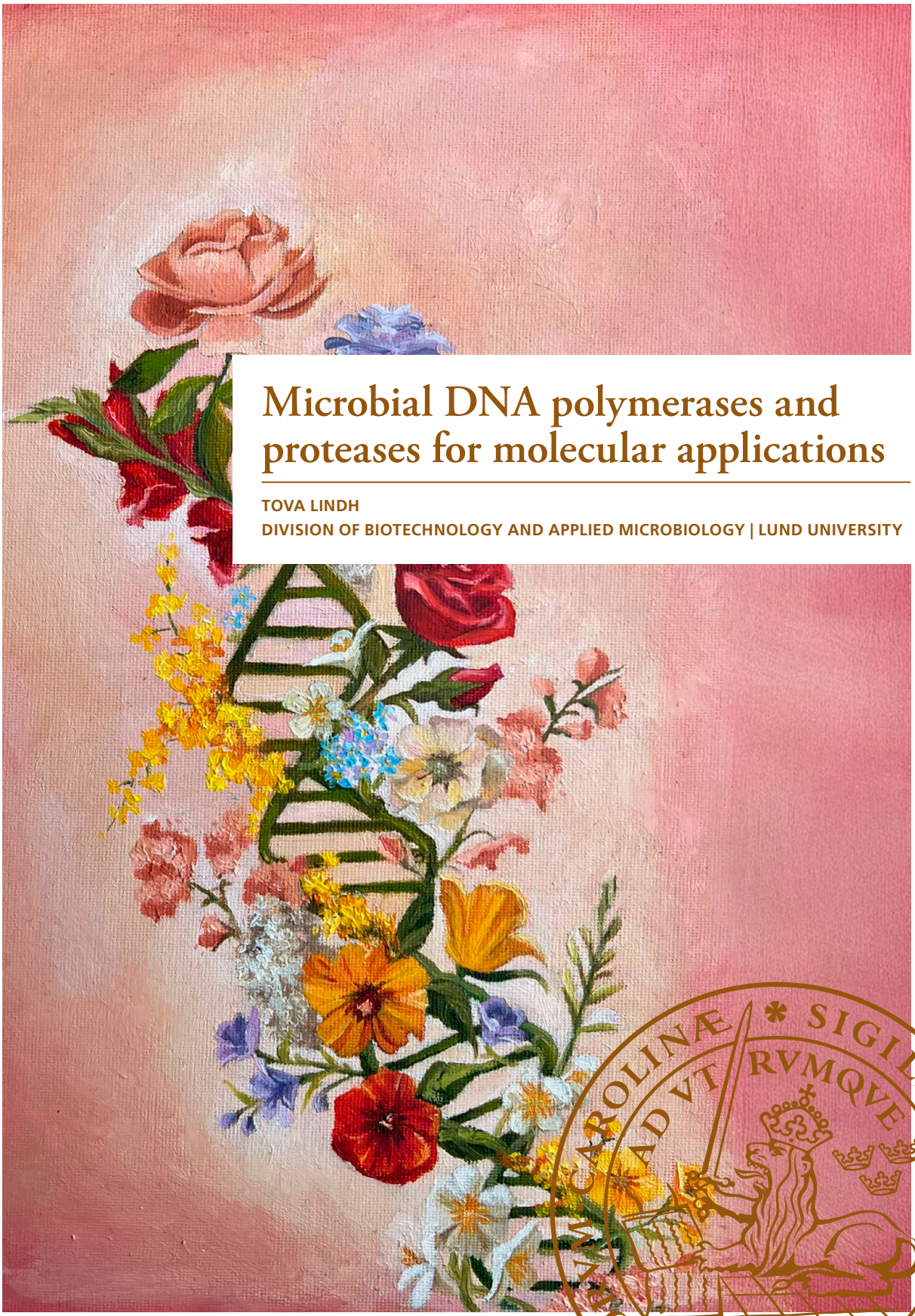
Microbial DNA polymerases and proteases for molecular applications

Lindh, Tova

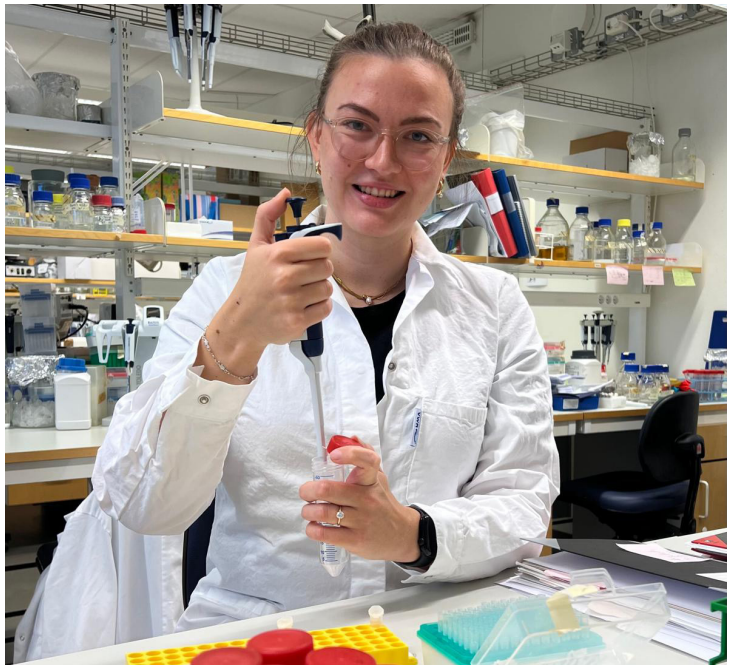
2025

Document Version:
Publisher's PDF, also known as Version of record



Microbial DNA polymerases and proteases for molecular applications

TOVA LINDH
DIVISION OF BIOTECHNOLOGY AND APPLIED MICROBIOLOGY | LUND UNIVERSITY



My PhD journey has truly been an adventure, and I've been fortunate to visit many places over the past few years. As someone who enjoys traveling, I've had the opportunity to attend international conferences, meetings, study visits, and even a research stay abroad. I've also been part of several groups where I got to perform lab work in different settings. No matter where I've been, my favorite place has always been the lab, regardless of its location.

Tova

Printed by Media-Tryck, Lund 2025 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Faculty of Engineering
Department of Process and Life Science Engineering
Division of Biotechnology and Applied Microbiology

ISBN 978-91-8096-142-4

Microbial DNA polymerases and proteases for molecular applications

Tova Lindh

LUND
UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Engineering at Lund University to be publicly defended on Friday 12th of December at 09.00 in Lecture Hall B, Department of Process and Life Science Engineering, Kemicentrum, Naturvetarvägen 22, Lund, Sweden

Faculty opponent
Dr. Jeppe Dyrberg Andersen

Department of Forensic Medicine, University of Copenhagen, Denmark

Organization: Lund University, Division of Biotechnology and Applied Microbiology

Document name: Doctoral Thesis **Date of issue:** 2025-12-12

Author(s): Tova Lindh **Sponsoring organization:**

Title and subtitle: Microbial DNA polymerases and proteases for molecular applications

Abstract: Enzymes serve as biological catalysts in a wide range of applications, from household products to industrial processes, research, diagnostics and forensic science. They accelerate reactions with high specificity under mild reaction conditions and are essential for life. Advances in recombinant DNA technology have enabled the use of microorganisms as efficient cell factories for production of heterologous enzymes. Proteases and DNA polymerases are important enzymes in biotechnology due to their roles in protein degradation and DNA synthesis, respectively. In this thesis work, I focused on challenges related to the production of microbial proteases and DNA polymerases. It involved exploring host interactions between yeast and recombinant bacterial proteases and developing a simplified protocol for in-house production of DNA polymerases. Additionally, this thesis also focused on characterizing polymerase performance, including error types and rates during in vitro DNA amplification.

Bacterial proteases were fused to Green Fluorescent Protein (GFP) and produced in the yeast Saccharomyces cerevisiae. Both plasmid-based and integration-based systems were evaluated for use in protease production. The protein production after induction with galactose was followed on a single-cell level using flow cytometry (FCM), enabled by the GFP fusion. The plasmid-based strains resulted in population heterogeneity with two subpopulations, where approximately half of the cells did not show fluorescence levels above the autofluorescence. On the other hand, the integration-based strains resulted in homogenous populations where almost all the cells produced GFP fusions that resulted in fluorescence above the autofluorescence level. IdeS, a cysteine protease with a unique substrate specificity towards human IgG, was successfully produced as detected by FCM, Western blot and an activity assay. BdpK, a broad-spectrum serine protease, was produced and fluorescence was detected. However, no protease activity was observed, likely due to the formation of inclusion bodies. SpeB, another cysteine protease with a broad substrate scope, posed a substantial fitness burden on the yeast cells, even before induction. This indicates leakiness of the promoter used. Compartmentalization of the proteases to the peroxisome was successful and could potentially be used as a strategy to protect the host cells from the proteolytic activity.

A simplified protocol for in-house production of DNA polymerases was developed, which only requires readily available laboratory reagents, for use in the event of supply shortage or disruptions. The recombinant polymerases were successfully produced in the Escherichia coli strain Rosetta (DE3) pLysS and purified using Immobilized Metal Affinity Chromatography (IMAC) and gravity flow. The produced DNA polymerases were then shown to be compatible with several molecular biology techniques, including PCR, colony PCR, quantitative PCR (qPCR) and library preparation for sequencing. Comparable performance was achieved across the polymerase variants produced.

The effects of polymerase characteristics on amplicon yield and PCR error formation in Short Tandem Repeat (STR) analysis, including single-base substitutions and stutter artefacts, were studied using the SiMSen-Seq library preparation method. Unique Molecular Identifiers (UMIs) were applied in the first barcoding PCR, which was then followed by a second PCR (adaptor PCR), enabling detailed analysis of polymerase-introduced errors. Six DNA polymerases with different characteristics were used. The SiMSen-Seq library preparation was successful for all polymerases when applied in the adaptor PCR. DNA polymerases with both proofreading activity and DNA-binding domains produced high STR amplicon yield and lower levels of base substitutions compared to DNA polymerases lacking these domains. Stutter levels were not clearly connected with proofreading or DNA-binding properties.

This thesis demonstrates that alternative microbial hosts and simplified production workflows can be efficiently used for the recombinant production of both proteases and DNA polymerases, with specific strategies to mitigate host stress and optimize yield. The simplified polymerase production protocol enables accessible and reliable enzyme synthesis using standard laboratory resources. Furthermore, detailed characterization of polymerase performance in STR analysis highlights the importance of enzyme properties in minimizing PCR errors.

Key words: DNA polymerase, protease, recombinant protein, GFP, flow cytometry, PCR, qPCR, sequencing, production protocol, *Saccharomyces cerevisiae*, BdpK, IdeS, SpeB, *Escherichia coli*, *Taq* polymerase, *Pfu* polymerase, Sso7d, fidelity, processivity, STR markers, base substitutions, stutter artefacts

Language: English **Number of pages**: 127

ISBN: 978-91-8096-142-4

Microbial DNA polymerases and proteases for molecular applications

Tova Lindh

LUND
UNIVERSITY

Cover art by Lovisa Jäghagen

Back cover photo by Fredrik Lund

Figures 1.1, 2.1, 2.2, 3.1, 3.2, 4.1, 4.2, 4.3, 4.4 and 4.7 are created in https://BioRender.com

Published by:

Division of Biotechnology and Applied Microbiology

Department of Process and Life Science Engineering

Faculty of Engineering

Lund University

Lund 2025

ISBN 978-91-8096-142-4 (print)

ISBN 978-91-8096-143-1 (electronic)

Printed in Sweden by Media-Tryck, Lund University,
Lund, 2025

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

Till Fredrik

Table of Contents

Abstract

Enzymes serve as biological catalysts in a wide range of applications, from household products to industrial processes, research, diagnostics and forensic science. They accelerate reactions with high specificity under mild reaction conditions and are essential for life. Advances in recombinant DNA technology have enabled the use of microorganisms as efficient cell factories for production of heterologous enzymes. Proteases and DNA polymerases are important enzymes in biotechnology due to their roles in protein degradation and DNA synthesis, respectively. In this thesis work, I focused on challenges related to the production of microbial proteases and DNA polymerases. It involved exploring host interactions between yeast and recombinant bacterial proteases and developing a simplified protocol for in-house production of DNA polymerases. Additionally, this thesis also focused on characterizing polymerase performance, including error types and rates during *in vitro* DNA amplification.

Bacterial proteases were fused to Green Fluorescent Protein (GFP) and produced in the yeast *Saccharomyces cerevisiae*. Both plasmid-based and integration-based systems were evaluated for use in protease production. The protein production after induction with galactose was followed on a single-cell level using flow cytometry (FCM), enabled by the GFP fusion. The plasmid-based strains resulted in population heterogeneity with two subpopulations, where approximately half of the cells did not show fluorescence levels above the autofluorescence. On the other hand, the integration-based strains resulted in homogenous populations where almost all the cells produced GFP fusions that resulted in fluorescence above the autofluorescence level. IdeS, a cysteine protease with a unique substrate specificity towards human IgG, was successfully produced as detected by FCM, Western blot and an activity assay. BdpK, a broad-spectrum serine protease, was produced and fluorescence was detected. However, no protease activity was observed, likely due to the formation of inclusion bodies. SpeB, another cysteine protease with a broad substrate scope, posed a substantial fitness burden on the yeast cells, even before induction. This indicates leakiness of the promoter used. Compartmentalization of the proteases to the peroxisome was successful and could potentially be used as a strategy to protect the host cells from the proteolytic activity.

A simplified protocol for in-house production of DNA polymerases was developed, which only requires readily available laboratory reagents, for use in the event of supply shortage or disruptions. The recombinant polymerases were successfully

produced in the *Escherichia coli* strain Rosetta (DE3) pLysS and purified using Immobilized Metal Affinity Chromatography (IMAC) and gravity flow. The produced DNA polymerases were then shown to be compatible with several molecular biology techniques, including PCR, colony PCR, quantitative PCR (qPCR) and library preparation for sequencing. Comparable performance was achieved across the polymerase variants produced.

The effects of polymerase characteristics on amplicon yield and PCR error formation in Short Tandem Repeat (STR) analysis, including single-base substitutions and stutter artefacts, were studied using the SiMSen-Seq library preparation method. Unique Molecular Identifiers (UMIs) were applied in the first barcoding PCR, which was then followed by a second PCR (adaptor PCR), enabling detailed analysis of polymerase-introduced errors. Six DNA polymerases with different characteristics were used. The SiMSen-Seq library preparation was successful for all polymerases when applied in the adaptor PCR. DNA polymerases with both proofreading activity and DNA-binding domains produced high STR amplicon yield and lower levels of base substitutions compared to DNA polymerases lacking these domains. Stutter levels were not clearly connected with proofreading or DNA-binding properties.

This thesis demonstrates that alternative microbial hosts and simplified production workflows can be efficiently used for the recombinant production of both proteases and DNA polymerases, with specific strategies to mitigate host stress and optimize yield. The simplified polymerase production protocol enables accessible and reliable enzyme synthesis using standard laboratory resources. Furthermore, detailed characterization of polymerase performance in STR analysis highlights the importance of enzyme properties in minimizing PCR errors.

Popular science summary

Microorganisms, like bacteria and yeast, are organisms that are too small to be clearly seen by the naked eye. Still, they are everywhere: in soil, water, on our bodies, and even in extreme environments like hot springs and deep-sea vents. Many of these microbes are incredibly useful to humans. Inside each microscopic cell there are thousands of enzymes, specialized proteins that speed up biochemical reactions with remarkable precision and efficiency. These enzymes turn cells into miniature factories, capable of producing, breaking down and transforming molecules that we can use. Enzymes are nature's molecular tools, and we can use them in a wide range of applications, from medicine to environmental science and biotechnology.

Think of a microbial cell as a production unit, a cell factory. Raw material or building blocks enter the cell and are processed by enzymes resulting in different products. We can program this natural machinery to specialize in producing certain proteins or chemicals by modifying the genes inside the cells. Different microbial cells can be used depending on what we want to produce. In this thesis, I developed and evaluated production strategies and applications for two important microbial enzymes, proteases and DNA polymerases, using cell factories.

Proteases and DNA polymerases are vital to maintain normal cell functions in all living organisms. We all use enzymes in our everyday life, often without realizing it. Proteases break down proteins into their building blocks, amino acids. In cells, this property is crucial for removal of damaged proteins, recycling of cellular components and as a defense mechanism against foreign proteins. Proteases are used in laundry detergents, in cheese making and in biotechnology they provide controlled cleaving of proteins. DNA polymerases are responsible for copying all the genetic material inside the cell and making sure that the copy is identical to the original. Polymerases can also be used outside of the cells, often in a technique called PCR. Maybe you recognize the word PCR and maybe you even took a PCR test to make sure you did not have Covid during the recent pandemic? PCR and DNA polymerases are used in forensics to analyze DNA samples, in paternity testing, and diagnostics to determine genetic variations, detect pathogens and in cancer screening, among others.

Production of these two microbial enzymes comes with some specific challenges. Proteases are tricky enzymes to produce since they cut other proteins into pieces,

and not all cell factories can handle this burden. *Saccharomyces cerevisiae*, or Baker's yeast, has traditionally been used by humans when baking bread, making wine or beer. It is a well-known microorganism, and several tools have been developed to genetically modify the yeast. With the use of the gene scissors, CRISPR/Cas9, genetic modifications can easily be done in *S. cerevisiae*. In this thesis, I evaluated Baker's yeast as an alternative cell factory to produce three bacterial proteases. To follow the protein production, I connected the proteases with a green fluorescent protein (GFP). This made it possible to measure how much fluorescence each cell produced, related to how much protease that was produced in each individual cell factory. One of the three proteases tested was successfully produced in yeast, one was produced but lacked activity and one severely affected the yeast viability.

DNA polymerases are generally produced in *Escherichia coli* using these bacteria as cell factories. When DNA polymerases are produced inside these cell factories, it is important that the final product is free from the cell's DNA. Otherwise, this DNA can interfere with the assay and can as an example lead to false positives, which gives us the wrong results. DNA polymerases are purified using highly sophisticated methods and instruments to remove cell debris and *E. coli* DNA. However, recent crises, such as the Covid pandemic, have highlighted the vulnerability of the supply chains, leading to shortage of critical reagents and enzymes. Therefore, I developed a simple polymerase production-chain that only requires materials that are commonly found in ordinary biotechnology laboratories that can be used in case of supply disruption. The production may be performed in 3 to 4 days and generates enough enzyme for 90 000 PCR reactions. The produced polymerases were successfully used in several different PCR-based DNA analysis methods, ranging from detection and quantification to sequencing.

We all leave DNA traces behind when handling items or touching surfaces, but to analyze it, we need to copy it to get sufficient amount of DNA for the analysis. The parts of our DNA that is analyzed for forensic DNA profiles are repetitive sequences, which makes it harder for the DNA polymerase to correctly copy the DNA. Since errors are introduced in the copies, this makes it extra complicated to analyze a sample from a crime scene, which might contain DNA from several different individuals.

I performed a study to evaluate the errors that the polymerases make during copying of these repetitive sequences. Six different polymerases were evaluated in a technique where all the individual starting DNA molecules are labeled with a unique barcode. This barcode is then copied together with the DNA sequence and all copies having the same barcode can then be grouped. There are two types of errors, random and systematic. The use of barcodes makes it possible to track how many errors were introduced and what type of error is the most common and then compare the different DNA polymerases to each other. I found that DNA polymerases that bind harder to the DNA and that can proofread the DNA copies performed better than

DNA polymerases without these characteristics, both in terms of how much DNA that was copied and in the random errors that were introduced. The systematic errors introduced were more even between the DNA polymerases. However, when creating these groups with the same barcode, the proportion of incorrect sequences was decreased, which has been shown previously.

Together, proteases and DNA polymerases exemplify how microbial enzymes, though invisible to the unaided eye, play a massive role in shaping the future of molecular science. Their natural efficiency and specificity make them indispensable tools to further understand and manipulate the building blocks of life. In this thesis, I showed how proteases and DNA polymerases can be produced in different cell factories using simple production methods and how the yeast cells were affected by the proteases. Additionally, DNA polymerases were evaluated in several PCR based assays and errors introduced by different polymerases were investigated, helping us understand how we can minimize these errors.

Populärvetenskaplig sammanfattning

Mikroorganismer, såsom bakterier och jäst, är organismer som är för små för att ses med blotta ögat. Trots detta finns de överallt: i jord, vatten, på våra kroppar och till och med i extrema miljöer så som heta källor och vulkaniska öppningar på havsbotten där varmt vatten strömmar ut. Många av dessa mikrober är användbara för oss människor. Inuti varje mikroskopisk cell finns tusentals enzymer, specialiserade proteiner som påskyndar biokemiska reaktioner med imponerande precision och effektivitet. Dessa enzymer gör cellerna till små fabriker som kan producera, bryta ner och omvandla molekyler och kemikalier som vi kan använda. Enzymer är naturens molekylära verktyg och används inom allt från medicin till miljövetenskap och bioteknik.

Tänk dig en mikrobiell cell som en produktionsenhet, en cellfabrik. Råmaterial eller byggstenar transporteras in i cellen och bearbetas av enzymer till olika produkter. Vi kan programmera cellernas naturliga maskineri genom att ändra generna i cellen så att den specialiserar sig på att producera specifika proteiner eller kemikalier. Olika mikrobiella celler används beroende på vad vi vill producera. I den här doktorsavhandlingen har jag utvecklat och utvärderat produktionsstrategier och tillämpningar för två viktiga mikrobiella enzymer, proteaser och DNA-polymeraser, med hjälp av cellfabriker.

Proteaser och DNA-polymeraser är avgörande för att upprätthålla normala cellfunktioner i alla levande organismer. Vi använder enzymer i vardagen, ofta utan att tänka på det. Proteaser bryter ner proteiner till deras byggstenar, aminosyror. I celler är detta viktigt för att ta bort skadade proteiner, återvinna cellkomponenter och som en försvarsmekanism mot främmande proteiner. Proteaser används i tvättmedel, vid osttillverkning och inom bioteknik för kontrollerad nedbrytning av proteiner. DNA-polymeraser kopierar allt genetiskt material i cellen och ser till att kopian är identisk med originalet. Polymeraser används även utanför celler, ofta i en teknik som kallas PCR. Du kanske känner igen ordet PCR, kanske tog du till och med ett PCR-test under pandemin? PCR och DNA-polymeraser används bland annat inom forensisk DNA analys, rättsmedicin, faderskapstester och inom diagnostik för bestämning av genetiska avvikelser, detektion av patogena mikroorganismer eller virus och inom cancerdiagnostik.

Att producera dessa två mikrobiella enzymer innebär dock vissa utmaningar. Proteaser är svåra att producera eftersom de bryter ner andra proteiner, vilket kan

leda till en stor belastning för cellfabriken. *Saccharomyces cerevisiae*, eller bakjäst, har länge använts vid bakning och framställning av vin och öl av människor. Det är en välstuderad mikroorganism och flera verktyg har utvecklats för att möjliggöra genetisk modifiering. Med gensaxen CRISPR/Cas9 kan genetiska förändringar enkelt göras i jäst. I detta arbete utvärderade jag jäst som alternativ cellfabrik för att producera tre bakteriella proteaser. För att följa proteinproduktionen kopplades proteaserna till ett grönt fluorescerande protein (GFP), vilket gjorde det möjligt att mäta hur mycket fluorescens varje cell producerade, ett mått på hur mycket proteas som producerades i varje enskild cellfabrik. Ett av proteaserna producerades framgångsrikt i jäst, det andra producerades men saknade aktivitet, och det tredje påverkade jästens livskraft negativt och kunde inte produceras.

DNA-polymeraser produceras vanligtvis i *Escherichia coli*, där bakterierna fungerar som cellfabriker. När polymeraser produceras är det viktigt att slutprodukten är fri från cellens eget DNA, eftersom detta kan störa analyser och till exempel ge falska positiva resultat. Polymeraser renas med avancerade metoder för att ta bort cellrester och DNA. Kriser som pandemin har de senaste åren visat hur känsliga våra leveranskedjor är, vilket lett till brist på viktiga reagenser och enzymer. Därför utvecklade jag en enkel produktionskedja för polymeraser som endast kräver material som finns i vanliga biotekniklaboratorier och som kan användas vid leveransstörningar. Produktionen tar 3–4 dagar och ger tillräckligt med enzym för cirka 90 000 PCR-reaktioner. De producerade polymeraserna användes framgångsrikt i flera PCR-baserade DNA-analyser, från detektion och kvantifiering till sekvensering.

Vi lämnar alla DNA-spår efter oss när vi hanterar föremål eller rör ytor, men för att analysera dem måste DNAt kopieras för att få tillräcklig mängd. De delar av vårt DNA som analyseras vid forensiska DNA profiler är repetitiva sekvenser, vilket gör det svårare för polymerasen att kopiera korrekt. Eftersom fel introduceras i kopiorna blir det extra komplicerat att analysera prover från brottsplatser, där DNA från flera olika individer kan förekomma.

Jag genomförde en studie för att utvärdera vilka fel polymeraser gör när de kopierar de här repetitiva sekvenserna. Sex olika polymeraser utvärderades med en teknik där varje ursprunglig DNA-molekyl märks med en unik streckkod. Denna streckkod kopieras tillsammans med DNA-sekvensen, och alla kopior med samma streckkod kan sedan grupperas. Det finns två typer av fel, slumpmässiga och systematiska. Streckkoderna gör det möjligt att spåra hur många fel som introducerats och vilken typ som är vanligast, och möjliggör jämförelser mellan polymeraserna. Jag fann att polymeraser med stark DNA-bindning och förmåga att korrekturläsa kopiorna presterade bättre än de utan dessa egenskaper, både vad gäller mängden kopierat DNA och antalet slumpmässiga fel. De systematiska felen var mer lika mellan polymeraserna. När grupper med samma streckkod skapades minskade andelen felaktiga sekvenser, vilket även visats tidigare.

Tillsammans visar proteaser och DNA-polymeraser hur mikrobiella enzymer, trots att de inte kan ses med blotta ögat, spelar en avgörande roll i utvecklingen av molekylär bioteknik. Deras naturliga effektivitet och specificitet gör dem till viktiga verktyg för att förstå och manipulera livets byggstenar. I detta arbete visade jag hur dessa enzymer kan produceras i olika cellfabriker med enkla produktionsvägar, hur jästceller påverkas av proteaser, samt hur DNA-polymeraser kan utvärderas i PCR-baserade analyser för att förstå och minimera fel.

List of papers

This thesis is based on the following research papers and manuscripts, which can be found at the end of the thesis and will be referred to by roman numerals throughout the text.

I. Expression of the bacterial enzyme IdeS using a GFP fusion in the yeast *Saccharomyces cerevisiae*

Lindh, T., Collin, M., Lood, R. & Carlquist, M. (2023).
Published in: P. Nordenfelt, & M. Collin (Eds.), *Bacterial pathogenesis: Methods and protocols* (2nd ed., pp. 131-146). Humana Press.
Doi: 10.1007/978-1-0716-3243-7_9

II. Functional insights from recombinant production of bacterial proteases in *Saccharomyces cerevisiae*

Lindh, T., Collin, M., Lood, R. & Carlquist, M. (2025).
Published in: *Microbial Cell Factories, 24*(1), Article 119.
Doi: 10.1186/s12934-025-02732-x

III. Production of DNA polymerases for advanced PCR applications using standard laboratory equipment

Lindh, T., Sidstedt, M., Mirow, M., Collin, M., Carlquist, M., & Hedman, J. *Manuscript.*

IV. DNA polymerase characteristics influence noise levels in sequencing of short tandem repeats

Lindh, T., Sidstedt, M., Kiesler, K. M., Vallone, P. M. & Hedman, J. *Manuscript.*

My contribution to the papers

I. I participated in the design of the study, and planned and performed all the experimental work. I performed the data analysis and visualization and wrote the manuscript.

