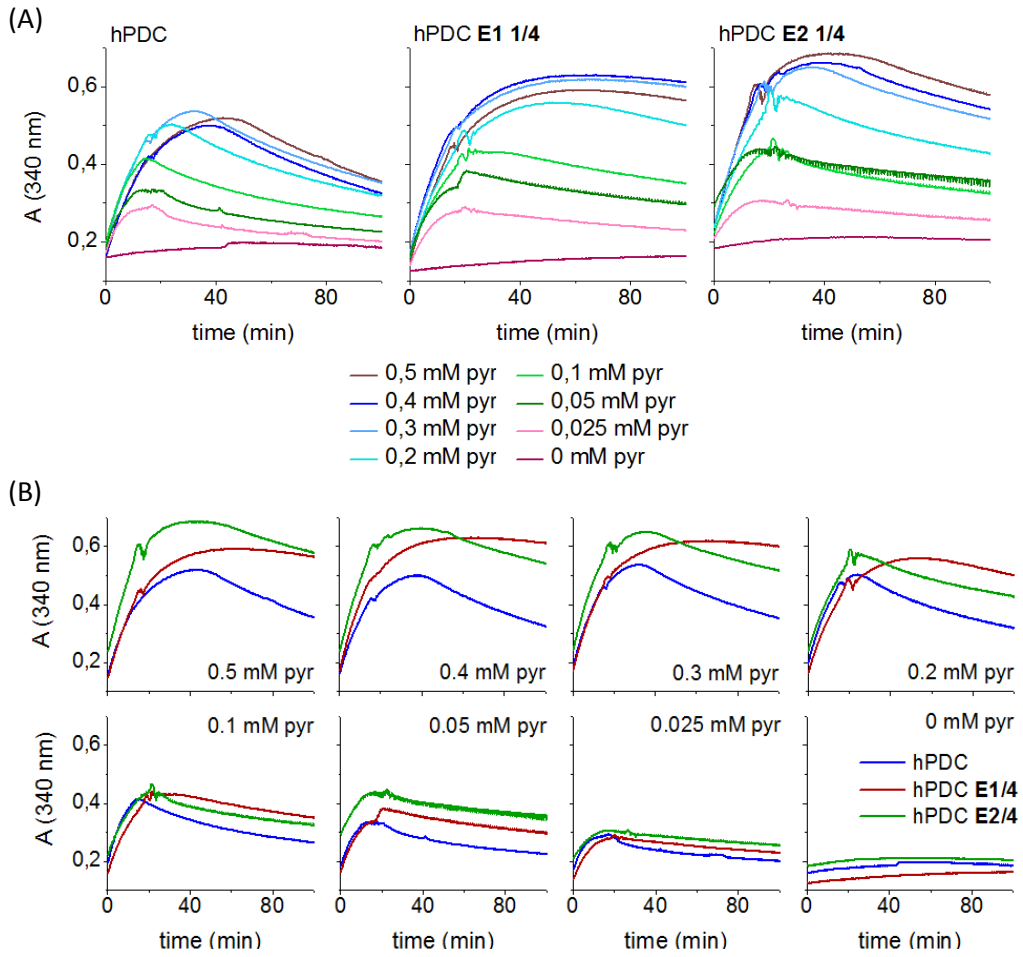


Figure 1.7: hPDC activity assay curves with varying E1 or E2-E3BP amount over a dilution series of pyruvate. This is a replicate to results presented in section 5.3. (A) shows data sorted according to subcomponent variation. (B) represents same data sorted according to pyruvate concentration as depicted in the graphs.