II. I participated in the design of the study, and planned and performed all the experimental work. I performed the data analysis and visualization and wrote the manuscript.

III. I participated in the conceptualization and design of the study and planned all the experiments. I performed or supervised all the experimental work. I performed the data analysis and visualization and wrote the manuscript.

IV. I participated in the design of the study. I performed all the experimental work, except for the first sequencing run. I performed the data analysis and developed scripts for visualization, and I wrote the manuscript together with my supervisors.

Abbreviations

AE	Amplification efficiency
BdpK	Serine protease from *Bdellovibrio bacteriovorus*
β-ME	β-mercaptoethanol
bp	Base pair
BSA	Bovine serum albumin
CE	Capillary electrophoresis
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats and the Cas9 nuclease
Cq	Cycle of quantification
DNase	Deoxyribonuclease
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DSBs	Double-stranded DNA breaks
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
ePTS1	Enhanced peroxisomal targeting signal type 1
FCM	Flow cytometry
FISH	Fluorescent *in situ* hybridization
FPLC	Fast procedure liquid chromatography
GFP	Green fluorescent protein
GRAS	Generally recognized as safe
HDR	Homology-directed repair
IdeS	Immunoglobulin G-degrading enzyme of *S. pyogenes*
IgG	Immunoglobulin G

IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
MS	Mass spectrometry
MFI	Mean fluorescent intensity
MPS	Massively parallel sequencing
Neq	*Nanoarchaeum equitans*
NHEJ	Non-homologous end joining
Ori	Origin of replication
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
Pfu	*Pyrococcus furiosus*
PTMs	Post-translational modifications
qPCR	Quantitative PCR
QPS	Qualified presumption of safety
RFP	Red fluorescent protein
RNase	Ribonuclease
RNA	Ribonucleic acid
RT	Reverse transcriptase
SEC	Size exclusion chromatography
SiMSen-Seq	Simple, multiplexed, PCR-based barcoding of DNA for sensitive mutation detection using sequencing
SNP	Single nucleotide polymorphism
SpeB	Streptococcal pyrogenic exotoxin B
SSB	Single strand DNA-binding protein
ssDNA	Single-stranded DNA
Sso7d	dsDNA-binding domain from *Sulfolobus solfataricus*
STR	Short tandem repeat
Taq	*Thermus aquaticus*
Tm	Melting temperature
UMI	Unique molecular identifier

Chapter 1
Introduction

The word enzyme was established by the German physiologist Wilhelm Kühne in 1877, derived from the Greek words *en* (meaning within) and *zume* (meaning yeast) to describe substances "within yeast" that were responsible for the fermentation process (Aehle, 2007; Kühne, 1877). Since then, significant advancements have been made in the discovery, extraction, and characterization of enzymes (Robinson, 2015). Enzymes act as biocatalysts, i.e., biological catalysts, and can increase reaction rates by up to a million-fold, depending on the specific reaction catalyzed (Adrio & Demain, 2014; Cooper, 2000). Many biochemical processes would not occur under life-compatible conditions without catalytic assistance from enzymes (Cooper, 2000). Cells contain thousands of distinct enzymes, whose specific activities determine which of the numerous potential biochemical reactions that are carried out at any given moment within the cellular environment. Enzymes accelerate biochemical reactions both within living organisms and in controlled environments outside cells (Cooper, 2000; Narayanan et al., 2023). Due to their catalytic nature, enzymes are effective at very low concentrations and catalyze reactions without being consumed or permanently altered (Robinson, 2015). In addition to their high catalytic efficiency, enzymes demonstrate remarkable specificity, typically catalyzing the conversion of a single substrate type, or a narrow range of structurally related substrates, into product molecules (Robinson, 2015). These characteristics have made enzymes important tools in several technological applications, such as in the development of industrial bioprocesses, diagnostic systems, and forensic science where reactions are catalyzed with great specificity (Adrio & Demain, 2014).

Enzymes and their role in biotechnology

Advancements in recombinant DNA technology and protein engineering have enabled the development of diverse industrial enzyme applications beyond their native functions (Sanchez & Demain, 2017). Further improvements in genomics, metagenomics, proteomics, and efficient expression and production systems have facilitated exploration of the rich biodiversity and discovery of novel microbial enzymes along with better engineered enzymes with enhanced catalytic properties (Adrio & Demain, 2014; Fasim et al., 2021).

Industries and markets where current applications involve enzymes include pulp and paper manufacturing, leather processing, detergent and textile production, pharmaceutical and chemical industries, food and beverage processing, biofuel generation, animal feed formulation, personal care products, molecular biology, and forensic science (Table 1.1) (Kirk et al., 2002; Liu & Kokare, 2017). Native host organisms are not always suitable for protein production and native enzymes may not be produced in sufficient quantities (Trono, 2019). Consequently, approximately 90% of industrial enzymes are recombinantly produced in optimized microbial host organisms, such as bacteria, yeasts or filamentous fungi (Adrio & Demain, 2014; Liu & Kokare, 2017; Robinson, 2015).

Microbial hosts remain the preferred systems for recombinant enzyme production due to their high yields, robust activity, reproducibility, and cost-effective production compared to insect and mammalian cells. The exponential growth of microorganisms in inexpensive media, combined with established and continuously improved protocols for process optimizations and strain improvements make them suitable for industrial use (Fasim et al., 2021).

Table 1.1: Microbial enzymes used in various industrial, medical and analytical applications. Adapted from Liu & Kokare, 2017 and Kirk et al., 2002.

Industry	Enzymes	Applications	References
Animal feed	β-Glucanase, phytase and xylanase	Increase total phosphorus content for growth and digestibility.	(Mitidieri et al., 2006; Tomschy et al., 2000)
Detergents	Amylase, cellulase, lipase, mannanase, and protease	Remove protein after staining, cleaning agents, color clarification, removing insoluble starch, fats and oils, and to increase effectiveness of detergents.	(Pandey et al., 2000; Wintrode et al., 2000)
Ethanol production	Cellulase, ligninase, and mannanase	Formation of ethanol	(Jolly, 2001)
Food, dairy and beverage	Amylase, amyloglucosidase, glucose oxidase, laccase, lactase, lipase, pectinase lipoxygenase, pectin methyl esterase, phospholipase, protease, transglutaminase, xylanase	Degradation of starch and proteins into sugars, production of low caloric beer, fruit juice processing, cheese flavor and production, milk clotting, glucose production from lactose, infant formulas and lactose removal from milk, dough stability and conditioning, modify visco-elastic properties, dough strengthening, bread whitening, bread softness and volume, and flour adjustment	(Gurung et al., 2013; Nigam & Singh, 1995)
Leather	Protease and lipase	Unhearing, bating, de-picking	(Binod et al., 2013; Saha et al., 2009)
Molecular biology	DNA ligase, restriction enzymes, polymerases, and proteases	Manipulate DNA in genetic engineering. DNA restriction and polymerase chain reaction (PCR).	(Nigam, 2013; Roberts et al., 2009) (**Paper I, II, III** and **IV**)
Paper and pulp	Amylase, lipase, protease, cellulase, hemicellulase, esterase, ligninase, and xylanase	Degrade starch to lower viscosity, aiding sizing, deinking, and coating paper. Cellulase and hemicellulase smooth fibers, enhance water drainage, and promote ink removal. Lipases reduce pitch and ligninase remove lignin to soften paper. Biofilm removal, contaminant control, and bleach boosting.	(Kirk et al., 2002; Kohli et al., 2001; Polizeli et al., 2005)
Personal care	Amyloglucosidase, glucose oxidase, and peroxidase	Antimicrobial and bleaching.	(Kirk et al., 2002)
Pharmaceuticals	Penicillin acylase, peroxidase, and protease	Synthesis of semisynthetic antibiotics, and antimicrobials	(Gurung et al., 2013; Roberts et al., 2007) (**Paper I** and **II**)
Starch and fuel	Amylase, amyloglucosidase, cyclodextrin-glycosyltransferase, glucose isomerase, protease, pullulanase, and xylanase	Starch liquefaction and saccharification, cyclodextrin production, glucose to fructose conversion, yeast nutrition, and viscosity reduction	(Kirk et al., 2002)
Textiles	Amylase, catalase, cellulase, keratinase, laccase, pectinase, peroxidase, and protease	Fabric finishing in denims, wool treatment, bleaching, bleach termination, degumming of raw silk (biopolishing), excess stain removal, and cotton softening.	(Liu et al., 2013; Saha et al., 2009)

Reactions that are catalyzed by enzymes can also be chemically catalyzed. However, there are several disadvantages with chemical synthesis processes, including limited catalytic performance, insufficient enantiomeric selectivity for stereoselective synthesis, and requirement for harsh reaction conditions including high temperature and pressure, acidic pH, and the use of organic solvents (Sanchez & Demain, 2017). In contrast, most enzymes are functional under mild reaction conditions, i.e., lower temperatures, neutral pH, and atmospheric pressure, and have a high stereo-selectivity and reaction rate (Abedi et al., 2011; Cooper, 2000). Enzymes can also be selected and modified to enhance stability, substrate specificity and specific activity, making them suitable as tools in biotechnology (Abedi et al., 2011; Adrio & Demain, 2014; Sanchez & Demain, 2017).

Despite the advantages of enzymatic reactions, enzymes have limitations in industrial and analytical applications. These limitations include requirements for co-factors, susceptibility to substrate or product inhibition, stability issues, and challenges in reuse of the enzymes across multiple reactions (Robinson, 2015). In research, diagnostic, and forensic applications, enzymes may be sensitive to inhibitors and contaminants, show reduced catalytic efficiency over time and suffer from high production costs for certain applications (Iyer & Ananthanarayan, 2008; Robinson, 2015; Sidstedt et al., 2020; Sidstedt et al., 2019; Uchii et al., 2019). Supply chains and centralized enzyme production have been disrupted and shown to be vulnerable in recent global crises, increasing the need for simple and robust production methods that can be rapidly employed in case of shortages. This also highlights the importance of exploring alternative host organisms for enzyme production which may offer advantages in terms of scalability, resilience, or resource efficiency. Investigating new hosts broadens the biotechnological toolbox and can strengthen preparedness for future disruptions in enzyme production systems.

Many enzymes either build or degrade biological polymers. For example, proteases catalyze the degradation of proteins and polymerases catalyze the synthesis of DNA strands. Proteases are widely used in food processing, detergent formulations, as biotechnological tools and for therapeutic antibody characterization (Contesini et al., 2018; Dyer & Weiss, 2022; Robinson, 2015; Sjögren et al., 2017). In molecular biology, polymerases are essential for DNA manipulation and sequencing, and they also enable genome editing technologies like CRISPR/Cas9 (Table 1.1) (Aschenbrenner & Marx, 2017; Ishino & Ishino, 2014; Rittié & Perbal, 2008; Robinson, 2015). Their sensitivity and selectivity make them valuable in diagnostics and forensic analyses, while their use in environmental biotechnology supports bioremediation and waste treatment (Liu & Kokare, 2017; McDonald et al., 2024; Robinson, 2015). Advances in synthetic biology and protein engineering continue to expand the applications of proteases and polymerases, driving innovation and sustainability across the biotechnology sector (Liu & Kokare, 2017; Rittié & Perbal, 2008).

Overview of proteases

Proteases are a large family of enzymes that catalyze the hydrolysis of proteins into peptides or individual amino acids and are widely utilized across various industries (Ward, 2011). In nature, proteases are ubiquitous and comprise approximately 2% of an organism's genome (Martinusen et al., 2025). They play critical roles in nearly all cellular processes, including managing protein life cycles, modulating cellular signaling, activating zymogens (inactive enzyme precursors), processing precursor proteins, recycling nutrients, and enabling adaptation to stress (Afrin et al., 2024; Bond, 2019; Culp & Wright, 2017; Sukharev et al., 1997). Proteases can be classified based on their catalytic residue, according to the specific type of reaction they catalyze or by their molecular structure and homology (Rawlings et al., 2014). Proteolytic enzymes are described as serine-, cysteine-, threonine-, aspartic-, glutamic-, asparagine- or metallo-proteases when classified based on the chemical mechanism of catalysis (Oda, 2011; Rawlings et al., 2014). Proteases that act on shorter peptides are sometimes referred to as peptidases. These often have their site of action at the ends of the amino acid chains. Endopeptidases cleave internal peptide bonds, exopeptidases target terminal peptide bonds, aminopeptidases specifically hydrolyze bonds at the amino-terminal end of peptides, and carboxypeptidases specifically hydrolyze bonds at the carboxy-terminal end of peptides (Figure 1.1) (Rawlings et al., 2014). Proteases represent over 60% of the global enzyme market, which requires efficient production (Dyer & Weiss, 2022; Sanchez & Demain, 2017; Ward, 2011). However, recombinant production of proteases includes several challenges. One of the main challenges is that their proteolytic activity can lead to degradation of essential cellular components in the non-native host organism, especially for proteases with a broad substrate scope, capable of degrading several different types of proteins (Komai et al., 1997; Kwon et al., 2011). Other challenges associated with recombinant protease production include formation of inclusion bodies, and aggregation of unfolded proteins.

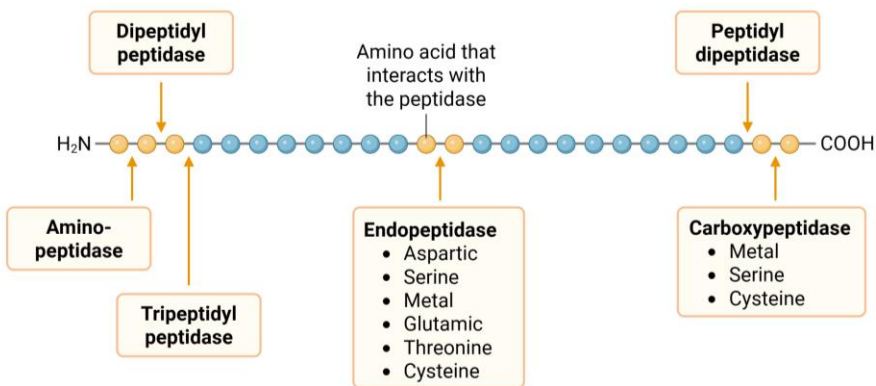


Figure 1.1: Types of peptidases and their sites of action.

The optimal production host may vary depending on the specific protease to be produced. In this thesis work, the yeast *Saccharomyces cerevisiae* was evaluated as a potential host for production of three bacterial proteases (BdpK, IdeS and SpeB) with different substrate specificities (**Paper I** and **II**). All three bacterial proteases degrade human immunoglobulin G (IgG) and are used as research tools, e.g., for therapeutic antibody characterization (Nelson et al., 2011; Sjögren et al., 2017; van Rijswijck et al., 2024; von Pawel-Rammingen et al., 2002). *S. cerevisiae* is a well-characterized host organism with known physiological behavior in large-scale processes. The usage of yeast as a production host also offers a way to study the effects of the proteases on a eukaryotic model system, which can be used to evaluate and develop new protease variants. In **Paper II**, detailed analysis was conducted on how proteases with varying substrate scopes affect the physiology of *S. cerevisiae*. The protease BdpK originates from the predatory bacterium *Bdellovibrio bacteriovorus* and demonstrates a broad-spectrum activity (van Rijswijck et al., 2024). The other two proteases, IdeS and SpeB, originate from the pathogenic bacterium *Streptococcus pyogenes.* IdeS is a highly specific protease with a unique specificity for human IgG, while SpeB is a broad-spectrum protease (Carroll & Musser, 2011; Happonen & Collin, 2024; Johansson et al., 2008; Vincents et al., 2004).

Overview of DNA polymerases

Polymerases are essential enzymes that catalyze the synthesis of nucleic acids and are fundamental in the process of DNA replication and RNA transcription in all living organisms (Akram et al., 2023; Case & Hingorani, 2017). Nucleotide building blocks are assembled into long chains of DNA or RNA, where polymerases use a template strand to direct sequence-specific synthesis, ensuring the correct inheritance of genetic information and precise gene expression (Case & Hingorani, 2017). Reverse transcriptases (RT) are RNA-dependent DNA polymerases that use RNA to generate DNA, and are commonly found in retroviruses (Clark et al., 2019). Polymerases are highly specific for their substrates, with many possessing proofreading capabilities that enhance the fidelity of nucleic acid synthesis and minimize the risk of mutations (Akram et al., 2023). Polymerases act in coordination with several other enzymes inside cells in order to synthesize nucleic acids in an accurate, efficient and well-regulated manner (Case & Hingorani, 2017). The diverse activities of polymerases are essential for cellular growth, maintenance, repair, and heredity, forming the molecular foundation for all biological systems (Garcia-Diaz & Bebenek, 2007; Laatri et al., 2024). Since the activity of polymerases is fundamental for all nucleic acid processing, their core structure and catalytic properties are highly conserved through evolution (Case & Hingorani, 2017).

Based on sequence homology and crystal structure analysis, DNA polymerases are classified into seven families: A, B, C, D, X, Y and RT (Ishino & Ishino, 2014; Steitz, 1999). Despite differences in overall structure and function, all families share a conserved catalytic domain, with similar catalytic residues responsible for the polymerization activity (Braithwaite & Ito, 1993; Steitz, 1999). The specificity and characteristics of DNA polymerases have made them important as biotechnological tools (Akram et al., 2023). DNA polymerases from thermophilic bacteria and archaea are commonly used in *in vitro* DNA polymerization for DNA analysis and as molecular biology tools where their thermostable properties are important (Bartlett & Stirling, 2003; Ishino & Ishino, 2014; Witte et al., 2018). Two of the most used DNA polymerases in polymerase chain reaction (PCR) applications are *Taq* polymerase from *Thermus aquaticus* and *Pfu* polymerase from *Pyrococcus furiosus* (Chien et al., 1976; Cline et al., 1996; Eckert & Kunkel, 1990; Lundberg et al., 1991; Tindall & Kunkel, 1988). Novel and engineered DNA polymerase variants are continuously developed, often tailored to the requirements of specific applications and providing improved activity, processivity, fidelity, or PCR inhibitor tolerance (Coulther et al., 2019; Ishino & Ishino, 2014; Sidstedt et al., 2020).

This thesis work includes recombinant production of *Taq* and *Pfu* polymerases, as well as *Taq*-Sso7d and *Pfu*-Sso7d using *Escherichia coli* as production host (**Paper III**). Sso7d is a double-stranded DNA-binding protein from *Sulfolobus solfataricus* previously shown to increase processivity when fused to DNA polymerases (Farooqui et al., 2023; Wang et al., 2004). A robust production protocol, feasible even with limited resources, was developed and the produced DNA polymerases were evaluated for use in PCR and qPCR assays and in library preparation for sequencing. Errors introduced during PCR and sequencing typically impair the limit of detection and in forensic DNA profiling it may lead to masking of DNA information from minor contributors in mixtures and faulty allele calling (Filges et al., 2019; Raz et al., 2019; Woerner et al., 2021). In forensic DNA profiling, microsatellites or Short Tandem Repeat (STR) markers are commonly used. The repetitive nature of STRs makes them highly susceptible to systematic artefacts called stutters. These are typically one repeat shorter than the true allele (Hauge & Litt, 1993; Meldgaard & Morling, 1997; Sparkes et al., 1996; Walsh et al., 1996). In this thesis, the effects of DNA polymerases with different characteristics on amplicon yield and PCR error formation, both systematic stutter artefacts and random base-substitutions, were studied in STR analysis (**Paper IV**).

Scope and outline of the thesis

The overall objective of this PhD thesis was to study recombinant production of proteases and DNA polymerases and their biotechnologically relevant characteristics. Specifically, my work included investigating the yeast host interactions with recombinant bacterial proteases, establishing a resource-efficient in-house protocol for polymerase production and characterizing polymerases as well as type and rate of errors during *in vitro* DNA polymerization. Compartmentalization of recombinant proteases was explored as a strategy to protect the yeast cell from proteolytic activity. Initial evaluation of a 28-plex STR assay utilizing unique molecular identifiers (UMIs) was performed, indicating potential, but also the need for further optimization. This work contributes to the research field by (i) investigating the use of yeast as a potential production host for bacterial proteases, addressing challenges associated with protease production, (ii) exploring single-cell monitoring of protein production through flow cytometry (FCM) by fusing proteases to Green Fluorescent Protein (GFP), (iii) developing a simple and practical protocol for DNA polymerase synthesis with limited resources only requiring readily available laboratory materials, and (iv) investigating how different DNA polymerase characteristics affect the levels of single-base substitutions and stutter artefacts in STR analysis.

Chapter 2 introduces the enzymatic mechanisms of action of proteases, their native activity and their biotechnological applications. Chapter 3 explores the enzymatic mechanisms behind the activity of DNA polymerases, their activity inside cells and their characteristics and applications in biotechnology. Chapter 4 focuses on the host organisms and methods for recombinant protein production used in my work, including expression systems and vectors. Challenges with recombinant enzyme production and strategies to improve protein production and purification are presented. Chapter 5 discusses how DNA polymerase characteristics impact amplicon yield and noise levels in STR analysis. Finally, the conclusions and outlook of the thesis work and its associated Papers I-IV are found in Chapter 6 and Chapter 7, respectively. This thesis is composed of four research papers, and the focus of each paper is described below.

In **Paper I**, a method to recombinantly produce the bacterial protease IdeS in the yeast *S. cerevisiae* is reported, together with the fusion to GFP to enable single cell measurements by flow cytometry. CRISPR/Cas9 was used to integrate the gene cassette in the yeast genome, applying the inducible *GAL1* promoter (*GAL1*p) to control protein production.

In **Paper II**, the method reported in Paper I was used to further explore yeast as a production host for bacterial proteases. There, three bacterial proteases (BdpK, IdeS and SpeB) with different substrate scopes were fused to GFP and produced in *S. cerevisiae*. Expression from a multicopy plasmid and from single-copy genome

integration were explored, and the protein production was monitored on a single-cell level through flow cytometry. The *GAL1*p was used to separate the growth and protein production phase and the physiological effects on the yeast host and protease activity were explored.

Paper III describes the development of a simple protocol to produce DNA polymerases, with minimal requirements on reagents and instrumentation. In-house synthesis of DNA polymerases with the developed protocol was performed and the polymerases were evaluated with PCR, colony PCR, real-time quantitative PCR (qPCR) and library preparation for subsequent sequencing. The library preparation used the SiMSen-Seq technique (described in Chapter 3), which enabled insights into base substitutions and stutter artefacts made by *Taq* and *Pfu* polymerase with or without the dsDNA-binding domain Sso7d.

Paper IV focuses on how different properties of DNA polymerases affect amplicon yield and the formation of PCR errors, including single-base substitutions and stutter artefacts. Six DNA polymerases with varying characteristics, such as proofreading capacity and DNA-binging domains, were applied in the two PCRs of the SiMSen-Seq STR assay. Sequencing and the use of UMIs enabled detailed insights into both the first few PCR cycles where genomic DNA act as template, as well as the later cycles where short STR amplicons act as template. This was achieved by looking at all STR sequences and consensus sequences grouped by the same UMI, reflecting the polymerase performance in the initial PCR cycles.

Chapter 2
Enzymatic activity and mechanisms of proteases

Proteases are critical components in cellular functions, playing key roles in protein turnover, signaling, and regulation (López-Otín & Bond, 2008). The proteolytic degradation of polypeptide chains, through hydrolysis of peptide bonds is essential for recycling of amino acids and for preventing the accumulation of damaged or misfolded proteins and thereby for sustaining life (Dos Santos, 2011; Hengge & Turgay, 2009).

Peptide bonds are generally stable in aqueous solution at neutral pH and room temperature, with hydrolysis occurring extremely slowly. Half-lives range from 7 to over 600 years depending on the bond environment (Kahne & Still, 1988; Radzicka & Wolfenden, 1996; Vlasak & Ionescu, 2011). The exact rate of hydrolysis of peptide bonds varies with the amino acid sequence, where certain sites such as glycine-glycine may be more susceptible to hydrolysis. However, spontaneous cleavage is rare under physiological conditions (Radzicka & Wolfenden, 1996; Vlasak & Ionescu, 2011). Hydrolysis rates increase under extreme pH conditions (acidic or basic) and elevated temperatures (Vlasak & Ionescu, 2011). For example, acid hydrolysis at 105 °C in 6 M hydrochloric acid for around 24 hours, breaks down most peptide bonds into the constituent amino acids (Roach & Gehrke, 1970). Some proteins, like ribonuclease A, can resist degradation and require even higher temperatures or longer times. Additionally, metal ions such as zinc and nickel can promote site-specific cleavage, and chemicals like cyanogen bromide and hydroxylamine enable targeted bond disruption for protein analysis (Krezel et al., 2006; Lyons et al., 2016).

Although non-enzymatic cleavage is possible, proteases offer several advantages for controlled hydrolysis as they act under mild, physiological conditions, exhibit high specificity for sequences or structural motifs, and allow precise and efficient protein degradation.

Biological function of proteases

Proteases have several critical roles inside cells, including regulating protein turnover, activation or inactivation of signal molecules, facilitating cellular responses and contributing to essential processes such as cell division and apoptosis (Bond, 2019; López-Otín & Bond, 2008). Proteases are involved in several processes to obtain functional proteins. Limited proteolysis often occurs during or after translation and is often essential for converting newly synthesized polypeptides into their functional forms (Rogers & Overall, 2013). The initial translation product is often referred to as a preproprotein, which includes both a signal peptide and a propeptide. Proteolytic cleavage of the signal peptide generates the intermediate proprotein and further processing results in the mature and active protein (Fikes et al., 1990). Proteins that are secreted or localized to organelles typically have an N-terminal signal peptide that directs them to their final cellular destination (Harwood & Kikuchi, 2022). The signal peptide is removed by proteolysis following translocation across membranes (Harwood & Kikuchi, 2022). Some proteins are synthesized as large precursors called polyproteins, consisting of several enzymes. These zymogens or proenzymes require proteolytic cleavage into smaller and functional polypeptides (Wetmore et al., 1992).

In microorganisms, proteases are also crucial for nutrient acquisition, environmental adaptation, competition, and pathogenesis, all of which contribute to their survival and interaction with the surrounding environment (Muszewska et al., 2017; Zhang et al., 2020). Proteolysis serves as a mechanism for nutrient recycling, occurring extracellularly and intracellularly. During extracellular digestion, enzymes break down proteins into peptides or amino acids to be metabolized (Mahmoud & Chien, 2018). Proteolytic activity in microorganisms can be regulated by nutrient availability (Hengge & Turgay, 2009). Limitations of carbon, nitrogen or sulfur can induce proteolytic activity in some microorganisms or microbial communities (Li et al., 2021).

Intracellularly, proteases are responsible for maintaining protein homeostasis and quality control by degrading misfolded or damaged proteins, modulating cell division, and responding to stress (Culp & Wright, 2017; Muszewska et al., 2017). This degradation also regulates cellular processes by removing enzymes and regulatory proteins that are no longer needed (Rogers & Overall, 2013). Some proteases are involved in signal peptide processing, vacuole maintenance, and intracellular protein recycling (Muszewska et al., 2017). The rate of protein degradation varies depending on the protein's function and condition. Abnormal proteins are typically degraded rapidly, while others are regulated based on metabolic needs (Goldberg, 2004). Proteases are also important in adaptation, as changes in the proteome can be advantageous as a response to harsh or changing environmental conditions (Kallazhi et al., 2025; Omnus et al., 2023).

Proteases are often synthesized as inactive precursors, zymogens, to prevent premature or unwanted proteolytic activity that could be harmful to the cell (Wetmore et al., 1992). Activation typically occurs in a specific cellular compartment or physiological context (Turk et al., 2012). As an example, in *S. cerevisiae*, protein degradation and recycling occur in the vacuole and at least seven distinct vacuolar proteases have been characterized and experimentally verified (Hecht et al., 2014; Rawlings et al., 2014). The vacuole is also vital under nutrient limitation and stressful conditions, where degradation of up to around 85% of the cell's intracellular protein content has been observed during nutrient starvation (Hecht et al., 2014; Teichert et al., 1989). Many vacuolar proteases are translated as precursors that are proteolytically cleaved to become mature and active proteases in the vacuole (Hecht et al., 2014; Jones, 1991; van den Hazel et al., 1996).

Catalytic mechanisms of proteases

Proteases described based on their catalytic mechanisms include serine-, cysteine-, threonine-, aspartic-, glutamic-, asparagine- and metallo-proteases (Figure 2.1) (Rao et al., 1998; Rawlings et al., 2014; Ward, 2011). In this thesis, I focused on one serine protease and two cysteine proteases (**Paper I** and **II**). Serine proteases are a large group characterized by the presence of a serine residue in their active site (Hecht et al., 2014; Rao et al., 1998). The nucleophilicity of the catalytic serine is typically dependent on a catalytic triad of Asp-His-Ser residues, which is one of the most thoroughly characterized catalytic motifs in enzyme biochemistry (Blow et al., 1969; Di Cera, 2009). However, there are several other catalytic triads and dyads within serine proteases (Di Cera, 2009; Hedstrom, 2002).

Cysteine proteases, also known as thiol proteases, share the catalytic mechanism involving a nucleophilic cysteine thiol in a catalytic dyad or triad (Rawat et al., 2021; Verma et al., 2016; Yang et al., 2023). Threonine proteases are enzymes harboring a threonine residue within the active site and use the secondary alcohol of their N-terminal threonine as a nucleophile to perform catalysis (Brannigan et al., 1995). Aspartic proteases use aspartic residues as ligands of the activated water molecule in the active site (Szecsi, 1992). In most aspartic peptidases, a pair of aspartic residues act together to bind and activate the catalytic water molecule. However, in some cases other amino acid residues replace the second aspartic acid (Hecht et al., 2014; Rawlings & Barrett, 2013a). Glutamic proteases contain one or two glutamic acid residues that bind and activate the catalytic water molecule for their proteolytic activity (Rawlings & Barrett, 2013a). Asparagine peptide lyases are the most recently discovered group of proteolytic enzymes. Their catalytic mechanism involves an asparagine residue acting as a nucleophile to perform a nucleophilic elimination reaction, rather than hydrolysis to catalyze the breaking of a peptide bond (Rawlings et al., 2011). Metalloproteases depend on divalent metal

cations, mainly zinc (Zn^{2+}), but sometimes cobalt, manganese, nickel, copper or iron, to activate a water molecule for their catalytic function, where the metal ion is coordinated to the protein usually via three ligands (Hecht et al., 2014; Rawlings & Barrett, 2013b). Besides the metal ion, at least one other residue is required for catalysis, usually consisting of glutamic acid (Rawlings & Barrett, 2013b). Aspartic, glutamic and metallopeptidases differ from serine, cysteine and threonine peptidases in that they utilize an activated water molecule as the nucleophile to cleave the peptide bond, rather than relying on the nucleophilic side chain of an amino acid residue in the active site (Rawlings & Barrett, 2013a). Catalytic residues include amino acids that function as ligands, either directly for the activated water molecule as seen in aspartic and glutamic proteases, or by binding one or two metal ions that in turn bind the activated water molecule in metalloproteases (Rawlings & Barrett, 2013a).

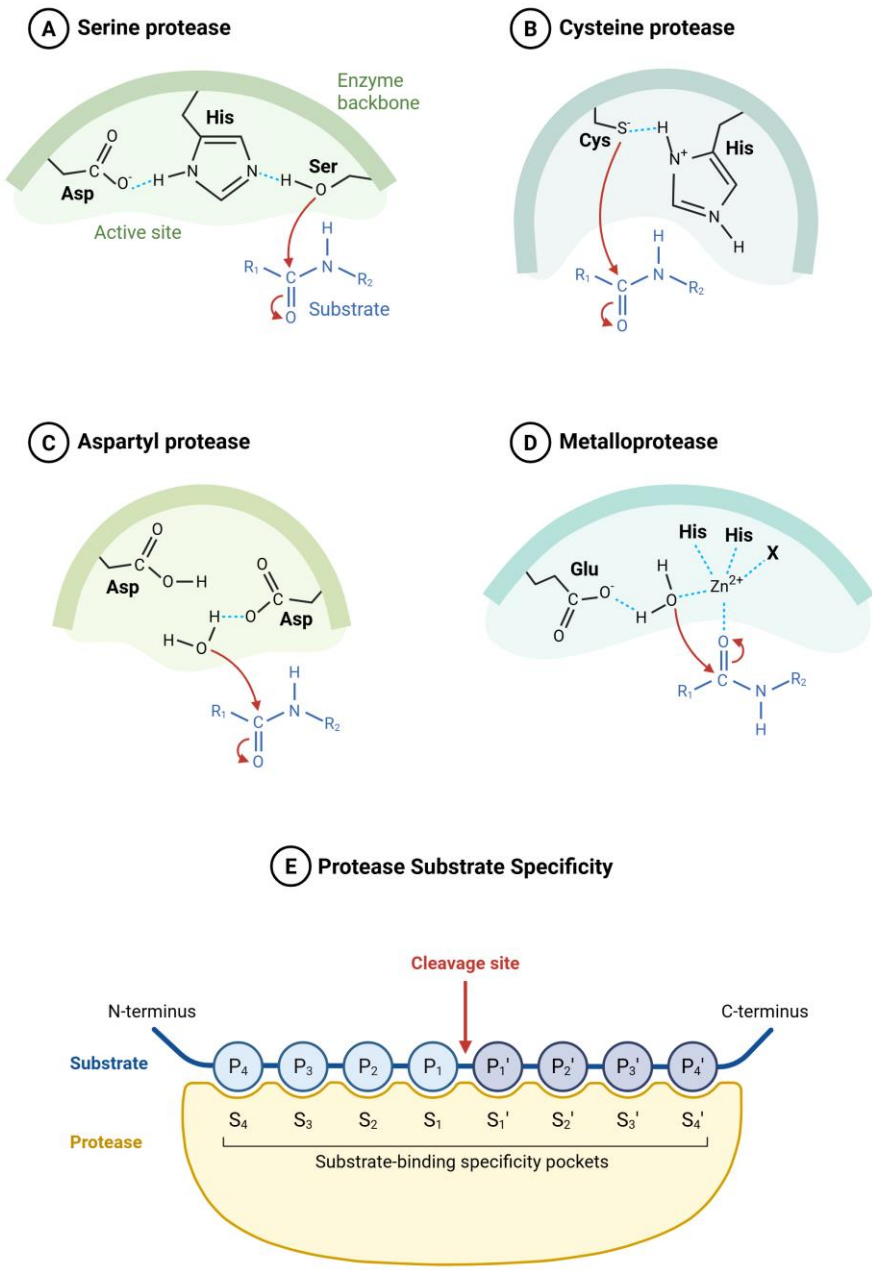


Figure 2.1: Examples of catalytic mechanisms for four types of proteases (**A**-**D**) and a schematic overview of protease substrate specificity (**E**).

Applications of proteases in biotechnology

Proteases are used in biotechnology as tools in several different types of applications. These applications include proteomics, diagnostic tools, antibody processing and characterization, specific removal of protein tags, degradation of proteins in RNA and DNA extraction and as therapeutic enzymes (Rao et al., 1998; Song et al., 2023). Trypsin is commonly used to digest proteins into smaller peptides for mass spectrometry (MS)-based proteomics (Dau et al., 2020). Digestion by trypsin occurs at the C-terminal of arginine and lysine residues, leaving a positive charge on the peptide C-terminus, which is advantageous for MS analysis (Dau et al., 2020; Woessmann et al., 2023). Trypsin is a serine protease, with high cleavage specificity and is stable under a wide variety of conditions (Olsen et al., 2004; Woessmann et al., 2023). Other proteases have also been explored for proteomics, sometimes in combination with trypsin in a sequential digestion (Dau et al., 2020; Giansanti et al., 2016). Site-specific cleavage for removal of protein tags, e.g., after protein purification, is commonly performed with tobacco etch virus (TEV) protease (Raran-Kurussi et al., 2017).

Proteases have essential roles in cells of all living organisms, and multiple pathological conditions including cancer, neurodegenerative disorders, inflammatory and cardiovascular diseases involve alterations in proteolytic systems (López-Otín & Bond, 2008). Overexpression or dysregulated activity of proteases in diseases can thus be used as diagnostic biomarkers for early detection as well as therapeutic targets by blocking of the appropriate proteases (López-Otín & Bond, 2008; Turk, 2006). Both protease inhibitors and proteases are utilized and approved as therapeutic agents (Al-Salama, 2020; Turk, 2006). Examples of protease inhibitors for clinical use are angiotensin-converting enzyme (AEC) inhibitors that were introduced in the 1980s and are used for treatment of several cardiovascular diseases (Turk, 2006). With the unique specificity for IgGs, the protease IdeS is also approved for clinical use against antibody-mediated transplant rejection and is used in ongoing clinical trials against other autoimmune conditions (Al-Salama, 2020; Happonen & Collin, 2024; Huang et al., 2022).

In DNA and RNA extraction, proteinase K, a broad-spectrum serine alkaline protease, named after its ability to degrade keratin, is commonly used (Morihara & Tsuzuki, 1975). Proteinase K can digest a wide range of proteins, including some that are resistant to other proteases like trypsin (Ebeling et al., 1974; Saenger, 2013). When isolating DNA or RNA, proteinase K is used to degrade unwanted proteins, such as nucleases and ribonuclease (RNases) and deoxyribonucleases (DNases), to obtain intact DNA or RNA (Gautam, 2022; Goldenberger et al., 1995). Proteinase K also breaks down the protein component in the cell membrane and nuclear proteins, allowing access to DNA and RNA in the nucleic acid extraction process (Hilz et al., 1975; Petsch et al., 1998). In the lysed sample, proteinase K also breaks down other chromosomal proteins and histones, which otherwise may interfere with

downstream applications such as PCR, RT-PCR, Sanger sequencing, Massively Parallel Sequencing (MPS), fluorescent *in situ* hybridization (FISH) and other molecular biology research and diagnostic applications (Shakoori, 2017; Sidstedt et al., 2020; H. Zhao et al., 2020).

Therapeutic antibodies represent a substantial and rapidly growing part of the biopharmaceutical market, where proteases are crucial in their characterization. Proteases such as papain, pepsin, Lys-C, IdeS, IdeZ, SpeB, and BdpK are routinely used to cleave antibodies into defined fragments, which is essential for structural characterization, therapeutic development and analytical assays (Beck et al., 2013; Happonen & Collin, 2024; Romei et al., 2023; Ryan et al., 2008; van Rijswijck et al., 2024). Analysis of antibodies by MS and chromatography is simplified by generating predictable antibody fragments through protease digestion, enabling evaluation of sequence, glycosylation patterns and structural features (Ryan et al., 2008; Schillinger et al., 2024). The controlled fragmentation obtained with proteases is used when studying antibody-antigen interactions (epitope mapping), specifically targeting of antibodies and purification processes for research and pharmaceutical production (Martinusen et al., 2025; Ryan et al., 2008).

The evaluated proteases BdpK, IdeS and SpeB

The bacterial proteases synthesized and evaluated in *S. cerevisiae* in this thesis (**Paper I** and **II**), BdpK, IdeS and SpeB, have different substrate scopes and catalytic properties (Table 2.1). The characteristics of these proteases and their ability to cleave human IgG with high specificity have made them valuable as biotechnological tools.

Table 2.1: Bacterial proteases evaluated in this thesis. The origin of the proteases, their size and their substrate specificity are presented. For substrate specificity, the sequence in parenthesis represents the cleavage site in the hinge region of human IgG. Adapted from Paper II.

Protease	Origin of protein	Size (kDa)	Substrate specificity	References
BdpK	*B. bacteriovorus*	32	Broad substrate specificity, cleaves human IgG1 above the hinge region. (KSCDK/THTCPPCP)	(van Rijswijck et al., 2024)
IdeS	*S. pyogenes*	38	IgG substrate specific, recognizes 3D structure and cleaves below the hinge region. No other substrates are reported. (CPAPELLG/GPSVF)	(Happonen & Collin, 2024; Johansson et al., 2008; Sjögren et al., 2017; Vincents et al., 2004; von Pawel-Rammingen et al., 2002)
SpeB	*S. pyogenes*	28	Broad substrate specificity, cleaves human IgG above the hinge region. (KTHT/CPPCPAP)	(González-Páez & Wolan, 2012; Gubba et al., 1998; Happonen & Collin, 2024; Lane & Seelig, 2016; Nelson et al., 2011; Persson et al., 2013; Sjögren et al., 2017)

BdpK is a serine protease derived from the predatory bacterium *B. bacteriovorus* (van Rijswijck et al., 2024). *B. bacteriovorus* is non-pathogenic to humans, and a Gram-negative bacterium that lives through preying on other Gram-negative bacteria (Im et al., 2018; Sockett, 2009). The life cycle of *B. bacteriovorus* consists of two different phases, a free-swimming nonreplicative attack phase where free living cells are searching for new prey, and a replicative growth phase inside the prey's periplasm where it is growing and replicating DNA, forming a bdelloplast structure, degrading the host's macromolecules and reusing them as nutrients and building blocks for its own growth and replication which occurs right before the lysis of the prey, releasing daughter cells into the medium (Herencias et al., 2020; Makowski et al., 2019). Due to *B. bacteriovorus'* predatory lifestyle, it has a high enzyme-to-genome ratio with many proteases, resulting in one of the highest reported densities of protease genes per genome among bacterial species (Bratanis et al., 2017; Rendulic et al., 2004). One of these proteases is BdpK, which is a broad-acting protease, that due to the conformation of the substrate, only cleaves human IgG1 at a single site above the hinge region generating intact Fab (fragment antigen-binding region) and Fc (fragment crystallizable region) fragments, even on mutated hinge regions (van Rijswijck et al., 2024). The enzymatic activity of BdpK is dependent on calcium ions as cofactor, while no reducing agent is required (van Rijswijck et al., 2024).

IdeS and SpeB are two secreted cysteine proteases from the pathogen *S. pyogenes* acting as virulence factors. Proteases are important virulence factors of pathogenic bacteria that play an important role in the colonization and circumvention of the host defense system (Maeda & Yamamoto, 1996; Vincents et al., 2004). A common mechanism used by many pathogenic bacterial species is direct interference with the adaptive immune system by cleaving antibodies (Collin & Kilian, 2014; Maeda & Yamamoto, 1996; Travis & Potempa, 2000). *S. pyogenes* is a group A *Streptococcus*, a well-characterized human pathogen and the causing agent of several mild infections like pharyngitis, as well as some severe life-threatening diseases like streptococcal toxic shock syndrome and sepsis (Nelson et al., 2011; Persson et al., 2013). A variety of immunomodulating proteases are produced, both secreted and surface-attached, allowing *S. pyogenes* to survive and proliferate in the human host (Collin & Kilian, 2014; Happonen & Collin, 2024; Persson et al., 2013). IdeS is well-characterized and has a unique specificity for human IgG (von Pawel-Rammingen et al., 2002). All subclasses of human, monkey, rabbit, and sheep IgGs are digested by IdeS and no other substrate than IgG is known. No cofactors or reducing agent are required for the proteolytic activity. The specificity of IdeS has been shown to be dependent on the protein-protein interaction where IdeS binds to the Fc region of the IgG before cleavage occurs and the IgG is hydrolyzed below the disulfide bridge in the hinge region (Johansson et al., 2008; Vincents et al., 2004).

The streptococcal cysteine protease SpeB was one of the first secreted enzymes to be identified and is among the most studied proteases from any pathogen (Nelson et al., 2011; Vincents et al., 2004). The proteolytic activity of SpeB is dependent on reduction of the catalytic-site cysteine (Collin & Olsén, 2001; Persson et al., 2013). Therefore, activation of SpeB in enzymatic *in vitro* assays is usually achieved through addition of a reducing agent such as dithiothreitol (DTT) or β-mercaptoethanol (β-ME) (Collin & Olsén, 2001; Persson et al., 2013). The physiological relevance of SpeB-mediated antibody hydrolysis has been questioned (Persson et al., 2013), while SpeB has been shown to have *in vitro* activity on several human proteins in the extracellular matrix and plasma (Nelson et al., 2011). The native protein is secreted as an inactive zymogen and then autocatalytically processed in several steps into the active form of the protease (Carroll & Musser, 2011; Happonen & Collin, 2024). SpeB is a broad-spectrum protease acting on a wide variety of substrates, cleaving peptide loops with a loose requirement of amino acid composition, as long as they are fairly accessible in solution, without cofactor requirement (Nelson et al., 2011).

The three proteases applied in my thesis work, BdpK, IdeS and SpeB (**Paper I** and **II**), are currently used as research and development tools for antibody production within the pharmaceutical industry as they all cleave human IgG around the hinge region (Figure 2.2) (Happonen & Collin, 2024; Johansson et al., 2008; Nelson et al., 2011; van Rijswijck et al., 2024; Vincents et al., 2004). The specificity of BdpK is determined by the tertiary structure of IgG, which exposes only a single lysine residue in the hinge region of IgG1, making the protease highly selective (van Rijswijck et al., 2024). This type of specificity resembles that of broad acting proteases, such as KGP or trypsin, but differs from others like papain, which cleave IgG1 above the hinge region (Wang & Wang, 1977). Papain has lower specificity when digesting human IgG1, as it targets multiple cleavage sites, making it less suitable for LC-MS-based IgG1 Fab clonal profiling (Collins & Khalili, 2022).

The strict specificity towards IgG makes IdeS a suitable and precise tool for IgG analysis for diagnostic and therapeutic monitoring (Happonen & Collin, 2024; Huang et al., 2022; Johansson et al., 2008; Lonze et al., 2018). The tertiary IgG structure where IdeS first binds to the Fc region makes the digestion of IgG by IdeS rapid and efficient at the lower hinge region between two glycine residues (Gly-236/Gly-237) or less efficient between glycine and alanine in human IgG2 (Sudol et al., 2022; Wenig et al., 2004). This digestion results in F(ab')2 and Fc fragments, while leaving other proteins and immunoglobulin classes (IgM, IgA, IgD and IgE) intact (von Pawel-Rammingen et al., 2002; Wenig et al., 2004).

SpeB, with broad proteolytic activity, cleaves IgG specifically in the hinge region, generating Fab and Fc fragments, under reducing conditions (Happonen & Collin, 2024; Nelson et al., 2011; Sjögren et al., 2017). The broad specificity of SpeB can degrade the heavy chains of several immunoglobulins (IgM, IgA, IgD and IgE) and other host proteins and the usage is therefore context specific (Collin & Olsén, 2001;

Gubba et al., 1998; Happonen & Collin, 2024). Several different subclasses of IgG from different species can be digested by SpeB and the generated fragments can be used in affinity studies, structural studies and glycosylation studies of Fab fragments (Blöchl et al., 2022).

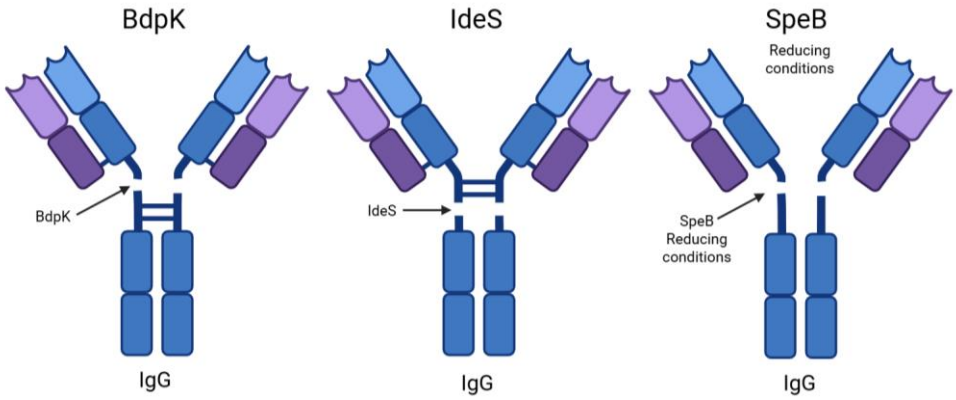


Figure 2.2: Cleavage site for the three proteases BdpK, IdeS and SpeB in the hinge region of human IgG. BdpK and SpeB cleave above the disulfide bonds in the hinge, while IdeS cleaves below the disulfide bonds in the hinge region.

Chapter 3
Enzymatic activity and mechanisms of DNA polymerases

DNA polymerases are vital enzymes in both prokaryotic and eukaryotic cells, responsible for synthesizing new DNA strands during genome replication and DNA repair (Akram et al., 2023). These processes are fundamental for inheritance, cell survival and genetic stability (Bebenek & Kunkel, 2004). DNA synthesis requires high fidelity and coordination, as errors can lead to mutations or genomic instability (Garcia-Diaz & Bebenek, 2007). While the formation of phosphodiester bonds between nucleotides is thermodynamically favorable, it is extremely slow under physiological conditions due to a high kinetic barrier, necessitating enzyme catalysis (Cherepanov & de Vries, 2003; Schroeder et al., 2006). DNA polymerases accelerate this otherwise slow chemical process, enabling rapid and highly specific synthesis of DNA with great accuracy and efficiency. The Klenow fragment, a large protein product of *E. coli* DNA polymerase I after cleavage by the protease subtilisin, was one of the first DNA polymerases to be characterized (Jacobsen et al., 1974; Klenow & Henningsen, 1970).

Biological function of DNA polymerases

During replication, DNA polymerases catalyze the addition of deoxyribonucleoside triphosphates (dNTPs) to the 3'-end of a growing DNA strand, using an existing strand as template (Berdis, 2009). This ensures accurate copying of genetic information during cell division. DNA polymerases function as part of large multi-enzyme complexes known as replisomes (Benkovic et al., 2001). The replisome coordinates the activities of various proteins including helicases, topoisomerases, primases, single strand DNA-binding proteins (SSBs), clamp loaders, sliding clamps, RNases, ligases, and DNA polymerases (Baker & Bell, 1998). These components work together to unwind DNA, stabilize single strands, initiate synthesis, and join fragments, enabling efficient and high-fidelity replication. The bacterial replisome consists of at least 30 proteins (Willey et al., 2019). DNA polymerases cannot synthesize DNA *de novo* and thus require DNA or RNA as a

starting material (Greci et al., 2022). Additionally, DNA polymerases require a free 3'-hydroxyl group to initiate dNTP incorporation and nucleotides are added to the 3'-end of the growing DNA strand (Berdis, 2009). The directional synthesis occurs at the replication fork, a dynamic structure where leading and lagging strand synthesis is coordinated. The leading strand is synthesized in the 5' to 3' direction, one nucleotide at a time, and the DNA polymerase moves in the same direction as the replication fork, continuously synthesizing DNA while the DNA unwinds ahead of the polymerase, only requiring one primer (Hamdan & van Oijen, 2010; Willey et al., 2019). The lagging strand on the other hand cannot be synthesized in the same direction as the movement of the replication fork since there is no free 3'-OH group to which a nucleotide can be added (Willey et al., 2019). The lagging strand is therefore synthesized discontinuously in the 5' to 3' direction, opposite to the direction of the replication fork, in a series of fragments called Okazaki fragments (Okazaki et al., 1968). Primases make many RNA primers along the template strand, which are extended by the DNA polymerase and eventually the Okazaki fragments are joined to form a complete strand. The many RNA primers must then be removed from the lagging stand, the gaps between the Okazaki fragments filled in by a DNA polymerase and ligated together by a ligase to form a complete DNA strand (Willey et al., 2019).

Microbial DNA polymerases enable rapid and accurate genome replication for cell division and adaptation to environmental changes, which is important to maintain short generation times (O'Donnell et al., 2013). Less accurate replication can in some cases be advantageous and lead to genetic variation through mutations that can drive evolution, adaptation and affect antibiotic resistance (Denamur & Matic, 2006; Martinez & Baquero, 2000).

Catalytic mechanisms of DNA polymerases

DNA polymerases have highly conserved structures, with similar overall catalytic subunits and little variation between species, independent of their subunit composition (Berdis, 2009; Joyce & Steitz, 1995). Based on amino acid sequence homology and crystal structure, DNA polymerases are divided into seven families (Table 3.1) (Akram et al., 2023; Case & Hingorani, 2017). The structure of most DNA polymerases is often compared to that of a partially closed right hand, consisting of "thumb", "palm", and "fingers" domains (Hamilton et al., 2001; Steitz, 1999). The function of the palm subdomain is the catalysis of the phosphoryl transfer reaction (Joyce & Steitz, 1995). Three conserved carboxylate residues in the DNA polymerase active site coordinate two Mg^{2+} ions essential for catalysis as cofactors (Figure 3.1 A) (Steitz, 1998). One of the metal ions facilitates deprotonation of the growing strand's 3'-OH group while the other one stabilizes the transition state during the nucleophilic attack on the incoming dNTP's α-

phosphate (Figure 3.1 A) (Steitz, 1998, 1999). The spatial arrangement of these catalytic residues is highly conserved between polymerases, reverse transcriptases and RNA polymerases (Joyce & Steitz, 1994; Steitz, 1999). The thumb subdomain is less conserved but has a common feature of largely helical structure and plays a role in positioning the duplex DNA and affects processivity and translocation (Hamilton et al., 2001). The thumb is believed to interact with the minor groove and with the incoming nucleotide. The fingers subdomain shows the greatest structural diversity of the three subdomains, where the finger domain may play a role in template fixation and template specificity (Joyce & Steitz, 1995). Like the thumb, it is likely to interact with the incoming nucleotide.

Table 3.1: DNA polymerase families, their functions and examples. Adapted from Willey et al., 2019.

Family	Examples	Functions
A	Bacterial DNA polymerase I	Replacements of RNA primers present in Okazaki fragments
B	Archaeal DNA polymerase B, Eukaryotic DNA polymerases α, δ, and ε	Replicative DNA polymerases in eukaryotes and in some archaea
C	Bacterial DNA polymerase III	Replicative DNA polymerase
D	Archaeal DNA polymerase D	Replicative DNA polymerase in some archaea, unique to archaea
X	Eucaryotic DNA polymerase β	DNA repair
Y	Bacterial DNA polymerase IV	DNA repair
Reverse transcriptase (RT)	Retroviral reverse transcriptase, Telomerase RT	RNA-dependent DNA polymerase

In addition to polymerization activity, some DNA polymerases also possess proofreading activity via 3’ to 5’ exonuclease domains. This feature enhances replication fidelity by removing incorrectly incorporated nucleotides during DNA synthesis (Eun, 1996; Hamilton et al., 2001). Protein sequence alignment and structural analysis suggest that the 3’ to 5’ exonuclease active site is highly conserved in such polymerases across prokaryotic and eukaryotic organisms (Hamilton et al., 2001; Ishino et al., 1994; Joyce & Steitz, 1994). Although the exonuclease catalytic site is conserved, the spatial arrangement between the exonuclease and polymerase active site can vary between polymerase families (del Prado et al., 2018). Hydrolysis of an incorrectly incorporated nucleotide occurs at the exonuclease active site, and the 3’-end of the growing strand must therefore be transferred from the polymerase active site to the exonuclease active site (Figure 3.1 B) (del Prado et al., 2018). This transfer can in some polymerases occur intramolecularly, without any dissociation of the polymerase/DNA complex. In other polymerases, the exonuclease activity requires a previous dissociation of the polymerase/DNA complex and then binding of the 3’-terminus at the exonuclease active site, a process called intermolecular proofreading (del Prado et al., 2018).