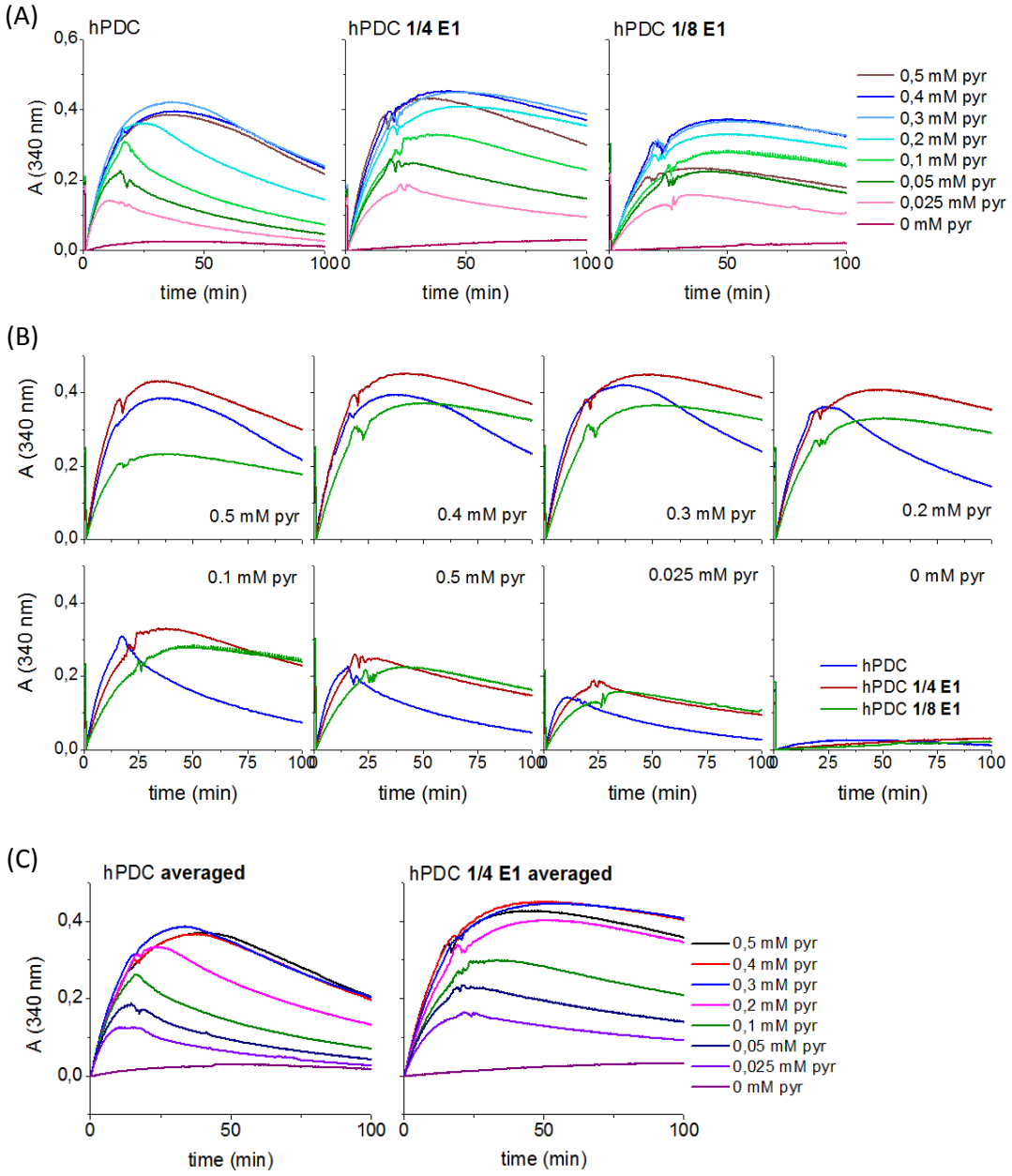


Figure 1.8: (A) and (B): Activity assay of hPDC with decreased amount of E1α-E1β by ¼ and ⅛. Start points were normalized to Y=0. Shown is A (340 nm) over time course of different pyruvate concentration. (A) sorted according to pyruvate dilution, (B) same data sorted according to E1 dilution. (C): Overlaid and normalized activity assay curves with usual hPDC ratio and hPDC with ¼ of E1 amount. Here, the data of (A) is overlaid with a replicate presented in Figure 5.4. The transient oscillation in enzyme activity is still visible.