DNA polymerases may also possess 5’ to 3’ exonuclease activity. *In vivo* this is used during lagging strand synthesis of replication and upon removal of RNA primers from Okazaki fragments (Willey et al., 2019). When nicks are present in the DNA strand, the polymerase adds nucleotides to the template while the 5’ to 3’ exonuclease simultaneously removes nucleotides ahead of the polymerase through hydrolysis of phosphodiester bonds, facilitating strand displacement synthesis (Hamilton et al., 2001).

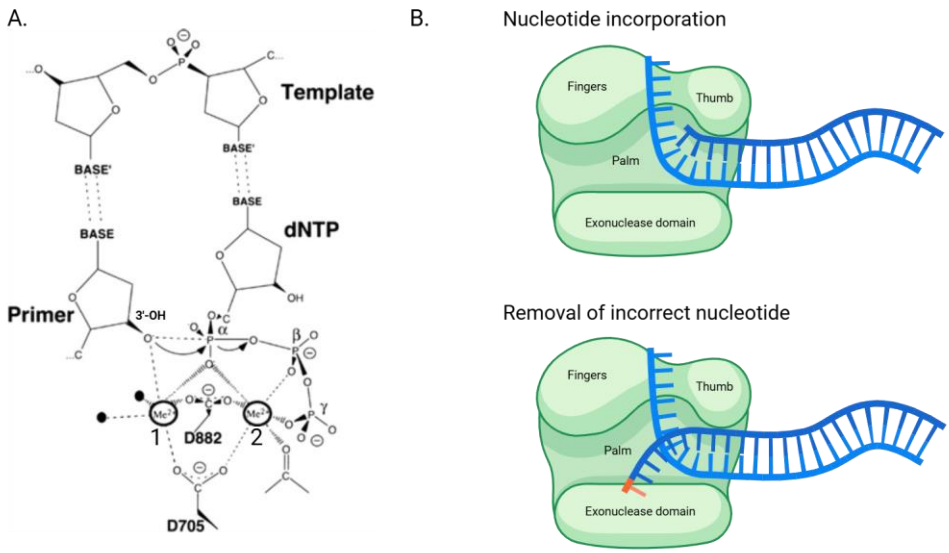


Figure 3.1: DNA synthesis by DNA polymerases. **A**) The two-metal ion mechanism of DNA polymerase marked as 1 and 2. Adapted from (Steitz, 1999). **B**) Model for polymerase DNA synthesis editing via the exonuclease domain.

An important characteristic of DNA polymerases used for *in vitro* DNA amplification in PCR applications is the thermostability (Hamilton et al., 2001). This makes them able to withstand the elevated temperatures of PCR. For this purpose, DNA polymerases have been isolated from thermophilic microorganisms, including thermophilic bacteria such as *T. aquaticus (Taq)*, *Thermus thermophilus (Tth)*, and *Geobacillus stearothermophilus (Bst)*, as well as hyperthermophilic archaea such as *P. furiosus (Pfu), Thermococcus kodakarensis* (*Pfx*/KOD) and *Thermococcus litoralis* (*Tli*/Vent) (Akram et al., 2023; Ishino & Ishino, 2014). These microorganisms have been isolated from high temperature environments like hot springs, deep sea vents, volcanically heated grounds and hydrothermal sediments (Ishino & Ishino, 2014). Molecular features contributing to the thermostability of these DNA polymerases include an increased number of ionic

interactions and salt bridges, greater hydrophobic core packing, shorter and more rigid surface loops, higher proportions of proline and arginine residues, and efficient DNA binding (Borhani & Arab, 2023; Yip et al., 1995). The thermostability of polymerases is primarily due to an increased number of non-covalent interactions, such as salt bridges and hydrogen bonds (Al-Mansoori et al., 2025). More ion pairs and salt bridges in the proteins strengthen the intramolecular cohesion and help resist thermal degradation (Al-Mansoori et al., 2025; Borhani & Arab, 2023). Enhanced packing of hydrophobic amino acid side chains in the core of the enzyme minimizes unfolding at high temperatures (Borhani & Arab, 2023). Structural modifications with shorter and more rigid surface loops reduce the flexibility and disorder that otherwise can lead to thermal unfolding of the protein and rigidity provided help maintaining the enzyme function (Rahban et al., 2022; Yip et al., 1995). Proline and arginine residues increase the protein's rigidity as an additional stabilization (Panja et al., 2020; Yip et al., 1995). DNA binding at high temperatures helps to maintain an active polymerase conformation which supports the activity under heat stress (Brown & LiCata, 2013; Datta & LiCata, 2003). Together, these features give the DNA polymerases their important and useful thermostable properties (Akram et al., 2023).

DNA polymerase-based methods in biotechnology

Since the invention of PCR by Nobel Prize recipient Kary B. Mullis in 1983, continuous development and new applications of PCR have emerged (Mullis, 1990; The Royal Swedish Academy of Sciences, 1993). PCR has revolutionized molecular research, diagnostics, forensics and biotechnology by enabling rapid and precise amplification of DNA (Dewangan et al., 2020). In molecular biology research, it supports genetic cloning, sequencing, and the analysis of gene expression among numerous other applications (Kaunitz, 2015; Pelt-Verkuil et al., 2008). In medicine, PCR enables precise identification of pathogens, characterization of microbial communities such as the gut microbiota, and the detection of genetic mutations, polymorphisms, and tissue types (Sibley et al., 2012). Applications of PCR also include DNA profiling that has transformed forensic science, and it has enabled the recovery and amplification of degraded DNA which has opened new opportunities in molecular paleontology, allowing analysis of specimens preserved as fossils and dried samples (Kaunitz, 2015).

Polymerase chain reaction

PCR provides exponential *in vitro* amplification of specific nucleotide sequences (Mullis & Faloona, 1987). Reagents required for a PCR reaction are template DNA, target-specific primers, DNA polymerase, dNTPs, divalent ions (Mg^{2+}), and a buffer

with suitable pH and ion content (Lorenz, 2012; Pelt-Verkuil et al., 2008). Thermal cycling is performed in a thermocycler starting with denaturation of the dsDNA, usually at around 95 °C. After the initial denaturation, repeated temperature cycles of denaturation, annealing of primers (temperature dependent on the melting temperature (Tm) of the specific primers) and polymerase-based DNA extension (elongation, usually at 72 °C). A final extension step is sometimes added to reach full adenylation of the amplicons. Several reaction parameters can be modified and optimized for the specific assay, such as the concentrations of the included reaction ingredients, the number of cycles and the annealing temperature (Dewangan et al., 2020).

The introduction of the thermostable *Taq* DNA polymerase in PCR made a great impact as it withstands the heat during the denaturation step and circumvents the need for addition of new polymerases in each cycle (Ishino & Ishino, 2014). *Taq* DNA polymerase shows the highest activity at 80 °C but still retains some activity at room temperature (Chien et al., 1976), which can lead to unspecific product formation. To overcome this, DNA polymerases that are inactive at room temperature have been developed (Ishino & Ishino, 2014). Hot-start DNA polymerases are designed to prevent non-specific amplification and primer-dimer formation by remaining inactive at room temperature. Activation occurs after a high-temperature step during PCR. This is commonly achieved by antibody-mediated inhibition or through chemical cross-linking (Paul et al., 2010). In antibody-mediated inhibition, a specific antibody or a smaller affibody blocks the polymerase activity until it is denatured by heat (Kellogg et al., 1994). In chemical cross-linking, the polymerase is temporarily inactivated by a reversible chemical modification that is removed during the initial denaturation step of the PCR (Birch et al., 1996).

Real-time quantitative PCR (qPCR) is a technique used for detection and quantification of nucleic acids (Heid et al., 1996). In qPCR, fluorescence detection is utilized to monitor the increase in nucleic acid amplicons in real time (Dymond, 2013). One of the main advantages with qPCR compared to conventional PCR is the immediate generation of results which removes the need for post-amplification analysis, reducing the risk for contamination due to manipulation of amplicons (Kaltenboeck & Wang, 2005). The fluorescence is measured after each amplification cycle and the general approaches for detection are DNA-binding fluorescent dyes or fluorophore-labeled oligonucleotide probes (Navarro et al., 2015). Amplicon detection using DNA-binding dyes relies on the increase in fluorescence when dsDNA is produced during the reaction, and the generation of the correct amplicons may be determined using melting curves (Mao et al., 2007; Wang et al., 2006). Sequence-specific detection of amplicons can be achieved by using fluorescently labelled probes that are either attached to the primers (e.g., hairpin primer-probes) or designed to anneal to the target amplicon (e.g., hydrolysis probes) (Navarro et al., 2015). The 5’ to 3’ exonuclease activity of for example *Taq*

DNA polymerase is used in hydrolysis probe assays (Dymond, 2013). In **Paper III**, I evaluated four in-house produced DNA polymerases using qPCR with the DNA-binding dye EVAGreen as well as a *Taq* polymerase with a hydrolysis probe. These polymerases were demonstrated to have the desired activity and purity for a range of molecular biology applications, including conventional PCR and qPCR. Quantification by qPCR is based on determining cycle of quantification (Cq) values from the amplification curves and the construction of a standard curve with reference material of known DNA quantity (Heid et al., 1996). The standard curve can then be used to determine the DNA concentration of unknown samples.

Development of new PCR applications since its invention includes multiplex PCR, which consists of more than one target in a single reaction, enabling simultaneous detection of several targets (Chamberlain et al., 1988; Henegariu et al., 1997). Multiplex PCR assays are more complex to develop and require careful primer design, adjusted primer concentrations and optimization of the reaction components and thermal cycling conditions to ensure even amplification of all targeted PCR products (Butler, 2005; Henegariu et al., 1997; Hill et al., 2009). Microsatellites or Short Tandem Repeat (STR) markers are used for forensic DNA profiling, applying multiplex PCR and capillary electrophoresis (CE) for detection (Jobling & Gill, 2004; Sparkes et al., 1996). Analysis with CE yields an electropherogram (EPG), where fragments are differentiated based on size and labelled with different fluorophores for detection (Butler et al., 2004). The intensity of each peak in the EPG reflects the amount of PCR product. However, STR profiling with PCR and CE can be limited when analyzing low-template or degraded DNA samples, complex mixtures or inhibitory sample matrices, where allele drop-out, peak imbalance and stochastic effects sometimes complicate interpretation (Dierig et al., 2024; Pedroza Matute & Iyavoo, 2025).

STRs are a common type of polymorphism in the human genome, which consists of 1 – 6 bases that are repeated a certain number of times (Gymrek, 2017; Shi et al., 2023). The number of repeats in a specific STR marker differs between individuals and can thus be used to generate DNA profiles with extremely low random match probabilities (Weber & Wong, 1993). The repetitive sequence of STR markers makes them highly susceptible to slipped strand mispairing in PCR, resulting in artefacts commonly referred to as stutters (Hauge & Litt, 1993; Meldgaard & Morling, 1997; Walsh et al., 1996). Stutter artefacts are problematic, as they may conceal the detection of minor contributors in crime scene stains with DNA from multiple contributors, so called mixtures (Agudo et al., 2022; Meldgaard & Morling, 1997).

Massively parallel sequencing

Massively Parallel Sequencing (MPS) has played an important role in molecular biology, by enabling cost-effective and fast analysis of microbial genomes as well

as metagenomic studies and sequence and mutation analysis in diagnostics (Biancalana & Laporte, 2015; Bruijns et al., 2018). MPS is employed in several new applications in forensic genetics, such as phenotype prediction, enhanced mixture analysis and body fluid identification (Alonso et al., 2018; Bruijns et al., 2018). Several targeted MPS kits for multiplex forensic STR analysis are commercially available (Bruijns et al., 2018; Dierig et al., 2024; Pedroza Matute & Iyavoo, 2025). One of the major advantages of using MPS in forensic STR analysis is the increased resolution, going from fragment length determination to sequence level analysis. Additional variants can be identified by analyzing the sequence of each allele (Pedroza Matute & Iyavoo, 2025). STR alleles with the same length can thus be differentiated by sequencing, while indistinguishable by CE (Alonso et al., 2018; de Knijff, 2019). Sequence variants in the flanking region of STRs can further increase the discrimination power (Devesse et al., 2020). However, this increased resolution may not provide the expected increased evidential strength (Staadig & Tillmar, 2019). Still, sequence data enables the differentiation between stutter artefacts and alleles of minor contributors in complex DNA mixtures (de Knijff, 2019). Additionally, MPS allows for simultaneous analysis of high numbers of loci with small fragment sizes since it is not dye- or fragment length-limited like CE, which is advantageous for example when dealing with degraded DNA samples (Alonso et al., 2018; de Knijff, 2019; Pedroza Matute & Iyavoo, 2025).

SiMSen-Seq MPS and UMIs in STR analysis

During PCR, errors such as base substitutions, and insertions and deletions may be introduced and accumulate over successive cycles (Potapov & Ong, 2017). Stutter artefacts also occur when analyzing repetitive sequences such as forensic STR markers. The use of unique molecular identifiers (UMIs), nucleotide barcodes added to the DNA, allows reads originating from the same original molecule to be grouped into a consensus read after sequencing (Kinde et al., 2011; Staadig et al., 2023; Ståhlberg et al., 2016). Barcodes can be ligated to the target DNA prior to selection and amplification, or incorporated during the initial PCR cycles, followed by a second adaptor PCR to increase yield and incorporate sequencing adaptors (Filges et al., 2019; Sloan et al., 2018). Post sequencing, the UMI information is utilized to correct for incorrect sequences. This is done by sorting all sequences containing the same UMI into groups and generating a consensus read, often based on the most common sequence variant. However, if the amplification artefact is introduced in the initial PCR cycles, an incorrect consensus read may be obtained. In forensic DNA analysis, UMIs have been applied for both single nucleotide polymorphism (SNP) and STR analysis (Sidstedt et al., 2024; Staadig et al., 2023; Woerner et al., 2021).

The UMI-based library preparation method Simple, Multiplexed, PCR-based barcoding of DNA for Sensitive mutation detection using Sequencing (SiMSen-

Seq) has previously been shown to result in a ten-fold reduction of PCR errors in STR analysis (Sidstedt et al., 2024). Both random base-substitutions and systematic stutter artefacts were reduced.

SiMSen-Seq includes two consecutive PCRs, a barcoding PCR followed by adaptor PCR (Figure 3.2). In the barcoding PCR, a combination of low primer concentrations, low amounts of DNA polymerase, increased annealing and extension times and four PCR cycles help to minimize errors while tagging the copies of each original template molecule with UMIs. The UMIs in the primers are protected by a stem-loop structure to avoid interaction with the template during annealing (Ståhlberg et al., 2016). The initial barcoding PCR uses genomic DNA as input template (Figure 3.2). There, three different polymerization processes give rise to the targeted short products labelled with UMIs: 1) generation of long products defined by the primer in one end, from genomic DNA, 2) generation of short product STR amplicons from long products and 3) generation of STR amplicons from other short amplicons. These three events have different efficiencies due to the different templates of different sizes, spanning from genomic DNA to short products. In regular PCR, such as qPCR, only the last process may be monitored. SiMSen-Seq, however, makes it possible to study the performance in the initial PCR cycles separately.

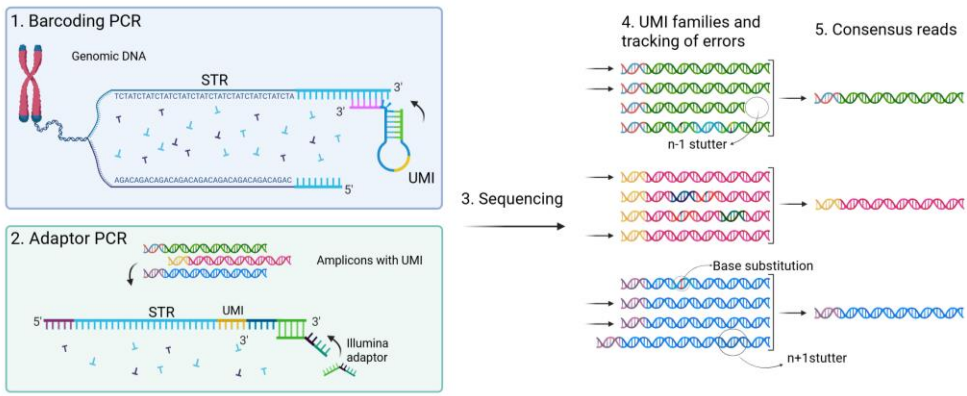


Figure 3.2: Schematic overview of the SiMSen-Seq STR method for construction of UMI families and generation of consensus reads. The correct molecules are indicated with an arrow in step 4. From Paper IV.

In the adaptor PCR, each UMI-labelled short product from barcoding PCR will generate a UMI family with reads having the same UMI or barcode (Figure 3.2). All products from the barcoding PCR are transferred to the adaptor PCR, and only the short products will serve as template for the adaptor primers used (Figure 3.2). Thus, only the third process will take place, generation of short products from other short products, and any erroneous sequences formed emanate from this process. The

consensus read is generated from the most common sequence of all STR reads with the same UMI. This will remove most of the errors introduced in the adaptor PCR.

By using one DNA polymerase as a reference in barcoding PCR and adaptor PCR (i.e., SuperFi II), respectively, any differences between the DNA polymerases with different properties will be connected to events happening in either of the two PCRs (**Paper IV**). Thus, it is possible to determine if the DNA polymerases are prone to form single-base substitutions and stutters in the initial PCR cycles, using genomic DNA as template, or in the later cycles, using short products as template. Using regular PCR would not enable this distinction, even if the PCR was split into two distinct thermal cycling programs, since, without the UMI approach, all three types of template molecules from the first PCR would serve as template for the second and it would not be possible to trace errors back to individual template molecules.

In **Paper III**, I used the in-house synthesized DNA polymerases *Taq*, *Taq*-Sso7d, *Pfu* and *Pfu*-Sso7d in SiMSen-Seq library preparation (Sidstedt et al., 2024). All four polymerase variants demonstrated the desired activity and enabled successful adaptor PCR amplification of STR markers, with similar performance between the polymerases (see Chapter 5). Applying UMIs and generating consensus reads reduced the proportion of stutter artefacts to around a third compared to before consensus read generation.

In **Paper IV**, I evaluated the effects of commercial DNA polymerases with different characteristics regarding amplicon yield and PCR error formation, considering both random base-substitutions and systematic stutter artefacts, in STR analysis (see Chapter 5). The SiMSen-Seq MPS library preparation method with UMIs (Sidstedt et al., 2024; Ståhlberg et al., 2016) was used to study polymerization errors emanating from individual DNA molecules and to study events in the early and late PCR cycles separately.

The SiMSen-Seq library preparation was successful for all polymerases, both in-house produced and commercial, when applied in the adaptor PCR (**Paper III** and **IV**). A reduction of both single-base substitutions and stutter artefacts was seen for all DNA polymerases when generating consensus reads. The challenging conditions in the barcoding PCR resulted in low yields of STR amplicons for some of the applied polymerases (**Paper IV**). By studying UMI families, errors emanating from individual genomic DNA template molecules were investigated. The usage of SiMSen-Seq and UMIs also helped to elucidate at what stage of the PCR the errors were formed: if the barcoding provided the correct sequence, then any errors noted emanated from the adaptor PCR.

The results presented in this thesis and in **Paper III** and **IV** were obtained by employing a 7-plex STR assay. However, to reach maximum impact and to be of practical use in forensic DNA analysis, a larger multiplex is desired. Using more DNA markers in forensic DNA analysis increases the chances of obtaining high likelihood ratios or low random match probabilities, which is especially important

for mixtures or when distinguishing between closely related individuals (Bille et al., 2013; Karlsson et al., 2007; Martín et al., 2014). While seven STR markers may be sufficient for single-source samples from unrelated individuals, more markers are needed to achieve reliable results in complex mixtures, or when potential contributors are relatives. Statistical power and discrimination capacity improve as additional markers are included (Bandah-Rozenfeld et al., 2025; Bleka et al., 2020; Navarro-López et al., 2022).

Initial experiments with an early-stage prototype 28-plex SiMSen-Seq STR assay, developed by Simsen Diagnostics AB, were performed, and promising results were obtained. The prototype includes the same markers as the ForenSeq MainstAY kit (Qiagen) (Stephens et al., 2023). Five different single-source DNA samples were analyzed, and 5 ng DNA was used as template. In the initial trials, 26 out of the 28 STR markers were detected with correct genotypes. However, the number of reads and consensus reads varied substantially between the markers, from only a few consensus reads up to more than 1400 consensus reads. This suggests that further optimization of the library preparation is needed. Regarding the two markers that were not detected, the reason why needs to be investigated. It could be due to the primer design or primer concentration, or due to how the data is bioinformatically handled including how the STR markers are defined in the pipeline. Continued efforts are ongoing to refine and optimize the assay for improved performance and reliability.

Produced and evaluated DNA polymerases

The DNA polymerases evaluated for in-house production and characterization in this thesis were *Taq* and *Pfu* polymerases including variants fused with the dsDNA-binding domain Sso7d (**Paper III**). *Taq* polymerase, from the thermophilic bacterium *T. aquaticus*, belongs to Family A and has 5'-3' exonuclease activity apart from 5'-3' polymerase activity (Chien et al., 1976; Eckert & Kunkel, 1990; Tindall & Kunkel, 1988). *T. aquaticus* was first isolated from hot springs in Yellowstone National Park. It is a Gram-negative rod and filamentous bacterium prone to aggregation that can form larger round bodies consisting of multiple cells connected by a combined outer envelope (Brock & Edwards, 1970; Brock & Freeze, 1969; Brumm et al., 2015). Its optimal growth temperature is between 65 °C and 70 °C, but it can tolerate temperatures between 40 °C and 80 °C (Brock & Freeze, 1969; Brumm et al., 2015). *Taq* polymerase was the first thermostable DNA polymerase to be discovered and characterized, and variants thereof remain among the most widely used DNA polymerases in PCR (Ishino & Ishino, 2014).

Pfu polymerase, from the hyperthermophilic archaeon *P. furiosus* belongs to Family B and has 3'-5' exonuclease activity (proofreading) and thus generates fewer base

substitution errors compared to *Taq* polymerase (Cline et al., 1996; Lundberg et al., 1991). *P. furiosus* was first isolated from geothermally heated marine sediments near volcanic vents at Vulcano Island, Italy (Fiala & Stetter, 1986). It is an irregular coccus with multiple flagella, giving it high motility. Its optimal growth temperature is around 100 °C, one of the highest known for life (Fiala & Stetter, 1986; Yip et al., 1995). Due to the high optimal growth temperature, proteins produced by *P. furiosus* are extremely thermostable (Farooq et al., 2024; Yip et al., 1995). *Pfu* polymerase and engineered variants are also commonly used in PCR, as well as mixtures of *Taq* and *Pfu* polymerases (Farooq et al., 2024).

Several strategies have been developed to engineer DNA polymerases to enhance their fidelity and processivity (Coulther et al., 2019; Ishino & Ishino, 2014). Fidelity is frequently expressed as the inverse of the error rate (fidelity = 1/error rate) and refers to the number of mis-incorporated nucleotides per total number of nucleotides polymerized. The fidelity of a polymerase is often expressed in relation to the fidelity of *Taq* DNA polymerase. Proofreading DNA polymerases like *Pfu* and KOD (from *T. kodakarensis*) exhibit approximately tenfold higher fidelity compared to *Taq* polymerase (Ishino & Ishino, 2014). The processivity of DNA polymerases is defined as the number of nucleotides incorporated in one binding event. One of the simplest approaches to increase processivity involves fusing the polymerase with a non-specific DNA-binding domain (Olszewski et al., 2017; Wang et al., 2004). For example, the dsDNA-binding Sso7d domain from *S. solfataricus* has been shown to enhance processivity (Farooqui et al., 2023; Wang et al., 2004). The Sso7d domain binds to dsDNA and stabilizes the polymerase–DNA complex, thereby reducing the likelihood of enzyme dissociation during amplification, while largely preserving the polymerase's catalytic properties (Farooqui et al., 2023; Ishino & Ishino, 2014). When Sso7d was fused to *Pfu* polymerase, the processivity was increased 9-fold, from an average of 6 nucleotides to 55 nucleotides per binding event (Fazekas et al., 2010; Wang et al., 2004).

The characteristics of these DNA polymerases and the ability to specifically catalyze DNA synthesis have made them valuable and well established as biotechnological tools.

In **Paper IV**, I used six different commercial DNA polymerases with distinct characteristics (see Table 2 in **Paper IV**). Three of these belong to family A and do not have 3' to 5' exonuclease activity. However, two of these are mixed with exonucleases to provide proofreading. The fidelity of the applied family A polymerases ranges from 1X to 9X compared to *Taq* DNA polymerase. The other three belong to family B and are proofreading polymerases. The fidelity of the included family B polymerases ranges from 26X to above 300X compared to *Taq* DNA polymerase.

The selected DNA polymerases represent a diverse set of enzymes with varying biochemical properties. Platinum SuperFi II and Phusion HS II are high-fidelity,

proofreading polymerases with 3' to 5' exonuclease activity and DNA-binding domains, with SuperFi II offering over 300-fold higher fidelity than *Taq* and Phusion HS II offering 52X fidelity compared to *Taq*, designed to minimize base substitution errors and improve accuracy and processivity. AccuPrime *Pfx* is a recombinant DNA polymerase with 3' to 5' exonuclease activity (26X fidelity compared to *Taq*), with proprietary accessory proteins to enhance primer-template hybridization. AccuPrime *Taq* High Fidelity combines *Taq* DNA polymerase with a proofreading enzyme and accessory proteins for improved PCR fidelity, yield, and specificity, offering a 9-fold higher fidelity compared to *Taq*. Ex*Taq* HS is a blend of *Taq* polymerase and a proofreading exonuclease providing 4.5X fidelity compared to *Taq*. Immolase is a novel, *Taq*-like family A DNA polymerase from an undisclosed organism. It lacks proofreading activity and is the only chemically inactivated hot start polymerase used. All other DNA polymerases have antibody or affibody mediated heat activation. All six polymerases are designed to require hot start activation to avoid unspecific product formation at room temperature.

Chapter 4
Challenges in recombinant enzyme production

Recombinant enzymes are fundamental in modern biotechnology, supporting applications ranging from pharmaceuticals and diagnostics to production of sustainable foods and research (De Brabander et al., 2023). However, achieving efficient and scalable production of functional enzymes presents several challenges, including protein misfolding, aggregation and inclusion body formation, insufficient post-translational modifications (PTMs), proteolytic degradation, low yield, and difficulties in recovering biologically active proteins (Beygmoradi et al., 2023; Bhatwa et al., 2021). The selection of appropriate host organisms, expression systems, vectors, and induction mechanisms is crucial to overcoming these challenges and optimizing protein quality and quantity (De Brabander et al., 2023; Palomares et al., 2004). There is also an increasing demand for resilience and civil contingency and a growing need for resource-efficient and cost-effective production pipelines for in-house production of molecular biology reagents. Availability of such simple and efficient protocols reduces dependency on commercial supply chains and enhances preparedness for situations where access to critical resources may be limited. Apart from production, cellular systems can also be used to study proteins, their functions, their effects on the host cell and for evaluation and development of new proteins. This chapter explores host organisms including tools for genetic manipulation, expression systems and vectors, and induction mechanisms, focusing on recombinant production of DNA polymerases and proteases.

E. coli remains one of the most extensively used microbial systems for heterologous protein production. Recombinant DNA polymerases are currently produced in *E. coli* and well-established production methods exist. However, challenges with complex production pipelines as well as proprietary formulations persist. Other challenges, like residual host DNA in the final DNA polymerase preparation can influence the usability of the produced DNA polymerase. In **Paper III**, I developed a simplified and resource-efficient *E. coli*-based production protocol and characterized the performance of the synthesized polymerases using several PCR-based assays.

S. cerevisiae has previously not been extensively studied as a production host for bacterial proteases. However, its well-characterized genetics and robust fermentation capabilities make it a suitable alternative for synthesis of proteases. In **Paper I** and **II** I investigated the applicability of *S. cerevisiae* for production of three different proteases. Additionally, I studied how bacterial proteases interact with a eukaryotic production host and how the proteases affect yeast physiology.

Host organisms for protein production

Choosing the proper production host for recombinant enzymes is an important part of developing a production process. Microbial hosts have different characteristics that need to be assessed depending on the desired product (Overton, 2014). Well-characterized microbial hosts and model organisms are commonly used in industrial production processes, e.g., *E. coli*, *Bacillus subtilis*, *S. cerevisiae*, *Aspergillus* spp. and *Komagataella phaffii* (previously *Pichia pastoris*) (Contesini et al., 2018; Johnson & Echavarri-Erasun, 2011; Ward, 2011). Advantages with microbial production systems are rapid growth to high cell densities with comparably low-cost and simple media, advancements in process scale up and high protein titer, rate, and yield compared to insect and mammalian cells (Karbalaei et al., 2020).

E. coli as a production host

E. coli is one of the most widely used hosts for recombinant protein production due to its fast growth and ability to reach high cell densities in simple and inexpensive media (Baneyx, 1999; Baneyx & Mujacic, 2004). *E. coli* is Gram-negative and one of the most extensively studied bacteria, both physiologically and metabolically, and was among the first organisms to have its genome sequenced (Blattner et al., 1997; Overton, 2014). It is typically the first choice for recombinant protein production, independent of the protein's origin (Baneyx, 1999; Baneyx & Mujacic, 2004). Some non-pathogenic strains are recognized as universally accepted commercial manufacturing hosts with a long history of safe use. These strains are considered as safe for industrial use under controlled conditions and compounds produced by these strains are considered safe by agencies such as the European Food Safety Authority (EFSA), the U.S. Environmental Protection Agency (EDA) and Food Standards Australia New Zeeland (FSANZ). Also, decades of optimization have established *E. coli* as a robust platform for large-scale bioprocessing (Huang et al., 2012; İncir & Kaplan, 2024). Genetic engineering is facilitated by a broad repertoire of molecular tools, ranging from classical plasmid-based expression systems to advanced genome editing strategies, including CRISPR/Cas technologies (İncir & Kaplan, 2024; Pontrelli et al., 2018).

Specific *E. coli* strains have been optimized for recombinant protein production. BL21 (DE3) combined with a multi-copy plasmid, is the preferred host strain for recombinant protein production (Heyde & Nørholm, 2021). The strengths of BL21 (DE3) include fast growth, diminished acetate production and deficiency in key proteases which minimizes proteolysis of heterologous proteins (Rosano et al., 2019). Furthermore, it harbors a phage T7 RNA polymerase gene, under control of the *lac*UV5 promoter. Heterologous genes on plasmids are controlled by the T7 promoter and gene expression is started upon addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) (see Induction mechanisms) (Rosano et al., 2019). Advancements in strain engineering include the addition of the pLysS plasmid, containing the bacteriophage T7 lysozyme gene, that reduces basal expression from the T7 promoter (Rosano et al., 2019). An additional advantage of the T7 lysozyme is its ability to facilitate the cell lysis required for protein purification (Studier, 1991).

The Rosetta (DE3) strain is derived from BL21 (DE3) and has additional rare tRNA genes to correct for codon bias (Rosano et al., 2019). Limitations with *E. coli* as a production host include the lack of PTMs and possibilities of large fractions of insoluble proteins due to protein misfolding and the formation of inclusion bodies (Baneyx & Mujacic, 2004; Bhatwa et al., 2021). However, inclusion body formation may also be utilized as a method to produce high quantities of pure proteins that are purified and later refolded into active enzymes (Clark, 2001; Singh et al., 2015). In this work, *E. coli* BL21 (DE3) and Rosetta (DE3) pLysS were used to produce recombinant DNA polymerases (**Paper III**). BL21 (DE3) was selected as the starting point for polymerase production and Rosetta (DE3) pLysS was then used to obtain tighter control of gene expression as basal expression of *Pfu* polymerase previously has been shown to be cytotoxic (Lu & Erickson, 1997) and to compensate for rare codons in the gene cassettes. Substantially higher concentrations of DNA polymerases were obtained when *E. coli* Rosetta (DE3) pLysS was used compared to BL21 (DE3) (**Paper III**).

S. cerevisiae as a production host

S. cerevisiae, or Baker's yeast, is widely used for recombinant protein production due to its Generally Recognized As Safe (GRAS) status, its well-characterized large-scale physiological behavior and its robustness and tolerance to industrial conditions (Han et al., 2018; Hou et al., 2012; Johnson & Echavarri-Erasun, 2011; Myburgh et al., 2023; Wang et al., 2017). It is a well-established model organism with a long history of use in wine, beer and bread making as well as for protein production. Being a eukaryote, it is capable of carrying out some PTMs, which are essential for the correct production of proteins that depend on them (Ferrer-Miralles et al., 2009; Johnson & Echavarri-Erasun, 2011). Additionally, *S. cerevisiae* offers genetic stability and ease of genetic engineering, with a range of tools and plasmid-

based systems for efficient genetic manipulation, including CRISPR/Cas9 (Borodina & Nielsen, 2014; Xie et al., 2018). Enhanced gene editing efficiency is achieved through a robust cellular DNA repair system, which facilitates precise genome editing (Haber, 2018). Limitations of *S. cerevisiae* include lower protein yield compared to *E. coli* as well as hyperglycosylation and a different glycosylation pattern compared to mammalian systems (Ferrer-Miralles et al., 2009). Until now, there has been limited research on production of bacterial proteases in yeast. Here, I used *S. cerevisiae* as a eukaryotic model system to study protein production, protease effects on the host cell and the protease substrate scopes that can be handled by *S. cerevisiae* (**Paper I** and **II)**. The bacterial proteases produced in *S. cerevisiae* resulted in different burdens on the cells (see Selection markers and copy number effects below).

CRISPR/Cas9

In this work, CRISPR/Cas9 was used to integrate heterologous protease genes into the genome of *S. cerevisiae* (**Paper I** and **II**). CRISPR/Cas technologies have been engineered and optimized for applications and genetic manipulation across a diverse range of microbial species. Gene editing tools such as the Cas9 endonuclease enable precise genome modifications by inducing double-stranded DNA breaks (DSBs), but the actual insertion or deletion of genetic material depends on the cell's endogenous DNA repair pathways (Figure 4.1) (Rainha et al., 2021). Guided by a single RNA (guidingRNA, short gRNA) composed of fused CRISPR RNA (crRNA) and transactivating crRNA (tracrRNA), Cas9 introduces a DSB three nucleotides upstream of a protospacer adjacent motif (PAM), typically NGG (Stovicek et al., 2017). For cell survival, the DSB must be repaired and in *S. cerevisiae*, DSBs are primarily repaired by non-homologous end joining (NHEJ) or homology-directed repair (HDR) (Haber, 2018). Unlike many other yeasts, *S. cerevisiae* exhibits particularly high HDR activity, which facilitates efficient gene replacement using homologous DNA sequences provided as linear fragments or integrative plasmids (Cai et al., 2019; Haber, 2018). The CRISPR/Cas9 system has further expanded the possibilities for yeast genome engineering, enabling simultaneous modifications at multiple loci (Stovicek et al., 2017). The main advantages of CRISPR/Cas9 are its precision, multiplexing capacity, and independence from selective markers (J. Zhao et al., 2020). The gene cassette of interest can be inserted in an intergenic region of the genome to avoid effects on the growth rate using CRISPR/Cas9 (Jessop-Fabre et al., 2016). The integration sites are designed not to interfere with any neighboring genes and are located at highly expressed parts of the genome, making the DNA easily accessible by the RNA polymerase (Mikkelsen et al., 2012).

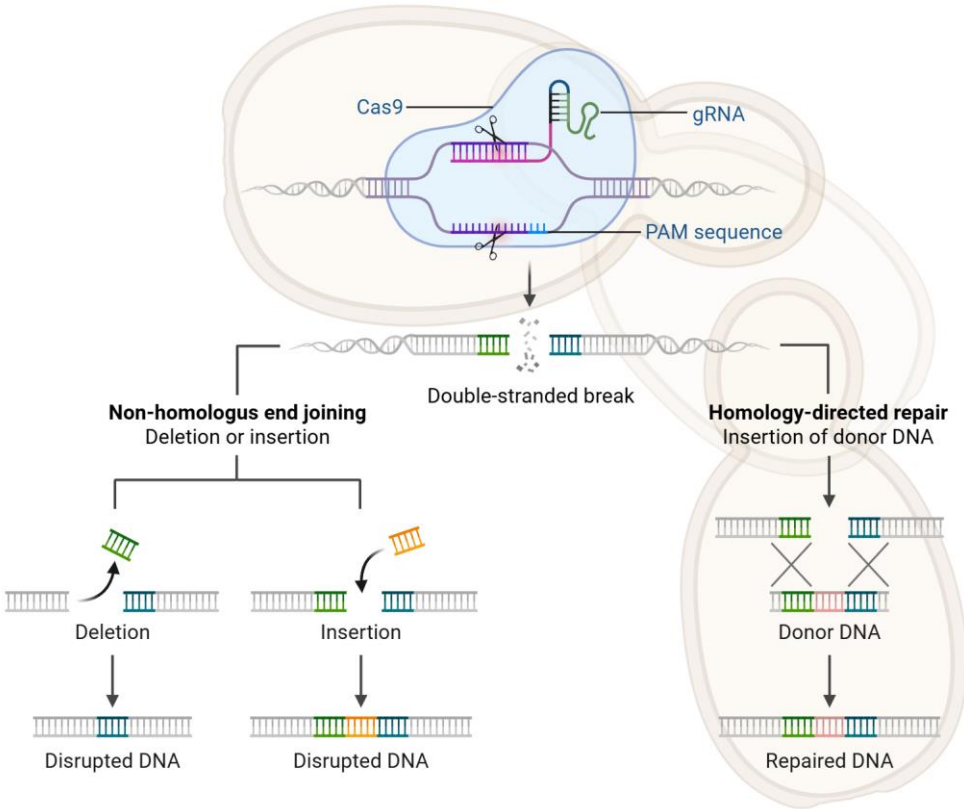


Figure 4.1: Schematic overview of CRISPR/Cas9 gene editing where the double-stranded break (DSB) is repaired either by non-homologous end joining (NHEJ) (left) or homology-directed repair (HDR) (right).

Other production hosts

One strategy to overcome some of the challenges encountered in recombinant enzyme production is to use an alternative production host (**Paper I** and **II**). Two other potential production hosts are *B. subtilis* and *K. phaffii*, with characteristics that may be advantageous in recombinant enzyme production.

B. subtilis is a robust Gram-positive bacterium widely used as a protein production host due to its GRAS and QPS (Qualified Presumption of Safety) status, rapid growth, and natural ability to secrete proteins directly into the culture medium (Liu & Yu, 2025; Su et al., 2020). Its lack of endotoxins and outer membrane simplifies downstream processing and makes it particularly attractive for industrial enzyme production (Su et al., 2020). A broad genetic toolbox, including protease-deficient and secretion-optimized strains, has been developed to improve heterologous protein yields, and genome-reduced variants further enhance secretion efficiency

(Liu & Yu, 2025). However, *B. subtilis* has limitations, including degradation of target proteins by residual proteases, challenges with proper folding of some heterologous proteins, and generally lower efficiency in producing complex eukaryotic proteins requiring PTMs (Liu & Yu, 2025; Su et al., 2020).

K. phaffii is an established eukaryotic yeast host for recombinant protein production thanks to the QPS and GRAS status on products produced in *K. phaffii*, its ability to grow to very high cell densities on inexpensive media, and its efficient secretion machinery that reduces background protein contamination (Bustos et al., 2022; Karbalaei et al., 2020). *K. phaffii* performs PTMs such as disulfide bond formation and glycosylation, making it suitable for the expression of complex proteins (Karbalaei et al., 2020; Spohner et al., 2015; Vijayakumar & Venkataraman, 2024). A range of strong and inducible promoters, together with stable genome integration, supports high-level and reliable expression (Bustos et al., 2022; Spohner et al., 2015). However, its native glycosylation patterns differ from those in humans and can lead to hypermannosylation, which may compromise therapeutic protein applications (Offei et al., 2022; Vijayakumar & Venkataraman, 2024). In addition, some strains produce extracellular polysaccharides that complicate downstream processing (Steimann et al., 2024).

Expression systems and vectors

Recombinant protein production relies on gene expression systems, the choice of which can vary depending on the host organism and the protein to be produced (İncir & Kaplan, 2024). The target gene is commonly expressed from a multi-copy plasmid to increase the protein yield. The choice of expression vector can influence the yield, and optimization is often required to reach necessary production levels (Lozano Terol et al., 2021; Schütz et al., 2023). Other central parts for controlling gene expression and protein production include some typical features of expression vectors, such as a strong constitutive or inducible promoter, a ribosome binding site in prokaryotes, or upstream activating sequences, enhancers or Kozak fragments in eukaryotes, selective antibiotic resistance markers or selective auxotrophic markers, a replication origin (ori) and fusion tags for protein purification, detection or localization.

Promoters, terminators and mRNA stability

The choice of promoter and induction mechanism influences protein yield and host physiology (İncir & Kaplan, 2024). Reasons for promoter selection include maximum expression during exponential growth, maximum expression during stationary phase, constitutive expression of a gene, induced expression of a gene,

controlled expression of a gene, moderate expression of a gene and temporal expression of a gene (Goldstein & Doi, 1995). It is not always suitable to rapidly produce high quantities of recombinant proteins, as this may compromise their quality and function through protein misfolding and aggregation into insoluble inclusion bodies, or lead to toxicity to the host cell, resulting in impaired growth and viability. Together, this lowers the overall protein yield and depletes the host cell's resources (Bhatwa et al., 2021). Proteins requiring post translational modifications will become non-functioning if these modifications are not performed and glycosylation patterns may change under extreme production rates (Beygmoradi et al., 2023; Gawlitzek et al., 1994; Schenk et al., 2008). Different promoters can be used to regulate protein production rate and timing. In **Paper I** and **II**, I used the inducible *GAL1*p promoter to control the time of induction with galactose, to separate the production phase from the growth phase. Recombinant production of cytotoxic proteins, as in the case of certain proteases, is favored by a separation of the growth and production phase. *GAL1*p is repressed by the presence of glucose and induced when galactose is available as the carbon source (Li et al., 2000). In **Paper III**, the polymerase production was regulated by induction with Isopropyl β-D-1-thiogalactopyranoside (IPTG). Both systems use strong promoters that are repressed and when induced result in strong activation of transcription and translation into proteins.

Promoters and transcription into mRNA play an important role in protein synthesis. The protein yield is also influenced by terminators, although this is not as well-studied (Wei et al., 2017). The terminator mainly influences the protein yield by affecting the mRNA stability and half-life, translational efficiency and localization of the mRNA (Ren et al., 2025; Vargas et al., 2023). The choice of terminator may be used to finetune gene expression for specific applications and metabolic engineering approaches. However, standard terminators were used in this thesis work.

After transcription to mRNA, translation is initiated to synthesize the polypeptide chain. In bacteria, the 5′ untranslated region of mRNA contains the Shine-Dalgarno sequence, part of the ribosome-binding site (RBS), located upstream of the start codon (Willey et al., 2019). This sequence pairs with the 16S rRNA of the 30S ribosomal subunit, facilitating the formation of the 70S initiation complex with the help of initiation factors and the initiator tRNA, aligning the ribosome to the start codon AUG (Willey et al., 2019). In eukaryotes, translation initiation is more complex, involving numerous initiation factors and interactions between the 5′ cap, 3′ poly-A tail, and associated proteins, interacting with each other, which folds the mRNA back on itself (Sonenberg & Hinnebusch, 2009; Willey et al., 2019). The 40S ribosomal subunit, bound to initiator tRNA and initiation factors, scans the mRNA for the start codon (Hinnebusch & Lorsch, 2012; Willey et al., 2019). Upon recognition, the 60S subunit joins to form the 80S ribosome, initiating elongation (Sonenberg & Hinnebusch, 2009; Willey et al., 2019). A key regulatory element in

eukaryotic translation is the Kozak sequence, located near the initiation site of translation and influences ribosome recognition and translation efficiency (Kozak, 1984, 1987; Xie et al., 2023). Modifications to the Kozak sequence have been used to fine-tune protein synthesis levels (Li et al., 2017; Xie et al., 2023).

Induction mechanisms

Induction systems are commonly used in recombinant protein production to regulate transcription initiation, and to control the timing and level of gene expression (**Paper I, II** and **III**) (Overton, 2014). Producing the target proteins continuously can be a burden to the host cell or reduce the yield. Induction can be used to keep the gene expression repressed until cell growth has reached a desirable state (Kastberg et al., 2022; Overton, 2014). Induction is then triggered by adding a specific molecule or by shifting growth conditions, enabling strong, coordinated production of the recombinant protein. This separates the growth and production phases, which is important in the case of production of a cytotoxic protein, like broad-spectrum proteases (**Paper II**) (Kastberg et al., 2022). Using induction can improve host viability and maximize protein yield (Bhatwa et al., 2021). Although induction facilitates coordinated protein production, problems with misfolding and aggregation of recombinant proteins can still occur. Mitigation strategies include usage of a weaker promoter, lowered inducer concentration, lowered temperature and prolonged time during induction to allow for slower but correct production and folding of the recombinant protein (Bhatwa et al., 2021).

One of the most used *E. coli* expression systems relies on the inducible T7 RNA polymerase, which enables high protein yields (**Paper III**) (Briand et al., 2016; Tabor & Richardson, 1985). The T7 RNA polymerase gene is integrated into the bacterial chromosome under control of the *lac*UV5 promoter and initially transcribed by the host's native RNA polymerase (Figure 4.2) (Studier & Moffatt, 1986). Expression is tightly regulated by the LacI repressor, which blocks transcription until induction. However, some promoter leakage and basal expression occur (Briand et al., 2016; Studier & Moffatt, 1986). Isopropyl β-D-1-thiogalactopyranoside (IPTG) is a synthetic analog of allolactose, a natural inducer of the *lac* operon in *E. coli* (Simas et al., 2023). Unlike allolactose, IPTG is not metabolized, ensuring a constant inducer concentration during cultivations. When added, IPTG binds to the LacI repressor protein, causing a conformational change that reduces LacI's affinity for the operator region of the *lac* operon (Lewis, 2005). When the LacI repression is relieved, it leads to production of T7 RNA polymerase. The T7 RNA polymerase then drives strong and specific transcription of the target gene, which is placed on an expression plasmid under control of the T7 promoter (Lopez et al., 1998; Tabor & Richardson, 1985). Since the T7 system amplifies transcription via a dedicated RNA polymerase, it enables extremely strong and tightly controlled expression of recombinant genes.

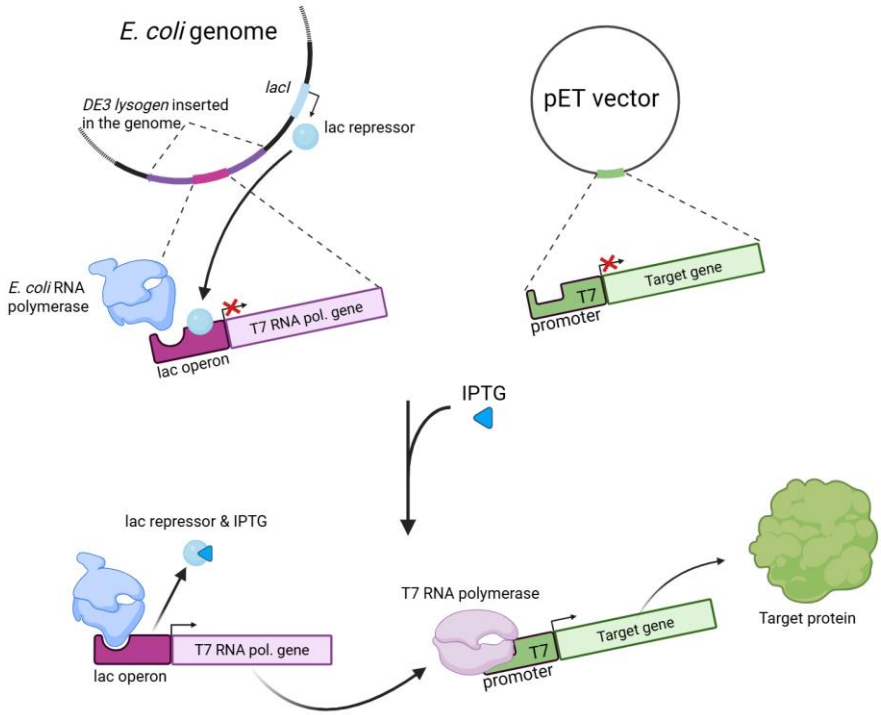


Figure 4.2: Schematic overview of IPTG induction of the T7 RNA polymerase expression system in *E. coli*.

In the yeast *S. cerevisiae,* the regulation of the *GAL* genes has served as an important model for regulation of transcription initiation in eukaryotic organisms (Willey et al., 2019). The *GAL* genes encode proteins required for the uptake and catabolism of galactose (Vollenbroich et al., 1999). Their expression is regulated by the transcriptional activator Gal4, which binds to upstream activating sequences located near each *GAL* gene. The activity of Gal4 is, in turn, inhibited by Gal80, which blocks transcription when galactose is absent or when glucose is present (Willey et al., 2019). In the presence of galactose, Gal3 can bind to galactose and ATP, causing a confirmational change, where Gal3 can move into the nucleus, bind Gal80 and escort it to the cytoplasm, relieving the Gal80 inhibition, allowing Gal4 to activate transcription (Lohr et al., 1995). Gal4 initiates this process by recruiting chromatin remodeling enzymes to expose the promoter region, followed by the recruitment of basal transcription factors (Sellick et al., 2008). Finally, RNA polymerase II binds to the promoter, enabling transcription to begin (Sellick et al., 2008). The regulation of yeast *GAL* genes exemplifies many of the features of eukaryotic gene regulations, including chromatin remodeling, chromatin modification, formation of activation complexes, and the indirect influence of activators and repressors on genes

(Stockwell & Rifkin, 2017). When galactose is the only available carbon source, the enzymes responsible for its metabolism are upregulated by a factor of 1000 compared to their expression in glucose (Lohr et al., 1995), making them some of the most tightly regulated proteins in yeast (Stockwell et al., 2015).

Galactose induction is a powerful and widely used strategy for recombinant protein production in *S. cerevisiae*, exploiting the tightly regulated *GAL1* promoter (*GAL1*p) (Da Silva & Srikrishnan, 2012). This was used in **Paper I** and **II** for separation of the growth and production phase for recombinant protease production. In *S. cerevisiae*, the presence of galactose activates the regulatory network described above which ends with Gal4p driving a strong transcription from *GAL1*p (Sellick et al., 2008). This results in robust production of target proteins under the control of *GAL1*p, typically yielding large amounts of recombinant protein within hours after galactose addition (Da Silva & Srikrishnan, 2012). The induction system allows for minimal background expression in the presence of glucose, providing precise temporal control over protein production (Sellick et al., 2008). The induction speed and tunability make galactose induction an essential tool for molecular biology and biotechnology (Stockwell et al., 2015). Even though *GAL1*p is reported to be tightly regulated, significant growth inhibition was seen among the cells carrying the *speB* gene before induction (**Paper II**). Galactose, in contrast to IPTG, is metabolized by *S. cerevisiae* when used as an inducer and its concentration in the culture will reduce over time (Sellick et al., 2008). However, galactose is metabolized slowly, and I found it to result in a long induction time (>24 hours) and high protein concentrations (**Paper II**). For protein production purposes, this gradual depletion poses no issue if sufficient protein levels are reached before harvest of the culture.

Selection markers and copy number effects

Selection markers are essential tools in molecular biology and genetic engineering, where they are used to identify and isolate successfully transformed cells and make sure that the plasmids are maintained by the cells (Stanbury et al., 2017). Antibiotic selection markers give resistance to specific antibiotics, e.g., ampicillin, kanamycin, or chloramphenicol, only allowing the cells that have incorporated the desired genetic construct to survive in the presence of the antibiotic. Resistance mechanisms can involve the production of enzymes that degrade the antibiotic, modify its target, or prevent its uptake, thereby neutralizing its effect (Muteeb et al., 2023). This method is highly efficient and widely used in both prokaryotic and eukaryotic systems. However, the antibiotic ampicillin can cause allergic reactions in some individuals and its use in production of therapeutic agents is therefore limited as even trace elements need to be removed (Carnes, 2005; FDA, 1998; Jourdan et al., 2020). Other antibiotics, like kanamycin, are often used instead of ampicillin for recombinant protein production. In contrast, auxotrophic selection markers rely on the restoration of a metabolic function in mutant strains that are unable to synthesize

a particular essential compound, e.g., amino acids such as uracil, histidine, or leucine (Stanbury et al., 2017). Cells transformed with a plasmid carrying the corresponding functional gene can grow in media lacking that compound, enabling selection without the use of antibiotics (Pronk, 2002). Auxotrophic markers are useful in yeast and other microbial systems, as an antibiotic-free alternative that can be advantageous in certain regulatory or environmental contexts as well as in scale up efforts (Pronk, 2002). Some antibiotics, especially targeting eukaryotes, are harmful to humans and auxotrophic markers can be a great alternative to those (Böttger et al., 2001; Mingeot-Leclercq & Tulkens, 1999). In **Paper II**, the auxotrophic marker for leucine was used. However, the burden of the plasmid carrying the gene *speB* was high and thus only a few cells in the culture harbored the plasmid, while other cells likely survived on leucine from the plasmid bearing cells or lysed cells. The cell culture with the plasmid carrying the gene *speB* was barely growing, even before induction. This suggests that trace-level transcription occurred, enough to cause the growth inhibition (Figure 4.3 A). Cells producing the other two proteases, BdpK and IdeS, as well as the control with only GFP, had similar growth profiles and resulted in two subpopulations. Only around half of the population produced GFP above autofluorescence levels. The loss of plasmids, even when the recombinant proteins did not pose an extreme burden on the cells, can be explained by cross-feeding of leucine or related biosynthetic intermediates and a decreased selection pressure (Hu et al., 2024; Pronk, 2002). The selection pressure and plasmid copy number can be increased using a deficient promoter like *LEU2*d (Pronk, 2002).