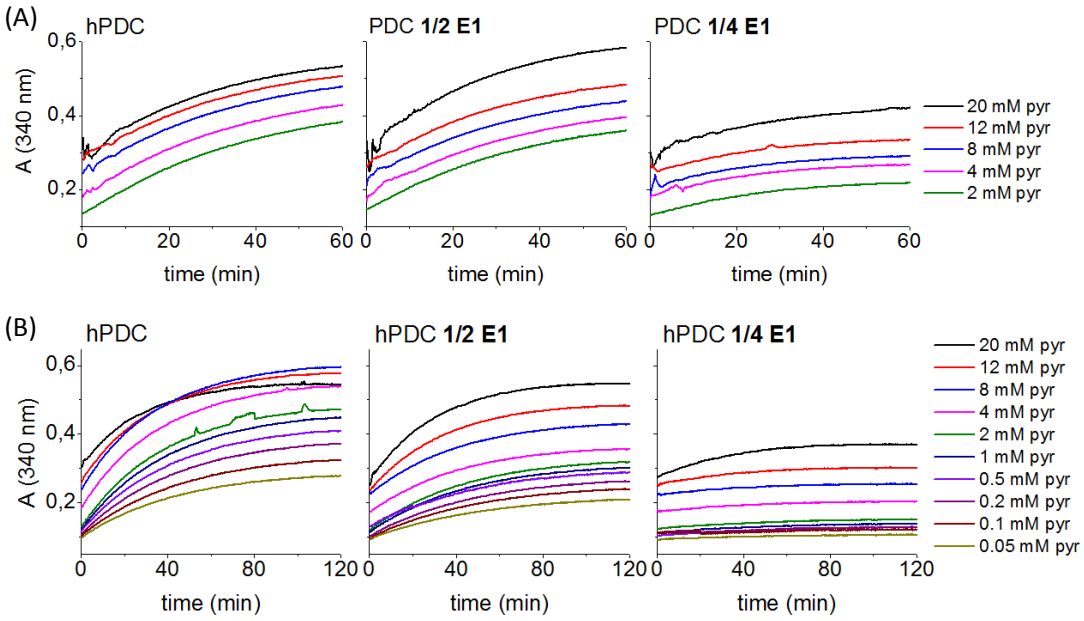


Figure 1.9: hPDC activity assay curves with ¼ of usual hPDC amount (5.66 µg total hPDC). Additionally E1 subcomponent was diluted as depicted above the graph. Pyruvate was added in different excess concentrations as indicated in the legend. (A) and (B) represent independent replicates. A delay in enzyme activity is visible for pyruvate concentration above 8 mM in (A), but not in (B). This data is a replicate and complements section 5.4.1.

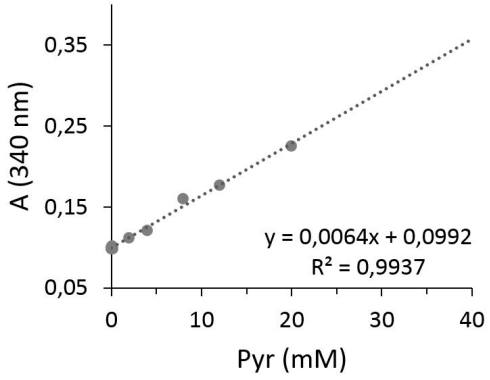


Figure 1.10: Pyruvate calibration curve. Different concentrations of pyruvate were spiked into hPDC activity assay buffer to record the baseline level of elevated pyruvate concentrations. A trend-line until 40 mM pyruvate is indicated by a dotted line.

Table 1.5: Parameters and their nominal values for hPDC modeling. Data gathered from Modak et al. (2002). Data marked with an (*) are from this study.

parameter unit	nominal value
$k1_{absolute}$ (s^{-1})	14.66
k2 (g $nmol^{-1}$ s^{-1})	21.6
k3 (s^{-1})	7
k4 (g $nmol^{-1}$ s^{-1})	16.2
k5 (g $nmol^{-1}$ s^{-1})	21.6
k6 (s^{-1})	2
k7 (g $nnom^{-1}$ s^{-1})	0.11
k8 (s^{-1})	0.2
k9 (g $nmol^{-1}$ s^{-1})	0.14
k10 (s^{-1})	0.06
$[E1]_0$ (nmol g^{-1})	5.56
$[E2]_0$ (nmol g^{-1})	18.52
$[E3]_0$ (nmol g^{-1})	1.85
$[K]_0$ (nmol g^{-1})	1.85
$k1_{effective}$ (nmol L^{-1} s^{-1})	2.769 *
$NADH_{decay}$ (s^{-1})	0.01 *
pyr_{feed} (s^{-1})	0 *
concentration	(nmol L^{-1})
rate	(nmol L^{-1} s^{-1})

Table 1.6: Reactions for hPDC modeling. Data gathered from Zeng et al. (2002). Data marked with an (*) are from this study.