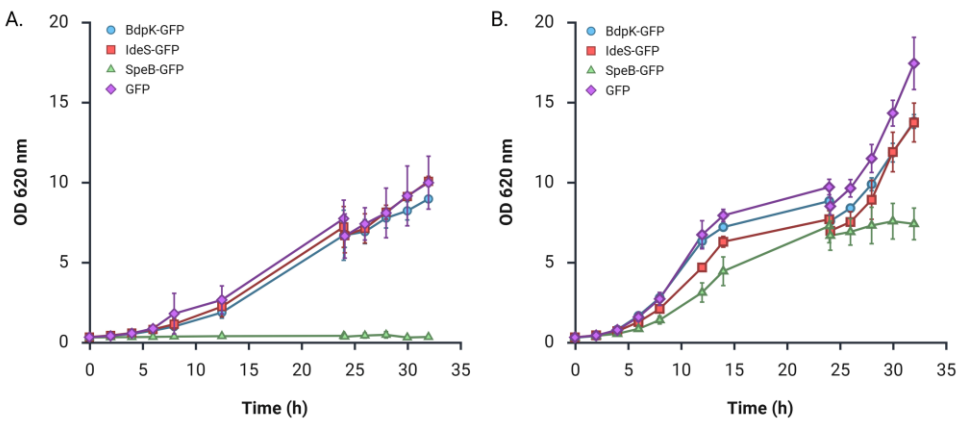


Figure 4.3: Aerobic batch growth of engineered yeast. **A**) Plasmid-based, and **B**) integration-based strains in YNB medium before and after induction of the *GAL1* promoter controlling protease gene expression. The dashed vertical line indicates the time of induction where galactose was added to a final concentration of 20 g/L. Symbols represent the strains producing BdpK-GFP (○), IdeS-GFP (□), SpeB-GFP (△), and GFP control (◊). Error bars correspond to standard deviation of biological triplicates. Adapted from Paper II.

The gene copy number is generally associated with protein synthesis levels. However, the relationship between gene copy number and protein yield is not always linear (Kim et al., 2015). The ori used on the plasmid can result in very different copy numbers inside the host cells (del Solar et al., 1998). Shuttle vectors that function across multiple hosts can contain several oris (Gnügge & Rudolf, 2017). Multi-copy plasmids are commonly used for recombinant protein production (İncir & Kaplan, 2024). In **Paper III**, pET vectors were used for recombinant polymerase production in *E. coli*, where pET vectors usually result in 15 to 20 plasmid copies per cell (Wall, 2009). In **Paper II**, multi-copy pESC plasmids with the auxotrophic leucine marker and the 2μ Ori that usually result between 10 and 40 copies per cell (Da Silva & Srikrishnan, 2012) were transformed into *S. cerevisiae* for recombinant protease production. Combinations of multi-copy plasmids and strong, inducible promoters generally result in high protein yields. In the context of protease production in yeast, previous studies have reported yields of up to 40 mg/L of yeast carboxypeptidase (CYP) when the *PRC1* gene is expressed under the strongly inducible *GAL1*p promoter on a multicopy plasmid (Nielsen et al., 1990; Rao et al., 1998).

To avoid selection markers, genes can be integrated into the genome of the production host. Integrative plasmids can be used for stable gene expression and protein production (Jensen et al., 2014). However, development of advanced tools, such as CRISPR/Cas9 have enabled precise and marker-free integration expanding the possibilities for precise genetic manipulation and co-production of several proteins (Jessop-Fabre et al., 2016). CRISPR/Cas9 mediated genome integration was performed in **Paper I** and **II**, where protease genes fused with GFP were integrated into the genome of *S. cerevisiae* for recombinant protein production. There, a larger proportion of the cell population produced GFP above autofluorescence levels compared to the production from plasmids, resulting in a homogeneous population. When one copy of the protease gene was integrated into the genome, the strains producing BdpK, IdeS and only GFP had similar growth profiles and a higher growth rate compared to the plasmid-based strains (Figure 4.3 B). The strain with the integrated *speB* gene grew before induction, although at a slower growth rate, and reached an OD_{620} value of around 7 after 24 hours The other strains reached OD_{620} values between 7.7 and 9.7 before the induction was started. After induction, the strains producing BdpK, IdeS and GFP continued to grow, and in contrast, the strain producing SpeB exhibited complete growth arrest after induction. Again, an extreme fitness loss was seen due to the production of SpeB (**Paper II**). No fluorescence from the SpeB-GFP fusion above autofluorescence levels was detected by flow cytometry, regardless of the expression system, likely due to the extreme burden of a small amount of active SpeB. When crude cell extract was incubated with pure SpeB, a large part of the proteome was degraded already after two hours of incubation (see Supplementary Figure S8-S10 in **Paper II**). The *in vitro* assay with cell extract and pure enzymes added showed that BdpK also degraded large parts of the proteome. This may

indicate that the protease BdpK produced in *S. cerevisiae* was inactive, since no fitness loss was observed although GFP was detected.

Reporter proteins and flow cytometry

In **Paper I** and **II**, I used GFP as a reporter protein to monitor protein production at a single-cell level using flow cytometry (FCM), and for detection through Western blot. A reporter protein is a protein that can be easily detected and indicate protein localization and gene expression. GFP is a common example of a reporter protein (Carter & Shieh, 2015). Fusion proteins with GFP and FCM analysis allow for real-time monitoring of protein production without the need for extraction. FCM analyses individual events (cells) in a liquid stream, passed through a laser beam, measuring light scatter and fluorescence (Figure 4.4). Light scattering is measured in two ways where forward scatter (FSC) provides an estimate of cell size and side scatter (SSC) indicates the granularity of, especially larger (mammalian), cells (Shapiro, 2005). Lasers of different wavelengths, such as the standard 488 nm blue laser, excite fluorophores whose emitted fluorescence is detected through wavelength-specific filters. The recorded events can be used both for phenotypic characterization and cell quantification (Rao et al., 2023). FCM can also be used with fluorescent dyes to obtain additional information. Propidium iodide (PI), for example, penetrates cells with damaged membranes and subsequently binds to DNA, allowing assessment of cell viability (Davey & Hexley, 2011). The resulting data, fluorescence intensities and scatter parameters for each event, can be analyzed using software such as FlowJo or open-source tools like Python, to summarize mean fluorescence intensities (MFIs), identify subpopulations, and in the case of **Paper I** and **II**, follow GFP levels over time after induction (Figure 4.5 and Supplementary Figure S1 in **Paper II**).

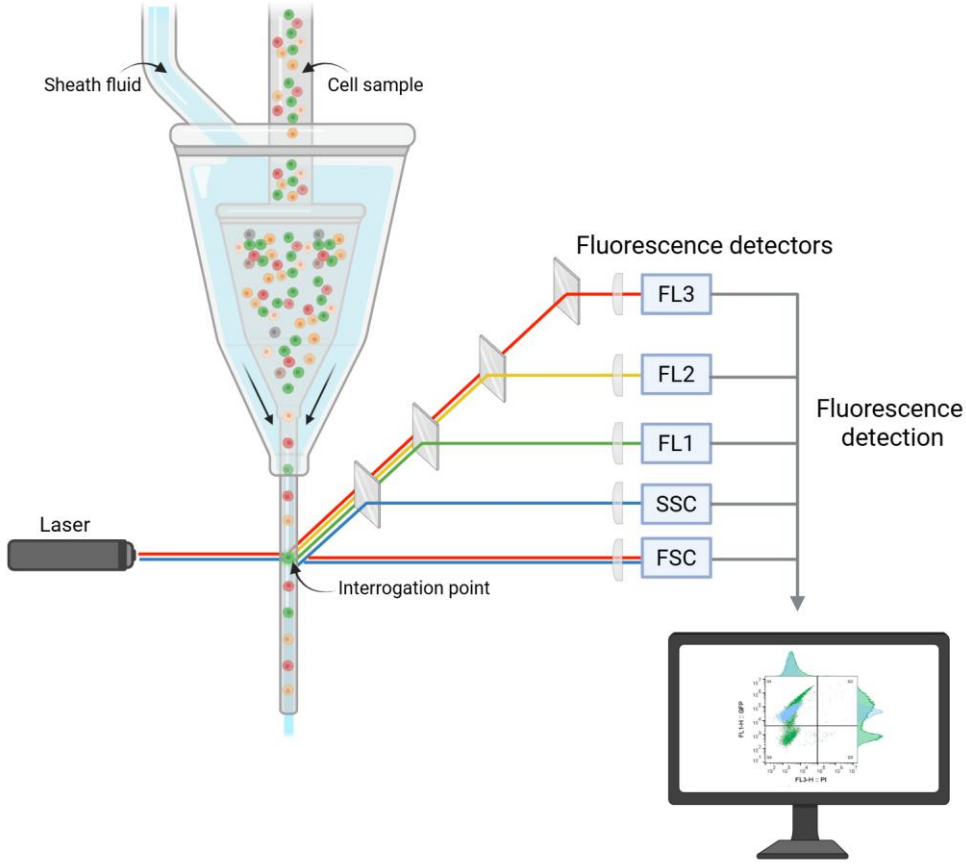


Figure 4.4: Schematic representation of a flow cytometry (FCM) instrument. The sheath fluid is circulated on the sides to push the cells in a single line as indicated with arrows. This way the cells reach the interrogation point one at a time where they are exposed to the blue and red excitation lasers and their light refraction and fluorescent emission is collected in the different channels. FSC: Forward scatter, SSC: Side scatter, FL1-FL3: representation of fluorescence collection channels.

As mentioned earlier, there was a difference in subpopulations depending on if the gene expression was from a multi-copy plasmid or from a single-copy genome integration. The two proteases BdpK and IdeS as well as the GFP control resulted in high fluorescence levels when produced in the plasmid-based strains. However, only about half of the cells produced GFP that reached fluorescence levels above the autofluorescence levels (Figure 4.5) (**Paper II**). The GFP positive population of the plasmid-based strains resulted in higher fluorescence levels compared to the integration-based strains. When looking at the whole cell population, the integration-based strains resulted in higher mean fluorescence intensities (MFI) compared to the plasmid-based strains. For protein production purposes, it is

important that all cells in the population produce the protein of interest in a desired amount.

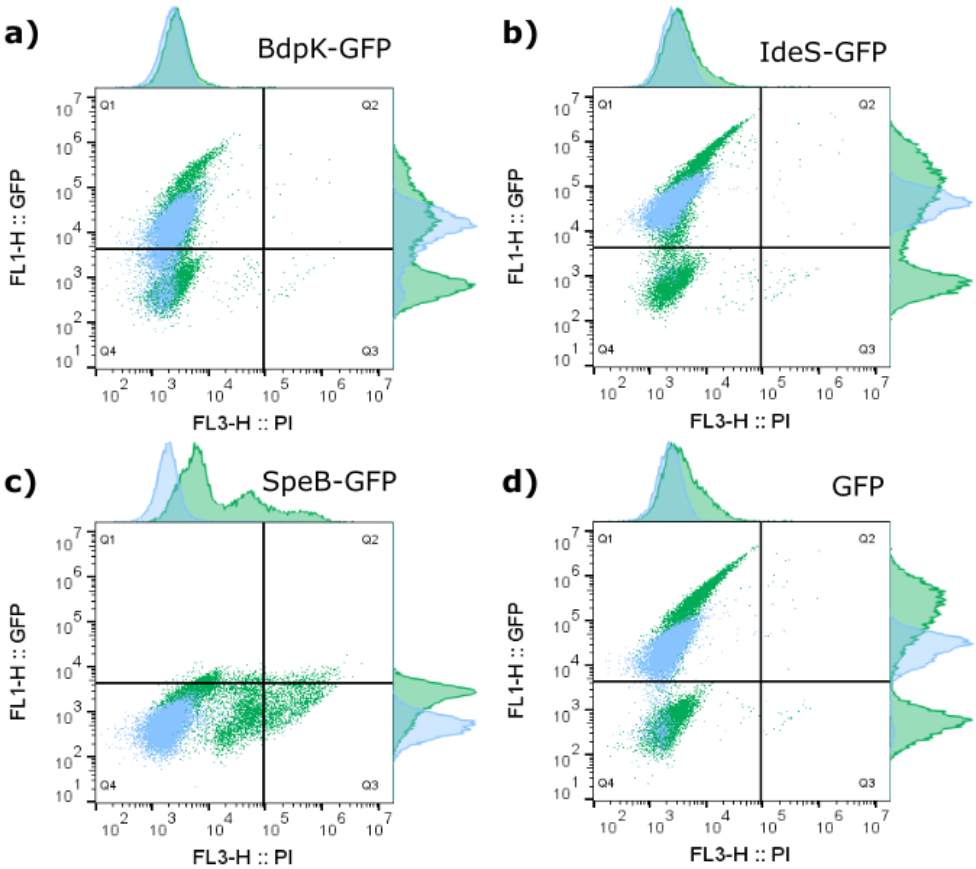


Figure 4.5: Flow cytometry analysis of yeast strains carrying genes for different bacterial proteases fused with GFP. Each dot represents a cell, and the distribution is plotted along with the histograms, showing the fluorescent intensity in the channels FL1-H (GFP) and FL3-H (PI). Q1: intact GFP + cells, where plus (+) means GFP positive, i.e., fluorescence above autofluorescence. Q2: permeable GFP + cells. Q3: permeable GFP - cells, where minus (-) means GFP negative cells, i.e., fluorescence at or below autofluorescence. Q4: intact GFP- cells. The green color corresponds to the multi-copy plasmid-based strains, and the blue color corresponds to the single-copy integrated strains. Samples taken 6 h after induction are shown. The samples shown are representative of three biological replicates. **A**) BdpK-GFP **B**) IdeS-GFP. **C**) SpeB-GFP. **D**) GFP control strains. Adapted from Paper II.

Signal peptides for protein localization

Signal peptides can be used to direct recombinant proteins to specific cellular compartments in eukaryotic cells or to facilitate secretion (Owji et al., 2018). Organelle-specific localization may be necessary for proper PTMs, such as glycosylation in the Golgi apparatus or disulfide bond formation in the endoplasmic reticulum (Kapp et al., 2009). Secretion of recombinant proteins is a common approach that can simplify downstream processing and purification, as it reduces the complexity of cell lysis and minimizes contamination from intracellular components (Owji et al., 2018). Additionally, compartmentalization can reduce cytotoxicity associated with certain proteins, for example by targeting enzymes to the peroxisome or vacuole (Sibirny, 2016; Weis et al., 2013). Expression of a broad-spectrum protease in yeast was found to have a significant negative impact on cell growth and fitness (**Paper II**). It has previously been found that organelle compartmentalization could circumvent toxicity of recombinant proteins (Bar-Peled & Kory, 2022; Yin et al., 2024).

Therefore, I explored peroxisomal localization of the protease-GFP fusions as a means to protect the cells from the proteolytic activity. A previous study showed an improvement in growth and product titer by compartmentalization of a toxic enzyme, norcoclaurine synthase, into the peroxisome of *S. cerevisiae* (Grewal et al., 2021). Heterologous proteins can be targeted to the peroxisome by the addition of a small signal peptide (peroxisomal targeting signal type 1, PTS1) to the C terminus of the protein of interest (Subramani, 1992). To obtain peroxisomal localization, I added the enhanced PTS1 (ePTS1) to the C terminus to generate BdpK-GFP-ePTS1 and IdeS-GFP-ePTS1. The N-terminal region of the peroxisomal membrane protein Pex22 fused to a Red Fluorescent Protein (RFP, mRuby2) was co-expressed as a fluorescent peroxisomal marker (Halbach et al., 2009). Localization of the protease fusions was verified through fluorescence microscopy (Figure 4.6). For IdeS, the fused protein was localized to the peroxisome as the RFP and GFP signals overlapped in the merged image. For BdpK, some of the protein was localized to the peroxisome. However, GFP signals not overlapping with the RFP signal were seen and may have emerged from the formation of inclusion bodies. This could explain the high GFP signals observed by flow cytometry, and the lack of fitness reduction and protease activity seen in **Paper II**. These results demonstrate that peroxisomal localization of proteases is feasible. However, further experiments are needed to confirm any protective effect, as well as to evaluate compartmentalization of SpeB, which imposed the greatest burden on the yeast cells.

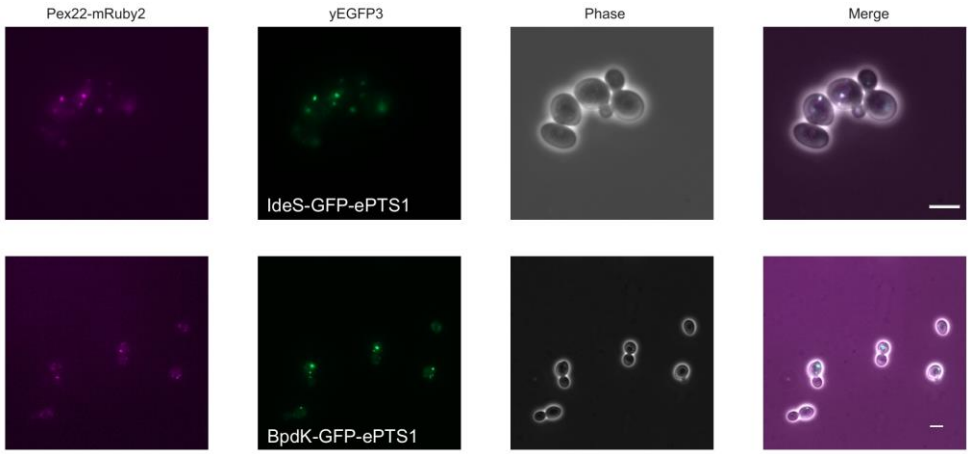


Figure 4.6: Fluorescence microscopy and phase contrast imaging for verification of peroxisomal localization. Results for IdeS are shown at the top row and for BdpK at the bottom row. Purple clusters in the first column indicate peroxisomal localization of the peroxisomal protein Pex22 fused to the red fluorescent protein (RFP) mRuby2, used as a control. Green clusters in the second column indicate presence of GFP-protease-ePTS1 fusions. The overlap of green and purple clusters in the fourth column indicates peroxisomal localization of the GFP-protease-ePTS1 fusion proteins.

Fusion tags for downstream purification

Fusion tags are important for effective downstream processing and purification of the produced protein (Sørensen & Mortensen, 2005; Yuan et al., 2025). Fusion tags such as poly-histidine (His) tags and Glutathione-S-Transferase (GST) tags are examples of affinity tags, used in protein purification (Arakawa & Akuta, 2025). His-tags usually consist of six to fourteen consecutive histidine residues at either terminus of the protein. Due to their small size, they can often remain fused with the target protein without interfering with its activity, intended application and biological function (Block et al., 2009; Charlton & Zachariou, 2008). Their terminal placement enhances accessibility for binding to immobilized metal ions during purification (Bornhorst & Falke, 2000). In contrast, GST is a large protein, and the fusion is frequently removed post-purification using site-specific proteases such as TEV protease (Gräslund et al., 2008; Yuan et al., 2025).

Before protein purification using affinity tags, the protein needs to be accessible. Both sonication and detergent-based lysis were evaluated for cell lysis to obtain the recombinantly produced DNA polymerases (**Paper III**). Sonication was, however, important both to lyse the cells and to shear the *E. coli* DNA. Fragmentation of the genomic DNA was necessary to obtain a manageable liquid cell lysate for downstream protein purification, as the intact genomic DNA formed a highly viscous, gel-like structure that was not possible to pellet by centrifugation and would interfere with purification.

In **Paper III**, His-tagged recombinant DNA polymerases were purified using Immobilized Metal ion Affinity Chromatography (IMAC) with gravity flow (Figure 4.7). This method relies on nickel ions (Ni^{2+}) bound to a chelating resin, such as nitrilotriacetic acid (NTA), which interacts with the His-tag via coordination bonds (Bornhorst & Falke, 2000; Porath, 1992). Upon loading the lysate, His-tagged proteins bind to the resin while unbound proteins are washed away. In **Paper III**, additional washing steps were added to reduce DNA contamination in the polymerase fraction. Elution was performed using imidazole, which competes with histidine residues for nickel binding, thereby releasing the target protein (Bornhorst & Falke, 2000).

The relatively simple purification method using IMAC with gravity flow was chosen in **Paper III** to develop a resource-efficient protocol for DNA polymerase production to be applied under constrained conditions, using only readily available laboratory reagents. Residual DNA in the final polymerase preparation can interfere with the intended applications and limit its usability. Assays involving bacteria and especially *E. coli* detection are limited by the presence of host cell DNA in the polymerase preparation. Unspecific product formation between the background DNA and primers or if shorter fragments of the residual DNA act as primers, complementary to longer parts of the background DNA, product formation can occur. If formation of unspecific products occurs and the length in base pairs (bp) or GC-content matches the intended product, the assay could result in false positives or masking the true products if the melt curves are similar. The additional washing steps added while the His-tagged DNA polymerase was bound to the column decreased the DNA concentration in the final polymerase preparation (**Paper III**). Another strategy to minimize DNA content in the final polymerase preparations is the addition of DNase, however, it needs to be efficiently removed as even traces of DNase in the final preparation may interfere with the target DNA in PCR. Efficient removal of DNases typically requires more advanced chromatography-based purification systems that separate the DNase from the DNA polymerase based on their molecular size, such as Fast Procedure Liquid Chromatography System (FPLC) coupled with size exclusion chromatography (SEC). Heat sensitive DNases are available but they are not compatible with large-scale production and reaction setups used for polymerase production due to their the high cost. As the protocol in **Paper III** was developed to require limited materials and readily available reagents, with the goal of supporting preparedness in the event of supply shortages, DNase and more advanced chromatography-based protein purification setups were excluded.

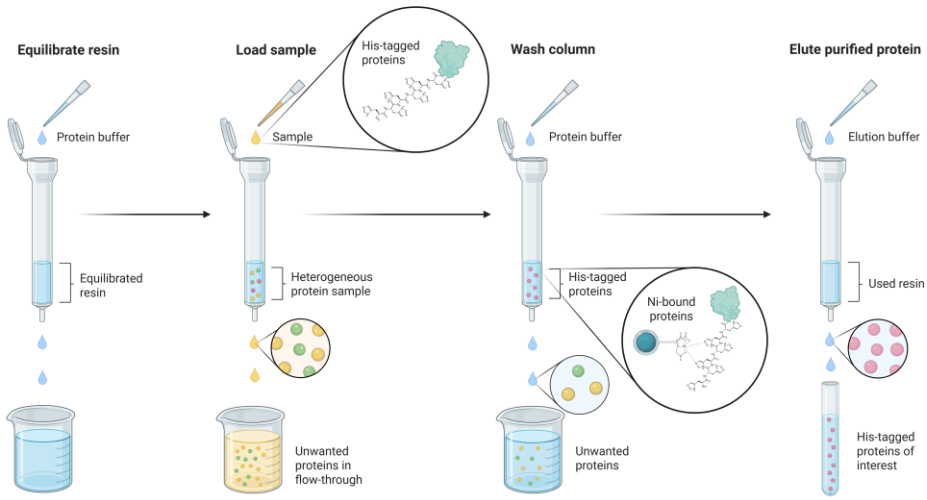


Figure 4.7: Schematic overview of the workflow of Immobilized Metal ion Affinity Chromatography (IMAC) with gravity flow for protein purification.

Ultimately, the choice of production host, expression system and vector or integration depends on the protein to be produced, required modifications, downstream applications, and production scale.

Chapter 5
DNA polymerase characteristics and PCR errors

DNA polymerase characteristics can impact the amplicon yield and polymerization errors formed in PCR assays. Errors introduced during amplification, such as base substitutions, insertions, deletions, or stutter artefacts can compromise the reliability of downstream analyses, leading to false-positive results, misinterpretation of genetic variations or incorrect allele calls in STR profiling (Cline et al., 1996; Hauge & Litt, 1993; Kebschull & Zador, 2015; Pienaar et al., 2006; Walsh et al., 1996). As described in Chapter 3, STR profiling with CE is not without limitations. Stutter artefacts may make it difficult to correctly call the alleles present in the sample, especially in crime scene stains containing DNA from multiple donors (Meldgaard & Morling, 1997). Base substitutions are generally not problematic in CE-based STR analysis, as the method relies on fragment length rather than sequence. However, in sequencing assays, base substitution errors become more apparent and can affect data interpretation. MPS-based STR analysis visualizes artefacts containing base substitution errors. The increased resolution compared to CE also leads to better allele discrimination, as sequence variants in both the repeat region and in the flanking regions are distinguishable (Agudo et al., 2022). Still, base substitutions and stutter artefacts formed in the PCR-based library preparation prior to sequencing may have an impact on the results. In **Paper IV**, six DNA polymerases with different characteristics were evaluated in terms of amplicon yield and polymerization errors in sequencing of STRs using SiMSen-Seq and UMIs. The use of STR markers enabled simultaneous investigation of random single-base substitutions and systematic stutter artefacts.

Amplicon yield

In a forensic or clinical context, an optimal limit of detection is crucial. It is therefore important to apply DNA polymerases that give both a high product yield and a minimum of errors. The amplicon yield is generally higher when using a *Taq* polymerase compared to a *Pfu* polymerase, which is one of the reasons why these

two enzymes often are combined in PCR (Ishino & Ishino, 2014). Blending the polymerases takes advantage of the higher yield of *Taq* polymerase and the superior fidelity of *Pfu* polymerase (Arezi et al., 2003). There is sometimes a trade-off between yield and specificity when optimizing PCR assays, as measures to increase specificity, e.g. raising the annealing temperature for more stringent primer binding, may reduce the overall yield. Amplicon yield is often high when using commercial DNA polymerases under ideal conditions, such as high template DNA concentrations, high quality samples, and optimized mastermix compositions. These favorable conditions, however, may not accurately represent the complexity of forensic or clinical samples. The presence of PCR inhibitors from the samples or the sample processing (Sidstedt et al., 2020; Sidstedt et al., 2019) as well as other challenging amplification conditions, such as the barcoding PCR in SiMSen-Seq, can significantly decrease the amplification efficiency depending on the applied DNA polymerase (**Paper IV**). Under these circumstances, DNA polymerase performance can vary greatly, making it important to select a polymerase that is tolerant to a wide range of conditions.

Two of the six evaluated DNA polymerases in barcoding PCR, SuperFi II (fidelity >300X vs. *Taq*) and Phusion HS II (52X vs. *Taq*), resulted in high STR amplicon yields when applied in the barcoding PCR (see Figure 2 A and B in **Paper IV**). These are enzymes with similar characteristics, such as a 3' to 5' exonuclease domain that provides high fidelity and a DNA-binding domain that increases the processivity. The low concentration of DNA polymerase and primers used in the barcoding PCR, in combination with large genomic DNA molecules serving as template and the few cycles applied sets high demands on the DNA polymerase performance. In the experimental setup, SuperFi II and Phusion HS II performed well in the barcoding PCR, resulting in large amounts of short product STR amplicons labelled with UMIs.

The other DNA polymerases AccuPrime *Pfx* (26X vs. *Taq*), AccuPrime *Taq* HF and Ex*Taq* HS (9X and 4.5X vs. *Taq* respectively), and Immolase, a novel family A DNA polymerase variant from an undisclosed organism, resulted in low STR amplicon yields when applied in the barcoding PCR. This indicates the importance of the DNA-binding domain or high fidelity on the product formation in the initial PCR cycles. However, when the same polymerases were applied in the adaptor PCR, less variation between the polymerases was observed, both in terms of raw read counts and STR amplicon yield (see Figure 2 E and F in **Paper IV**). The short product STR amplicons from the barcoding PCR act as template molecules in the adaptor PCR and this, together with the less challenging PCR conditions with higher concentrations of polymerase and primers and 27 PCR cycles, seemed to even out some of the differences between the polymerases. Thus, high fidelity and processivity did not have a substantial positive effect on the yield of STR amplicons in adaptor PCR.

Single-base substitutions

Single-base substitution errors arise from incorrect nucleotide incorporation that is not corrected by proofreading and subsequently propagates through successive PCR cycles. These errors are random and are generally more common for polymerases without a 3' to 5' exonuclease domain (Eckert & Kunkel, 1991). The polymerase fidelity varies significantly between enzymes. For example, *Taq* DNA polymerase, which lacks 3' to 5' exonuclease activity, introduces incorrect nucleotides more frequently than *Pyrococcus*-based DNA polymerases. SuperFi II and Phusion HS II are *Pfu*-based polymerases, that exhibit proofreading capabilities and have DNA-binding domains, that reduce base substitution errors (Cline et al., 1996; Coulther et al., 2019).

In the barcoding PCR, the proportion of base substitutions (referred to as "other errors" in Figure 5.1) was significantly lower after consensus read generation (reflecting the sequences generated in the barcode PCR) when using the high fidelity polymerases SuperFi II and Phusion HS II, compared to AccuPrime *Taq* HF and Immolase (**Paper IV**). Both AccuPrime *Taq* HF and Immolase have lower fidelity compared to SuperFi II and Phusion HS II and resulted in proportions of other errors, mainly base substitutions, around 2% after consensus read generation (Figure 5.1 B). This was between 3.1 and 7.6 times higher than SuperFi II and Phusion HS II. The proofreading of AccuPrime *Taq* HF is performed by a separate proofreading enzyme. This provided less efficient correction than 3' to 5' exonuclease domains within the polymerases themselves. Immolase lacks proofreading activity altogether, why more base substitution errors were expected for this enzyme.

When different enzymes were applied in the adaptor PCR, AccuPrime *Pfx* resulted in a higher proportion of overall errors, including stutters, compared to the other polymerases. The percentage of base substitutions was also higher, although the fidelity of AccuPrime *Pfx* is higher than three of the enzymes displaying lower proportions of base substitutions (Figure 5.1 C). The more even results between the different polymerases applied in the adaptor PCR, after consensus read generation, both in terms of single-base substitutions and stutter artefacts, shows the importance of correct and efficient amplification in the initial PCR cycles (Figure 5.1 D). The generation of consensus reads reduced the proportion of incorrect sequences substantially, highlighting how UMIs can be used to lower the incidence of random base-substitutions as well as systematic stutter artefacts bioinformatically.

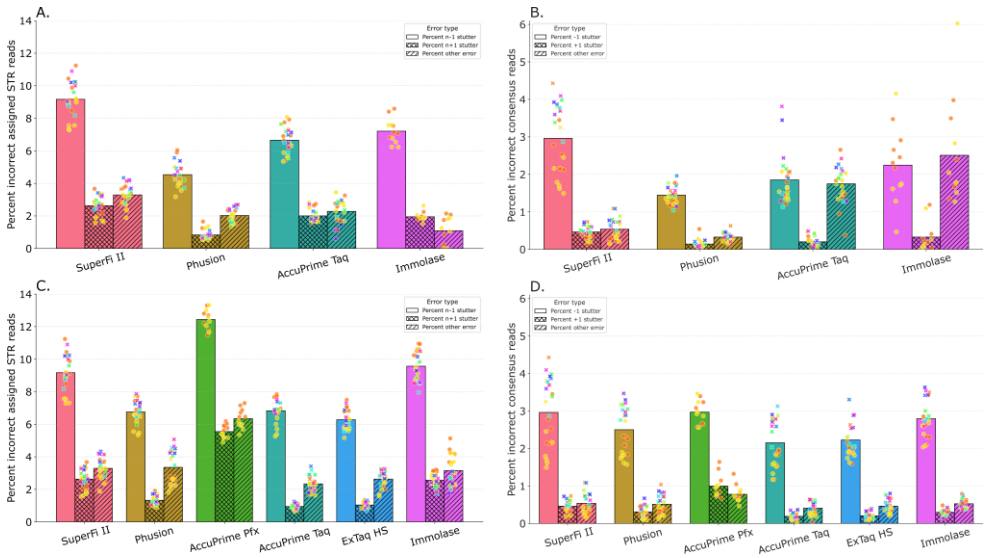


Figure 5.1: Error proportions in barcode PCR (**A** and **B**) and adaptor PCR (**C** and **D**) for STR reads (A and C), and consensus reads (B and D). Errors are categorized into n-1 stutters (left, no pattern), n+1 stutters (middle, crossed pattern), and other errors such as base substitutions and rare stutter variants (right, striped pattern). The circles and crosses represent individual samples, where the circles are sequenced at Laboratory A and the crosses sequenced at Laboratory B. The different colors represent different DNA samples as input DNA. SuperFi II was used in the barcoding PCR when studying the effects of different DNA polymerases on adaptor PCR, and vice versa. In A and B n=24 for SuperFi II, Phusion HS II and AccuPrime Taq HF and n=12 for Immolase. In C and D n=24 for SuperFi II, Phusion HS II, AccuPrime Taq HF, ExTaq HS and Immolase and n=12 for AccuPrime Pfx. From Paper IV.

Stutter artefacts

The systematic nature of stutter formation makes it possible to remove them by applying thresholds at a certain percentage of the parental allele, i.e., number of reads in sequencing or fluorescence peak heights in CE. Such thresholds filter out the most common stutters, i.e., n-1 and n+1 variants. This works well in forensic analysis of single-source samples but comes with the risk of removing any true alleles that are present at low ratios such as in crime scene stains with DNA from multiple donors or cancer variants in clinical diagnostics. Efforts have been made to reduce the amounts of stutters, either by adapting the DNA polymerase or the thermal cycling program (Courtney et al., 2024; Loh et al., 2024), by removing them bioinformatically post analysis (Sidstedt et al., 2024; Woerner et al., 2021), or by removing the PCR step (Fungtammasan et al., 2015).

Different DNA polymerases yield varying stutter ratios, although the relationship between stutter formation and factors such as fidelity has remained unclear (Yamanoi et al., 2021). In a previous study, SuperFi, a high-fidelity, *Pyrococcus*-

like family B DNA polymerase with a DNA-binding domain, outperformed two other high-fidelity polymerases, while a *Taq*-based polymerase with much lower fidelity produced nearly identical stutter levels as SuperFi (Yamanoi et al., 2021). Research has shown that high concentrations of *E. coli* SSB can reduce slippage during amplification (Viguera et al., 2001). Additionally, biologically compatible solutes like betaine or sorbitol have been found to reduce stutter artefacts in mono- to penta-nucleotide repeats when used as PCR additives. These effects were observed across polymerases from various sources, including *Taq* and *Pfu* (Coticone & Bloch, 2014). The potential connection between stutter formation and the processivity of thermostable DNA polymerases used for amplification of STR markers is not fully elucidated. A *Taq* polymerase fused with the bacteriophage T7 thioredoxin binding domain and the thioredoxin protein resulted in a conformational change that substantially reduced stutter formation (Courtney et al., 2024). This suggests that enhanced processivity may help minimize stutters. Maintaining thioredoxin in its reduced state using DTT was needed to minimize stutters. The processivity of the bacteriophage T7 polymerase is an effect of a conformational change following binding of thioredoxin to the thioredoxin binding domain, as it, unlike other polymerases lacks a protein clamp to enable processive polymerization (Davidson et al., 2003; Lovett, 2007). Apart from processivity, strand-displacement activity has also been implicated in reducing stutter levels, as demonstrated by *Tli* DNA polymerase, which exhibits inducible strand-displacement (Viguera et al., 2001).

In **Paper IV**, a total of 12 different single-source DNA samples with published or known STR profiles were used for sequencing of STR markers. Since the profiles were known, this allowed for differentiation between true allele variants and PCR-introduced artefacts. Systematic stutter artefacts were observed across all DNA polymerases, including incorrect consensus reads and erroneous sequences within UMI families that resulted in correct consensus reads (Figure 5.1). Phusion HS II resulted in the lowest proportion of stutters. The increased processivity through a DNA-binding domain as well as high fidelity through the 3’ to 5’ exonuclease domain seemed to influence the amplicon yield. Polymerases lacking 3’ to 5’ exonuclease activity or DNA-binding domains gave similar or lower stutter levels compared to SuperFi II. Thus, no clear relationship between these characteristics and stutter formation were seen.

Performance of in-house produced DNA polymerases in SiMSen-Seq library preparation

In **Paper III**, the in-house produced DNA polymerases were applied in library preparation for subsequent sequencing of STR markers using the SiMSen-Seq STR method (Sidstedt et al., 2024). SuperFi II was used in the barcoding PCR and in-house produced polymerases in the adaptor PCR. The overall performance of the polymerase variants was consistent and comparable to each other and the mean proportion of STR reads were above 40%. Before consensus read generation, the mean proportions of n-1 stutter ranged from 6% to 7.5%. Other errors, including n+1 stutters, base substitutions and more complex stutter variants ranged from 5.5% to 6.5%. These levels are comparable to or even a bit lower than the error rates for the commercial polymerases (Figure 5.1 C). Super Fi II was used in the barcoding PCR when evaluating both the commercial polymerases and the in-house enzymes in adaptor PCR. After consensus read generation, the proportion of n-1 stutters was around 2% and other errors around 0.5%, reflecting the performance of SuperFi II in the early PCR cycles and again similar to the results of the commercial enzymes (Figure 5.2 D). A higher amount of genomic template DNA (20 ng) was used for the in-house synthesized DNA polymerases compared to the commercial ones (1 ng), leading to more amplicons from the barcoding PCR acting as template molecules in the adaptor PCR. However, two samples where 1 ng input DNA was used were sequenced successfully, one for *Pfu* polymerase and one for *Pfu*-Sso7d polymerase, showing that the in-house polymerases are compatible with lower amounts of template DNA in the barcoding PCR. Further optimization would be required to obtain similar performance as the commercial DNA polymerases in the terms of amplicon yield. Some long unwanted products were seen when the produced polymerases were used, which was not the case for SuperFi II (see Figure 5 in **Paper III**). This could be due to the residual *E. coli* DNA in the polymerase preparations, possibly generating PCR products through unspecific binding of STR primers or by short *E. coli* DNA fragments acting as primers themselves. Still, more than 40% of the reads were recognized as STRs, which is comparable to several commercial DNA polymerases (**Paper IV**). Thus, the *E. coli* DNA seemed to have a limited negative effect on the performance of the assay. However, the residual DNA may cause larger problems if the in-house polymerases are to be used in the barcoding PCR.

Chapter 6
Conclusions

In this thesis, I have focused on recombinant production of proteases and DNA polymerases as well as their biotechnological applications. My work spanned from introduction of heterologous genes in *S. cerevisiae* and *E. coli* and development of protocols for recombinant protein production with limited resources to assessing how the host organism is affected by the recombinant enzymes as well as evaluation of DNA polymerase performance in applications ranging from conventional PCR to library preparation for sequencing of STR markers. The impact of DNA polymerases with varying properties on amplicon yield and PCR error formation, both random base-substitutions and systematic stutter artefacts, was investigated in the context of STR analysis. The main conclusions are summarized below.

- The yeast *S. cerevisiae* can be used to produce microbial proteases with specific substrate scopes, such as IdeS. The broad-spectrum protease BdpK was produced in the yeast and resulted in high fluorescence levels when fused to GFP, although the probable formation of inclusion bodies suggests challenges in proper protein folding or solubility.
- Production of the protease SpeB was found to pose an extreme burden on *S. cerevisiae* cells, affecting both growth and viability negatively. The negative effect was likely due to the broad substrate scope of SpeB.
- GFP fusions combined with PI staining and FCM analysis provide a powerful method to monitor both protein production and physiological effects on the host cell without the need for cell lysis and protein purification. This can be used for optimization of the protein production process, to follow the induction over time and to track subpopulations that do not produce the desired protein.
- I developed a resource-efficient protocol that enables the production of *Taq*, *Taq*-Sso7d, *Pfu* and *Pfu*-Sso7d DNA polymerases in-house using readily available laboratory materials and reagents. This protocol can be employed in case of supply chain disruptions with limited access to specialized equipment. Large quantities of active polymerase were produced. Enough *Taq* polymerase to perform 90 000 PCR reactions were obtained from a 200 mL cell culture. These polymerases are suitable for a range of PCR-

based applications, from conventional amplification to qPCR and library preparation for sequencing. The performance was comparable across the different in-house produced polymerases, supporting a robust production method.

- The production pipeline was improved to minimize residual DNA in the polymerase preparations. Residual *E. coli* DNA in polymerase preparations can interfere with downstream assays, leading to nonspecific amplification products. Preparations of *Taq*-Sso7d and *Pfu*-Sso7d polymerases contained more residual DNA, likely due to the DNA-binding domain retaining DNA during the protein purification process. Addressing this is crucial for improving assay specificity and reliability.
- Polymerase characteristics affect the amplicon yield and formation of PCR errors, both random single-base substitutions and systematic stutter artefacts. Through the usage of UMIs and SiMSen-Seq, it was possible to study the first few PCR cycles where genomic DNA is the main template as well as the later cycles where short STR amplicons are the primary template. Polymerases with high fidelity through a 3' to 5' exonuclease domain and increased processivity through a DNA-binding domain produced high STR amplicon yields and low levels of base substitutions when applied in the initial PCR cycles. These characteristics are important in the first few cycles to maximize yield and minimize errors.
- Systematic stutter artefacts were seen among all polymerases applied, regardless of their characteristics. Phusion HS II resulted in the lowest proportion of stutter artefacts, while SuperFi II resulted in the highest, despite their similar characteristics. *Taq*-based polymerases generated lower proportions of stutter artefacts compared to SuperFi II and no clear correlation between proportion of stutters and polymerase characteristics such as fidelity and processivity was seen.

Overall, this thesis demonstrates the versatility and challenges of using *S. cerevisiae* as a host organism for microbial protease production, highlighting both successful production as well as cellular stress response. The observed burden from proteolytic activity, particularly with SpeB, shows the need for targeted localization or secretion strategies to mitigate intracellular damage. The development of a simplified protocol for in-house DNA polymerase production offers a cost-effective and robust alternative in case of supply shortage for various PCR applications. However, residual *E. coli* DNA in the polymerase preparations presents a hurdle to assay specificity. Finally, the study shows that polymerase characteristics are important for accuracy in amplification of repetitive sequences, such as STR markers.

These results provide a foundation for future efforts in optimizing host organism systems, enzyme production, and a deeper understanding of stutter formation in early and late PCR cycles for both research, diagnostic and forensic applications.

Chapter 7
Outlook

This thesis has addressed key aspects of recombinant production of microbial DNA polymerases and proteases for molecular applications, and this chapter highlights unresolved questions and challenges for further investigation. Future research directions and remaining challenges are summarized below.

Recombinant protease production in S. cerevisiae

When a plasmid-based system was used for production of proteases in *S. cerevisiae*, a large subpopulation of non-producing cells was observed. In contrast, this was not observed when an integration-based system was used, where only a small subpopulation of non-producing cells was present. The GFP positive sub-population of the plasmid-based cells reached higher fluorescence intensities, likely due to the increased gene copy number. However, the cells grew at a lower growth rate compared to the integration-based strains. The gene copy number is generally associated with protein concentration but the relationship is not always linear. A strategy to reach higher protein concentrations, and benefit from production of the whole cell population and higher growth rate could therefore be to integrate several copies of the protease gene into the genome of *S. cerevisiae*. This could be done with CRISPR/Cas9 where several gene copies can be integrated into the genome at the same time. Due to the highly efficient HDR in *S. cerevisiae*, too many copies with identical sequences can lead to genetic instability, and experiments to find an optimal number of gene copies for high protein production and maintaining genetic stability would be necessary.

Strategies to mitigate the toxic effect include targeting proteases to specific organelles. Compartmentalization of the produced proteases may be used to protect the cells from the proteolytic activity of the proteases with a broad substrate spectrum. Localization of the proteases to the peroxisome was successful, however more work is needed to confirm if the cells are protected from the proteolytic activity, as well as the effect of compartmentalization of SpeB that posed an extreme burden on the yeast cells. Other organelles, like the vacuole, where the native proteases of *S. cerevisiae* are located, or secretion of the produced proteases could also be investigated as a strategy to protect the host cells and produce high amounts of recombinant proteases.

When pure BdpK was incubated with crude cell extract, degradation of large parts of the proteome was observed, while no negative physiological effects were seen on the strains producing BdpK, suggesting that inactive protease was produced. In the fluorescence microscopy pictures, aggregations of GFP that could be inclusion bodies were seen, which would explain why no activity was measured and no negative effect on the host cells were seen. Further experiments are needed to confirm this also, if inclusion bodies are formed, that could possibly be used as a production strategy to produce large quantities of pure protein. The inclusion bodies may be purified from cell lysate and later refolded into solubilized functional proteins.

*GAL1*p is a native tightly regulated promoter. However, the severe impact on growth on the cells harboring the *speB* gene, even before induction, suggests that some basal transcription occurs prior to induction, although under the limit of detection with GFP and FCM. Due to this drastic effect, likely caused by trace levels of SpeB, this protease could potentially be used as a sensitive indicator of basal transcription from promoters before induction. This could then be used as a tool to identify and develop promoters with tight regulations or repressors to turn off transcription.

Another approach to avoid the protein formation before induction could be to change to another inducible synthetic promoter with an insulator that regulates the promoter even tighter and prevents leakiness. There are several available synthetic promoters, where unintended basal expression has been minimized by an insulator sequence upstream of the TATA-box (Tominaga et al., 2024). This could potentially be used in combination with compartmentalization or secretion for production of cytotoxic proteins like SpeB.

Alternatively, if the strategies presented above do not improve protease production, other host organisms like *K. phaffii* or *B. subtilis* could be investigated due to their efficient secretion systems.

DNA polymerase production, characterization and PCR noise levels

Sonication was used to lyse the cells and to fragment the *E. coli* DNA. The fragmentation was necessary to obtain a liquid for further protein purification as the DNA formed a gel-like, highly viscous clump too thick to handle in the centrifugation and purification steps. Different sonication protocols may fragment the DNA to varying degrees. There may be an optimal level of fragmentation that minimizes background *E. coli* DNA carried over with the polymerase during purification, resulting in less DNA in the final polymerase preparation. The additional washing steps added while the polymerase was bound to the IMAC column decreased the DNA concentrations in the final preparations and it is possible that a higher level of DNA degradation could lead to even more DNA being washed away. An altered sonication protocol would be a simple method to evaluate and

include if improvements are seen as both sonication and washing are existing steps in the developed polymerase production protocol.

DNase needs to be efficiently removed as degradation of DNA when added into a PCR reaction would interfere with the accuracy of the assay. DNases can efficiently be removed from polymerase preparations through advanced chromatography systems such as FPLF and SEC based on the differences in molecular size. In this work, DNase was not used to degrade the host DNA from the polymerase preparations, however a strategy to evaluate could be the addition of DNase before protein purification and determination if the DNase efficiently can be removed from the polymerase while bound to the IMAC column. Would the addition of DNase to the developed simplified production protocol with efficient removal be possible without traces of DNase ending up in the final DNA polymerase preparation?

With the higher sequence resolution of the STR markers obtained in **Paper IV**, the data could be further evaluated to investigate the single-base substitutions within the UMI families. Are there certain bases or positions in the amplicons that are more prone to incorrect nucleotide incorporation and does this differ between the applied DNA polymerases?

The in-house produced DNA polymerases were used in the adaptor PCR in the SiMSen-Seq assay. Future studies could use these polymerases in the barcode PCR of the SiMSen-Seq assay to investigate the differences between the *Taq* and *Pfu* based polymerases with or without the DNA-binding domain Sso7d when it comes to formation of systematic stutter artefacts and single-base substitutions in the first few PCR cycles. This would allow for more precise control of the polymerase characteristics and what roles they play in stutter formation in STR analysis.

A recently engineered polymerase variant from the marine archaea *Nanoarchaeum equitans* (*Neq*), which belongs to family B, includes two point mutations that increased the fidelity (*Neq*2X) (Hernández-Rollán et al., 2024). Fusion with Sso7d (*Neq*2X7) enhances the processivity but reduces the fidelity of the polymerase, indicating a trade-off between fidelity and processivity. Since the relationship between stutter formation and fidelity is unclear and processivity has been shown to decrease stutter formation, the *Neq* polymerase variants could be evaluated in STR analysis to further investigate the relationship between polymerase characteristics and stutter formation. The *Neq*2X7 DNA polymerase was shown to better tolerate several common PCR inhibitors compared to *Neq*2X and *Pfu*-Sso7d polymerases. For evaluation of polymerase performance in STR analysis, tolerance to inhibitors found in soil, sediment and blood, commonly found in crime scene samples, would need to be evaluated as well.

Another interesting recently engineered DNA polymerase is a *Taq*-based polymerase, fused with a bacteriophage T7 thioredoxin binding domain and the polypeptide thioredoxin (Courtney et al., 2024). This fusion resulted in a conformational change that led to substantially reduced stutter formation. Future

method development could therefore include the UMI-based 28-plex SiMSen-Seq prototype in combination with a polymerase that has been shown to produce low levels of stutter, such as Phusion HS II and the novel engineered variant. Both the UMI SiMSen-Seq method and the engineered *Taq*-based polymerase variant have been shown to substantially reduce stutters and the combination of the two would therefore be interesting to evaluate and investigate if stutter artefacts could be eliminated entirely with this approach.

Acknowledgements

To all my past and current colleagues and students at TMB, thank you for being part of this journey the last couple of years. I've enjoyed going to work and it's thanks to the friendly, warm and encouraging environment created by everyone at TMB. The daily lunches and fikas in the kitchen, the supportive and inspiring discussions, the sometimes-well-needed breaks, small chats in the office, the after-works, the workouts and the welcoming feeling are some of the things I will miss, thanks to all of you.

Like all research, this thesis would not have been possible without funding, and I am grateful for the financial support that made my research possible. Parts of this work related to proteases were financially supported by Sweden's Innovation Agency VINNOVA (reference number 2020–05038) and Genovis AB. Work related to DNA polymerase production and evaluation was financially supported by two separate 2:4 grants distributed by the Swedish Civil Contingencies Agency: *"Nationell laboratorieförmåga för totalförsvaret vid kris och höjd beredskap (FBD)",* and *"Robust och uthållig laboratoriediagnostik i kris och krig (FBD)".*

I would also like to thank Ernhold Lundströms stiftelse, Per Skårmans stiftelse, Civilingenjör Hakon Hanssons stiftelse, and the Royal Physiographic Society in Lund for the travel and research grants that enabled my participation in international conferences over the years. Lastly, I am grateful to Irisstipendiet for the generous grant that supported my research visit at the National Institute of Standards and Technology in the United States.

During my PhD studies, I've had the luxury of having five supervisors that all have been great mentors that helped and supported me throughout the last 4.5 years. Thank you for all the motivation, inspiration and guidance. In the order I got to know you:

Magnus, I've never met anyone with the same level of enthusiasm and endless ideas for research projects as you have. Thanks for all the encouraging project meetings resulting in new and exciting plans and scientific discussions. It has been great to have you as my main supervisor during the first two years of my PhD, and then as co-supervisor with great input on my new projects. Throughout the years, I've really appreciated the open door to your office and that you always make time for a small chat whenever I've needed it.

Rolf, I'm really glad that you gave me the opportunity to start my PhD journey at Genovis. You have always encouraged me to take on new challenges and were the co-supervisor I needed when I started. Thanks for always making me feel welcome and a part of the R&D group at Genovis. I enjoyed being a part of academia as well as industry and appreciated the combination of the two. Thanks for always responding promptly and helpfully to all my questions, you have been a great support throughout the projects.

Mattias, I'm happy that you agreed to be one of my co-supervisors when I asked, maybe without realizing that it would mean a bit more work for you. Thanks for all the valuable input on all of my projects and for making me a small part of Infection Medicine and B14, where I've enjoyed being in the lab and learning new techniques. I've appreciated your support and supervision since the start and the continuation during the second part of my PhD as well as the invitations to event hosted at Infection Medicine.

Johannes, I will always be grateful for the opportunity to continue my PhD with you. Thanks for the endless support and encouragement. Thanks for letting me grow and improve as a researcher. Thanks for introducing me to the world of forensics and DNA analysis. Thank you for always believing in me, inspiring me to try new things, and checking in on how I'm doing. I'm glad that you encouraged me when I wanted to go to the US and that you made my research stay at NIST such a positive experience. I've appreciated our weekly meetings, working on manuscripts together and the insightful discussion that always helped move the project forward. Thank you for everything you taught me.

Maja, I've always seen you as my supervisor during the second half of my PhD and I'm happy to have called you my co-supervisor the last year. I'm grateful for your positivity and the energy you bring to my projects. Thank you for always welcoming me to NFC in Linköping and for the introduction to the forensic community on conferences. It has been a pleasure to work with you, and I truly appreciated the time and effort you have put into my projects. Thank you for your reassurance and cheering me on when I went to NIST, it wouldn't have been the same without you.

I was first introduced to TMB when I started my engineering studies in 2016, where I met **Jenny** and **Daniel** that presented the research in applied microbiology at the division during my first week at LTH and my curiosity for microbiology started to grow. Thank you for encouraging me to continue with microbiology. And to all the seniors that have been at TMB during my time there, thank you for the warm, friendly, supporting and inspiring environment that you have created where students can thrive!

Jenny, I'm glad that you were my supervisor during my internship at BioInvent and encouraged me to take the opportunity to try something different for my Master thesis, that led me to TMB. Thank you for always standing up for us students, your kindness and open office have always been appreciated. **Daniel**, thank you for being

a great mentor and a great friend. Thank you for everything you taught me, for always giving small questions as much attention as the large ones and for inspiring me to become a better researcher. Your guidance has shaped not only my research but also how I approach challenges and collaboration. Thank you for being such an important part of this experience. **Marie**, thank you for teaching me how to be a bit more French and to express my needs more clearly. I've appreciated your input on my projects as a department representative as well as your much-missed presence at the daily fikas. **Linda**, thank you for tips and trick related to supervising students, I believe I became a better supervisor thanks to you. I also appreciate the Gerdahallen classes throughout the years. **Cathrine**, thank you for always caring about TMB and us PhD students. Thank you for taking over as the department representative and coming with input on my projects from a different angle. I've appreciated having you around and learning about drinking water as well. **Ghjuvan**, I admire how you always are checking in on how everyone is doing, and I hope to take after that habit. Thank you for starting the Bioinformatics meetings, I think they will help many to get started and learn from each other. **Gunnar**, your interesting and encouraging questions during seminars have been a source of inspiration and have often led to new ideas and perspectives, thank you. **Ed**, I'm glad that you always appreciated my gluten free baking and gave me a reason to try new recipes. Thank you for always having interesting questions at seminars and being around for fika. **Sebastian**, thank you for the interesting chats at fika and at Christmas lunches. I clearly remember when you started at TMB and talked about your motorcycle, hopefully I will have some more time to ride mine when I'm done with this thesis. See you on the road. **Peter**, although you already left TMB before I started as a PhD student, I've appreciated the chats we had and the advises you gave when you visited Lund, both before and after I started the PhD. **Anette**, thank you for the nice chats at fika and always helping with all the administrative stuff that I couldn't figure out how to do, from course registration to travel expenses and providing lists of students. **Christer**, it's incredible that one person can know so much about literally everything. You have solved countless problems, from microwaves to analytical scales that stopped working, from building shake flask holders from shelves to solving how to get WSL to work on my computer. Thank you for always taking the time to chat, solve problems and share interesting anecdotes about TMB in the past. **Emma**, thank you for always being so positive and asking how everyone is doing. I've appreciated your focus on finding solutions and your help with all the administrative tasks these past 1.5 years. **Carl**, as the head of division of BAM, thank you for showing an interest in the projects at +3 and caring about all the PhD students.