R1	E1_TPP + Pyr $\rightarrow E1 + NADH^*$
R2	E1 + E2LSS $\rightarrow E2 + E1\_TPP$
R3	E2 $\rightarrow E2LSSHH$
R4	E2LSSHH + E3 $\rightarrow E2LSS + E3$
R5	E1 + E3 $\rightarrow E3S2AcH + E1\_TPP$
R6	E3S2AcH $\rightarrow E3$
R7	K' $\rightarrow K; E3S2AcH$
R8	K $\rightarrow K'$
R9	E1 $\rightarrow E'; K$
R10	E' $\rightarrow E1$
$NADH_{decay}$	NADH $\rightarrow$ *

1.4 Additional data to structural analysis of hPDC

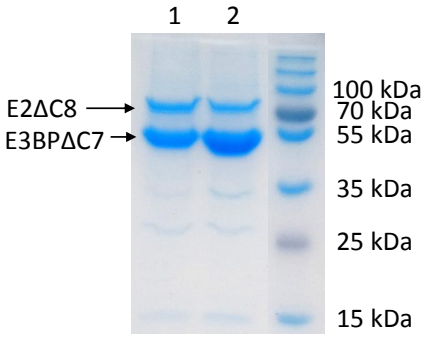


Figure 1.11: Overexpressed and His-tag purified hPDC core variants as described in section 6.2. E2ΔC8-E3BPΔC7 is presented in both lanes, whereas the His-tag for purification is located at the N-terminus of E3BP in lane 1, and shifted to the C-terminus of E3BP in lane 2. The N-terminal His-tagged protein (N-His-E2ΔC8-E3BPΔC7) was constructed and published by Guo et al. (2017), The C-terminal His-tagged protein (E2ΔC8-E3BPΔC7-C-His) corresponds to the core variant produced by Khansahib (2018). This picture is modified according to Khansahib (2018).

Table 1.7: Theoretical size, purity, yield and ratio of E2:E3BP protein expression. Values either refer to an average of duplicate or a single measurement. hPDC core variants are coexpressed from one vector and purified by His-tag located at E3BP. Thereby considerable fluctuations regarding purity, yield and ratio exist. Expression culture in each case was 400 mL. E2:E3BP ratio was judged via densitometric analysis of SDS-PAGE. Table according to Ilhan *et al*, submitted to Bioscience Reports, 2019.

Variant	Purity (%)	Yield (µg)	Ratio (E2:E3BP)
WT	78	159	0.14
$E2_{\Delta C8}$-$BP_{\Delta C7}$	70	333	0.19
$E2_{\Delta C8}$-$_{\Delta N77}$E3$BP_{\Delta C7}$	41	90	0.52
$E2_{\Delta C8}$-$_{\Delta N128}$E3$BP_{\Delta C7}$	53	91	0.6
$_{\Delta N123}E2_{\Delta C8}$-E3$BP_{\Delta C8}$	63	348	-
$_{\Delta N210}E2_{\Delta C8}$-E3$BP_{\Delta C8}$	67	415	0.18
$_{\Delta N236}E2_{\Delta C8}$-E3$BP_{\Delta C8}$	62	355	0.2
$_{\Delta N260}E2_{\Delta C8}$-E3$BP_{\Delta C8}$	67	338	0.15
$_{\Delta N123}E2_{\Delta C8}$-$_{\Delta N77}$E3$BP_{\Delta C8}$	63	367	0.97
$_{\Delta N210}E2_{\Delta C8}$-$_{\Delta N77}$E3$BP_{\Delta C8}$	54	149	-
$_{\Delta N236}E2_{\Delta C8}$-$_{\Delta N77}$E3$BP_{\Delta C8}$	49	145	0.53
$_{\Delta N260}E2_{\Delta C8}$-$_{\Delta N77}$E3$BP_{\Delta C8}$	40	158	0.66
$_{\Delta N123}E2_{\Delta C8}$-$_{\Delta N128}$E3$BP_{\Delta C8}$	53	208	0.64
$_{\Delta N210}E2_{\Delta C8}$-$_{\Delta N128}$E3$BP_{\Delta C8}$	57	231	1.06
$_{\Delta N236}E2_{\Delta C8}$-$_{\Delta N128}$E3$BP_{\Delta C8}$	48	289	-
$_{\Delta N260}E2_{\Delta C8}$-$_{\Delta N128}$E3$BP_{\Delta C8}$	46	204	-