I'm deeply grateful for all the incredible people I've met during my PhD, this thesis would not have been possible without your support.

Nina and **Raquel**, thank you for taking me under your wings when I first started at TMB. Over the years, you two became very important colleagues and friends and my time at TMB would not have been the same without you. Thanks for bringing me to pottery and teaching me valuable skills outside of the lab as well. **Nina**, thanks for checking in on how I'm doing, especially now while writing. Thanks for grabbing a Boba Tea with me in Malmö and for introducing me to new cafés. **Raquel**, I deeply miss you in the office and in the lab. Thank you for the small chats, for the great advice and always putting a smile on my face. To both of you, thank you for always knowing what I need, even without me saying it, and for all the small gestures. Let's do another FaceTime call soon, I miss having you both nearby!

Nikhil and **Krishnan**, thanks for all the fun we had inside and outside of the lab and for becoming friends. It was great to have the two of you in the lab. **Nikhil**, your calmness is something I admire and try to remind myself to take after you. **Krishnan**, I'm impressed by how you can know so much about everything in the lab and I'm grateful for all the help with troubleshooting I got from you. **Tage**, thanks for helping me solve programming problems by saying different commands to try, I eventually got it to work. Who would have thought that we both would end up with a PhD in Applied Microbiology when we sang Bacon songs or when we were on the Guild board together? Let's celebrate with a bottle of flintbubbel like in the old days! **Viktor**, TMB is not quite the same without you and without Kanelbulle tasting, Easter Egg hunts, a giant FabALACTICA at my lab bench and the strong opinions about ketchup in the lunchroom, but I'm convinced that the impression you left, and your spirit will stay, and give so many more a great time at TMB. Thanks for being a great friend and contributing to many fun memories. **Bruna** and **Gisele**, you two came like a huge package of energy and enthusiasm. **Bruna**, it felt like the world was so small when we met in Vancouver and realized that you would come to TMB. I'm grateful for the time we had together at TMB, all the fun but also long days in the lab. Then it felt like forever until we would write our theses, but here we are, just a little bit left. **Gisele**, you might trick many since you are calm on the outside, but I'm happy I got to know the energetic side of you as well. Thank you for being a great office mate and for joining Bodypump after a year, I look forward to getting more updates from Brazil. **Isabela**, thanks for everything you taught me about protein purification, you are the type of mentor I aspire to become. Your friendship and hospitality are something I will always cherish. **Julia S**, thanks for all the fun we had, your energy and the bbq after-works with a distance. **Lisa**, thanks for showing me the qPCR and everything I needed to know in the Pathogen lab and for always having new aliquots ready for me to use. Thanks for all the fun times during teaching and for being a great friend and former colleague. **Chiara**, thanks for the time we shared in the student room and for all the fikas outside of work. **Sofia**, thanks for continuing my work on the pink yeast, thanks for all the fun we

had in the lab and thanks for becoming a friend. I will come and visit you in Italy one day and hope to get a private tour at Barilla or at least in Parma and Mantova. **Paulo**, thank you for all the joy and positive energy you brought to the division, along with interesting discussions during the meetings and seminars. **Matilda**, thank you for the weekly updates with pictures of your cats, and it has been great to follow your knitting projects, as soon as I'm done with this thesis, I will continue my crocheting projects and keep you updated. Thank you for all the fun times in the lab and for all your hard work with structuring protocols, it helps a lot. **Isabel**, it sometimes feels like you came to TMB as a master student at the same time as I did and you have been a key part of TMB since you started. Your empowering words after one of my seminars still makes me happy. You have an impressive ability to see everyone and I'm happy to call you a friend. **Iria**, I knew we would have fun together from the beginning. Thank you for being a great office mate, for taking care of the plants in the office, for all the fun times, and for all the encouragement and boosting when things have felt overwhelming. Thank you for making teaching and plate pouring fun to do together. Thank you for making me go to Gerda, even in the mornings. **Julia B** and **Sofie**, thank you for performing the many experiments with my polymerases, even though they didn't work out and thank you for making the forensic group flourish, I've been in great company in the pathogen lab during the last year. **Julia**, I'm impressed by how you can dive into new projects and techniques. Thank you for all the chats in the lab and at fika. **Sofie**, thank you for all the encouragement and for making me follow both beach volleyball and table tennis more closely. And thank you for going to Tabata with me. Thank you both for all the laughter and the great support! **Xiao**, it has been great to get to know you and I'm glad that you bring so much enthusiasm to the division and various activities around it. I still get a smile on my face when I think about our first online chat before you started. And thanks for the Chinese crackers, they are among my favorite snacks. **Jacob**, thank you for keeping the yeast work going, I look forward to seeing your peroxisomal results. It has been great to share the office with you and your positive attitude is contagious. Thank you for keeping the after-works going and arranging board game nights, it really makes TMB a better place. **Christina**, it has been great to share the office with you when you are in Lund and to have all the interesting discussions between writing. Thank you for the new research perspectives and lunches at Bryggan. **Caroline**, it is always nice when you come to TMB and hopefully there is more lab work to come, without contamination. Thanks for your enthusiastic and interesting seminars, it has been a great way to learn more about drinking water microbiology. **Essi**, it's great that someone like you started at TMB and you are a true source of joy. Thank you for engaging in all sorts of activities that we do together at the division. **Khaled**, thank you for showing me how to write my name in Arabic, you might have to show me again… Thank you for the enthusiasm and curiosity you show toward everyone's research. **Tadele** and **Getnet**, thank you for teaching me about Ethiopia, I will try the restaurant you recommended in Lund. To all the **PhD students at BAM**, thanks for the fun social

events, for being great people, for making good things happen and for supporting each other. I'm glad I got to know you all! **Amanda** and **Cecilia**, I will always be grateful for the teaching course we took together and the not so fun group project that gave me two incredible friends. Thanks for all the nice lunches and walks, thanks for all the tips about research grants, thanks for all the fun discussions and thanks for supporting and cheering me on when I've been stressed. My PhD journey has definitely been better with the two of you. **Edoardo**, thank you for being a great friend since the beginning of our PhDs and I hope it won't take 4 years until the next time we make homemade pasta.

Thank you to all the students we have had at TMB over the years, you have all been an important part of the division and what we do, and I'm happy I got to know so many of you. To all the students I've supervised in Microbiology and in Technical Biology, I hope your interest in microbiology started to grow, just like mine did, and that you enjoyed the courses as much as I did. Thank you for the curious questions about microbiology and my research. To my students, **Sid, Elsa** and **Milena**, thank you for choosing to do your projects with me. I enjoyed supervising you and I hope that I was able to teach you something valuable. **Sid** and **Elsa**, thank you for working with me on different parts of my yeast projects and I hope you learned something that you brought or your respective PhD journeys. **Milena**, thanks for working with me on my DNA polymerase projects and contributing to great results and figures, for many fun days in the lab and for becoming a friend.

To my past and present office mates, **Raquel, Tage, Mikael, Gisele, Christina, Jacob** and **Iria**, thanks for making the office a nice and fun place to work at and for allowing me to create a jungle.

Thanks to the **TMB running club**, the **Tabata gang** and **Gerdahallen crew** for making me exercise in great company during all these years. It has been so much more fun to go to the gym together with all of you!

To everyone that has been a part of the **Microbial Engineering group**, thank you for the great start of Friday mornings during all these years. To everyone that has been a part of the **Forensic group**, thank you for all the interesting discussions.

Thank you, **Courtney**, for helping me with fluorescence microscopy, creating beautiful pictures of my yeast cells and for the interesting discussions.

To Tetra Pak at TMB, **Birgitta, Hanna, Ilias** and **Ravinder**, thank you for the interesting chats at lunch and fika, covering everything from milk spoilage to olive oil production.

Danne, thanks for all the fun stories while waiting for the elevator.

Thanks to the R&D group at Genovis, **Rolf, Maria, Fredrik Leo, Emil** and **Pearse**, for making me feel like a part of the group, for all the fun meetings and your encouragement. Maria, thank you for always make me a "take away bag" when I

needed enzymes. **Emil** and **Pearse**, thanks for all the help in the lab at Genovis. **Fredrik Leo**, my fellow industrial PhD student in the group, thanks for showing me how work in the academia and the industry can be combined.

To everyone at **Infection Medicine**, thank you for sharing the lab with me and always showing me around when I needed something I couldn't find. Special thanks to **Maria** and **Isabella** for helping me out in the lab and for inviting me to YoungInfect@LU events. And thanks to **Martin** for de-staining my gel when I had to leave.

Thank you to everyone at **NFC** in Linköping that showed me around, explained their work and helped me in the lab there. Thank you, **Arvid**, for patiently helping me with the bioinformatic pipeline to analyze my data. I've learned a lot from you. Thank you, **Erland**, for always inviting me and checking in on how my projects are going. And **Adam**, although you are at RMV, thanks for showing me around and explaining your work and teaching me about SNP panels.

Thank you to everyone in the **Applied Genetics group at NIST** for providing a friendly and encouraging environment during my research stay and supporting my work. Special thanks to **Pete** and **Kevin**, for inviting me, for all your guidance, for helping me to push the boundaries of my research and for teaching me new skills. If I ever get the opportunity, I would love to come back!

To everyone that has been a part of the **FBD** projects, thank you for your kindness, support and motivation when I've presented my work at meetings, it has been a pleasure to work with all of you.

To **Niki** and **Stefan** at Simsen Diagnostics, thank you for providing the prototype assay and letting me perform the initial experiments at this early stage.

I also want to thank both the yeast and the forensic genetics community that I've had the privilege of meeting at international conferences. Thanks for all the fruitful discussions and valuable input on my work.

Thanks to all others who I did not mention but have been helpful in different ways.

Till alla mina vänner som har stöttat mig och fått mig att fullständigt koppla bort jobbet, tack! Utan er hade den här avhandlingen inte varit möjlig. **Sofia**, tack för att du varit min vän sedan dagis och alltid ställt upp för mig. Tack för alla fina upplevelser vi fått genom åren och tack för att du säger till att nu är det dags för en paus. **Johanna**, som vi har suttit ihop genom åren sedan grundskolan. Tack för alla kul stunder när jag behövt tänka på något annat. **Elsa, Kajsa, Natali, Linn, Lovisa, Johanna** och **Johan**, mitt Kattegäng, tack för all uppmuntran genom åren. **Linn**, tack för att du alltid är en källa till glädje och motivation. Tack för att du är en fin vän, tack för att du kollar hur jag mår och tack för alla otroliga minnen vi har tillsammans. **Lovisa**, tack för att du målat den otroliga framsidan till min avhandling, jag vet att du kämpade med den, men den blev jättefin! Tack för alla

fina promenader och för att jag får vara en del av dina barns liv och ta del av glädjen de sprider. **Johanna**, tack för alla fina telefonsamtal och spontana besök i både Lund och Stockholm. Tack för all pepp och tips på vad jag kan göra när jag blir klar. **Johan Harryson**, tack för alla skratt och för att du alltid tar tiden till att ses. **Sofie, Fanny** och **Vesa**, tack för att ni är så fina vänner och alltid visat förståelse för mig och varandra i alla olika situationer. Tack för härliga tider på BioInvent, under Lundakarnevalen och i Göteborg. Tack för alla fina stunder. Jag ser alltid fram emot att hänga med er. Tack **Steph** för alla härliga fikor och luncher i Stockholm, ska höra av mig i bättre tid nästa gång vi kommer upp. **Johan Hallbeck**, tack för att du varit en så bra vän sedan vi började på LTH tillsammans och tack för all den otroliga maten du lagat. Tack **Manfred**, för att jag vet att du kommer uppskatta min avhandling. **Amanda, Amanda, Linn, Melina** och **Y Nhi**, tack för alla fina stunder tillsammans, tack för att ni drog med mig i Lundakarnevalen en gång till och tack för fina firanden genom åren. **Klara**, tack för alla långa och välbehövliga promenader, fikor och middagar. Till mitt löpargäng, **Lovisa, Elin, Camilla**, **Mayra** och **Alice**, tack för att ni sett till att jag gått hem från jobbet i tid på tisdagar och för alla kilometrar under året.

Till hela min familj, tack för att ni varit det bästa stödet jag kan tänka mig. **Mamma** och **Pappa**, tack för att ni alltid trott på mig och uppmuntrat mig till att allt är möjligt. **Mamma**, tack för att du inspirerar mig, både inom forskningen och som människa och alltid finns ett samtal bort när jag behöver prata. Till mina småbröder som inte är så små längre, **Esbjörn, Arvid** och **Hjalmar**, tack för att ni alltid ställer upp för mig, oavsett om det gäller att bli hämtad sent, göra pannkakor till frukost, bygga och svetsa ett bord eller frakta hem stolar efter ett köp på Marketplace. Ni betyder otroligt mycket för mig. **Esbjörn**, tack för att du alltid öppnar upp ditt hem för oss, för att du fick oss att bestiga Kebnekaise och för att du ibland ger mig en reality check om att det jag håller på med är ganska häftigt. **Arvid**, tack för att du alltid har en praktisk lösning liggandes och inspirerar mig till att bygga saker själv. **Hjalmar**, tack för att du hänger med på diverse av mina förslag, från gymmet till plantagen och för att du går med på att vattna mina blommor under sommaren. Utan er hade jag inte varit den jag är idag.

Till hela familjen **Lund/Ahlgren**, tack för att ni alltid varit så välkomnande, stöttande och nyfikna på mitt arbete. **Johan** och **Ragnhild**, tack för att ni är de bästa svärföräldrarna jag kan önska mig och för att jag alltid kan komma till er när jag behöver en paus. **Arvid** och **Johanna**, tack för allt fint ni gjort för mig under åren och för alla middagar som slutat mycket senare än planerat.

Fredrik, min följeslagare i livet. Tack för att du alltid varit min trygga punkt och för att du alltid kan få mig på bra humör. Det har varit fantastiska år tillsammans på TMB och jag kommer alltid vara tacksam för att vi doktorerade samtidigt. Tack för att du alltid varit på min sida och att jag får dela livet med dig. Tack för att du firat alla framgångar med mig och alltid uppmuntrat mig till att fortsätta när det känts tungt. Jag ser fram emot alla våra kommande äventyr!

References

Abedi, D., Zhang, L., Pyne, M., & Perry Chou, C. (2011). 1.03 - Enzyme Biocatalysis. In M. Moo-Young (Ed.), *Comprehensive Biotechnology (Second Edition)* (pp. 15-24). Academic Press. https://doi.org/10.1016/B978-0-08-088504-9.00111-2

Adrio, J. L., & Demain, A. L. (2014). Microbial Enzymes: Tools for Biotechnological Processes. *Biomolecules*, *4*(1), 117-139. https://doi.org/10.3390/biom4010117

Aehle, W. (2007). *Enzymes in industry: production and applications*. John Wiley & Sons.

Afrin, S., Tamanna, T., Shahajadi, U. F., Bhowmik, B., Jui, A. H., Miah, M. A. S., & Bhuiyan, M. N. I. (2024). Characterization of protease-producing bacteria from garden soil and antagonistic activity against pathogenic bacteria. *The Microbe*, *4*, 100123. https://doi.org/10.1016/j.microb.2024.100123

Agudo, M. M., Aanes, H., Roseth, A., Albert, M., Gill, P., & Bleka, Ø. (2022). A comprehensive characterization of MPS-STR stutter artefacts. *Forensic Science International: Genetics*, *60*, 102728. https://doi.org/10.1016/j.fsigen.2022.102728

Akram, F., Shah, F. I., Ibrar, R., Fatima, T., Haq, I. u., Naseem, W., Gul, M. A., Tehreem, L., & Haider, G. (2023). Bacterial thermophilic DNA polymerases: A focus on prominent biotechnological applications. *Analytical Biochemistry*, *671*, 115150. https://doi.org/10.1016/j.ab.2023.115150

Al-Mansoori, A. K. K., Kang, S. Y., Thangadurai, S., & Rahim, R. A. (2025). Exploring enzyme thermostability enhancement through molecular modification: What remains to be known? *Journal of Molecular Structure*, *1337*, 141852. https://doi.org/10.1016/j.molstruc.2025.141852

Al-Salama, Z. T. (2020). Imlifidase: First Approval. *Drugs*, *80*(17), 1859-1864. https://doi.org/10.1007/s40265-020-01418-5

Alonso, A., Barrio, P. A., Müller, P., Köcher, S., Berger, B., Martin, P., Bodner, M., Willuweit, S., Parson, W., Roewer, L., & Budowle, B. (2018). Current state-of-art of STR sequencing in forensic genetics. *Electrophoresis*, *39*(21), 2655-2668. https://doi.org/10.1002/elps.201800030

Arakawa, T., & Akuta, T. (2025). Beyond Purification: Evolving Roles of Fusion Tags in Biotechnology. *Current Issues in Molecular Biology*, *47*(9), 768. https://www.mdpi.com/1467-3045/47/9/768

Arezi, B., Xing, W., Sorge, J. A., & Hogrefe, H. H. (2003). Amplification efficiency of thermostable DNA polymerases. *Analytical Biochemistry*, *321*(2), 226-235. https://doi.org/10.1016/S0003-2697(03)00465-2

Aschenbrenner, J., & Marx, A. (2017). DNA polymerases and biotechnological applications. *Current Opinion in Biotechnology*, *48*, 187-195. https://doi.org/10.1016/j.copbio.2017.04.005

Baker, T. A., & Bell, S. P. (1998). Polymerases and the Replisome: Machines within Machines. *Cell*, *92*(3), 295-305. https://doi.org/10.1016/S0092-8674(00)80923-X

Bandah-Rozenfeld, D., Voskoboinik, L., Rashkovski, K., Avrahami, K., Schayek, H., & Avlas, O. (2025). The power of STR discrimination: Two brothers with high-profile similarity. *Journal of Forensic Sciences*, *70*(4), 1622-1628. https://doi.org/10.1111/1556-4029.70060

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology*, *10*(5), 411-421. https://doi.org/10.1016/S0958-1669(99)00003-8

Baneyx, F., & Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. *Nature Biotechnology*, *22*(11), 1399-1408. https://doi.org/10.1038/nbt1029

Bar-Peled, L., & Kory, N. (2022). Principles and functions of metabolic compartmentalization. *Nature Metabolism*, *4*(10), 1232-1244. https://doi.org/10.1038/s42255-022-00645-2

Bartlett, J. M., & Stirling, D. (2003). A short history of the polymerase chain reaction. In *PCR protocols* (pp. 3-6). Springer.

Bebenek, K., & Kunkel, T. A. (2004). Functions of DNA Polymerases. In *Advances in Protein Chemistry* (Vol. 69, pp. 137-165). Academic Press. https://doi.org/10.1016/S0065-3233(04)69005-X

Beck, A., Wagner-Rousset, E., Ayoub, D., Van Dorsselaer, A., & Sanglier-Cianférani, S. (2013). Characterization of Therapeutic Antibodies and Related Products. *Analytical Chemistry*, *85*(2), 715-736. https://doi.org/10.1021/ac3032355

Benkovic, S. J., Valentine, A. M., & Salinas, F. (2001). Replisome-Mediated DNA Replication. *Annual Review of Biochemistry*, *70*, 181-208. https://doi.org/10.1146/annurev.biochem.70.1.181

Berdis, A. J. (2009). Mechanisms of DNA Polymerases. *Chemical Reviews*, *109*(7), 2862-2879. https://doi.org/10.1021/cr800530b

Beygmoradi, A., Homaei, A., Hemmati, R., & Fernandes, P. (2023). Recombinant protein expression: Challenges in production and folding related matters. *International Journal of Biological Macromolecules*, *233*, 123407. https://doi.org/10.1016/j.ijbiomac.2023.123407

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X. Z., & Zhou, T. (2021). Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in *Escherichia coli* and Strategies to Address Them for Industrial Applications. *Frontiers in Bioengineering and Biotechnology*, *9*, 630551. https://doi.org/10.3389/fbioe.2021.630551

Biancalana, V., & Laporte, J. (2015). Diagnostic use of Massively Parallel Sequencing in Neuromuscular Diseases: Towards an Integrated Diagnosis. *Journal of Neuromuscular Diseases*, *2*(3), 193-203. https://doi.org/10.3233/jnd-150092

Bille, T., Bright, J.-A., & Buckleton, J. (2013). Application of Random Match Probability Calculations to Mixed STR Profiles. *Journal of Forensic Sciences*, *58*(2), 474-485. https://doi.org/10.1111/1556-4029.12067

Binod, P., Palkhiwala, P., Gaikaiwari, R., Nampoothiri, K. M., Duggal, A., Dey, K., & Pandey, A. (2013). Industrial enzymes-present status and future perspectives for India. *Journal of Scientific and Industrial Research*, *72*(5), 271-286.

Birch, D. E., Kolmodin, L., Wong, J., Zangenberg, G. A., Zoccoli, M. A., McKinney, N., & Young, K. K. Y. (1996). Simplified hot start PCR. *Nature*, *381*(6581), 445-446. https://doi.org/10.1038/381445a0

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., & Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, *277*(5331), 1453-1462. https://doi.org/10.1126/science.277.5331.1453

Bleka, Ø., Just, R., Le, J., & Gill, P. (2020). An examination of STR nomenclatures, filters and models for MPS mixture interpretation. *Forensic Science International: Genetics*, *48*, 102319. https://doi.org/10.1016/j.fsigen.2020.102319

Blöchl, C., Holzner, C., Luciano, M., Bauer, R., Horejs-Hoeck, J., Eckhard, U., Brandstetter, H., & Huber, C. G. (2022). Proteolytic Profiling of Streptococcal Pyrogenic Exotoxin B (SpeB) by Complementary HPLC-MS Approaches. *International Journal of Molecular Sciences*, *23*(1), 412. https://www.mdpi.com/1422-0067/23/1/412

Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J., & Schäfer, F. (2009). Chapter 27 Immobilized-Metal Affinity Chromatography (IMAC): A Review. In R. R. Burgess & M. P. Deutscher (Eds.), *Methods in Enzymology* (Vol. 463, pp. 439-473). Academic Press. https://doi.org/10.1016/S0076-6879(09)63027-5

Blow, D. M., Birktoft, J. J., & Hartley, B. S. (1969). Role of a Buried Acid Group in the Mechanism of Action of Chymotrypsin. *Nature*, *221*(5178), 337-340. https://doi.org/10.1038/221337a0

Bond, J. S. (2019). Proteases: History, discovery, and roles in health and disease. *Journal of Biological Chemistry*, *294*(5), 1643-1651. https://doi.org/10.1074/jbc.TM118.004156

Borhani, S., & Arab, S. S. (2023). Investigation of thermal stability characteristic in family A DNA polymerase - A theoretical study. *Progress in Biophysics and Molecular Biology*, *182*, 15-25. https://doi.org/10.1016/j.pbiomolbio.2023.05.003

Bornhorst, J. A., & Falke, J. J. (2000). [16] Purification of proteins using polyhistidine affinity tags. In *Methods in Enzymology* (Vol. 326, pp. 245-254). Academic Press. https://doi.org/10.1016/S0076-6879(00)26058-8

Borodina, I., & Nielsen, J. (2014). Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. *Biotechnology Journal.*, *9*(5), 609-620. https://doi.org/10.1002/biot.201300445

Böttger, E. C., Springer, B., Prammananan, T., Kidan, Y., & Sander, P. (2001). Structural basis for selectivity and toxicity of ribosomal antibiotics. *EMBO Reports*, *2*(4), 318-323. https://doi.org/10.1093/embo-reports/kve062

Braithwaite, D. K., & Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Research*, *21*(4), 787-802. https://doi.org/10.1093/nar/21.4.787

Brannigan, J. A., Dodson, G., Duggleby, H. J., Moody, P. C. E., Smith, J. L., Tomchick, D. R., & Murzin, A. G. (1995). A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature*, *378*(6555), 416-419. https://doi.org/10.1038/378416a0

Bratanis, E., Molina, H., Naegeli, A., Collin, M., & Lood, R. (2017). BspK, a Serine Protease from the Predatory Bacterium *Bdellovibrio bacteriovorus* with Utility for Analysis of Therapeutic Antibodies. *Applied and Environmental Microbiology*, *83*(4). https://doi.org/10.1128/aem.03037-16

Briand, L., Marcion, G., Kriznik, A., Heydel, J. M., Artur, Y., Garrido, C., Seigneuric, R., & Neiers, F. (2016). A self-inducible heterologous protein expression system in *Escherichia coli*. *Scientific Reports*, *6*(1), 33037. https://doi.org/10.1038/srep33037

Brock, T. D., & Edwards, M. R. (1970). Fine Structure of *Thermus aquaticus*, an Extreme Thermophile. *Journal of Bacteriology*, *104*(1), 509-517. https://doi.org/10.1128/jb.104.1.509-517.1970

Brock, T. D., & Freeze, H. (1969). *Thermus aquaticus* gen. n. and sp. n., a Nonsporulating Extreme Thermophile. *Journal of Bacteriology*, *98*(1), 289-297. https://doi.org/10.1128/jb.98.1.289-297.1969

Brown, H. S., & LiCata, V. J. (2013). Enthalpic switch-points and temperature dependencies of DNA binding and nucleotide incorporation by Pol I DNA polymerases. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1834*(10), 2133-2138. https://doi.org/10.1016/j.bbapap.2013.06.021

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis*, *39*(21), 2642-2654. https://doi.org/10.1002/elps.201800082

Brumm, P. J., Monsma, S., Keough, B., Jasinovica, S., Ferguson, E., Schoenfeld, T., Lodes, M., & Mead, D. A. (2015). Complete Genome Sequence of *Thermus aquaticus* Y51MC23. *PLoS One*, *10*(10), e0138674. https://doi.org/10.1371/journal.pone.0138674

Bustos, C., Quezada, J., Veas, R., Altamirano, C., Braun-Galleani, S., Fickers, P., & Berrios, J. (2022). Advances in Cell Engineering of the *Komagataella phaffii* Platform for Recombinant Protein Production. *Metabolites*, *12*(4), 346. https://www.mdpi.com/2218-1989/12/4/346

Butler, J. M. (2005). Constructing STR multiplex assays. *Methods in Molecular Biology*, *297*, 53-66. https://doi.org/10.1385/1-59259-867-6:053

Butler, J. M., Buel, E., Crivellente, F., & McCord, B. R. (2004). Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis*, *25*(10-11), 1397-1412. https://doi.org/10.1002/elps.200305822

Cai, P., Gao, J., & Zhou, Y. (2019). CRISPR-mediated genome editing in non-conventional yeasts for biotechnological applications. *Microbial Cell Factories*, *18*(1), 63. https://doi.org/10.1186/s12934-019-1112-2

Carnes, A. E. (2005). Fermentation design for the manufacture of therapeutic plasmid DNA. *BioProcess International*, *3*(9), 36-44.

Carroll, R. K., & Musser, J. M. (2011). From transcription to activation: how group A *Streptococcus*, the flesh-eating pathogen, regulates SpeB cysteine protease production. *Molecular Microbiology.*, *81*(3), 588-601. https://doi.org/10.1111/j.1365-2958.2011.07709.x

Carter, M., & Shieh, J. (2015). Chapter 12 - Making and Using Transgenic Organisms. In M. Carter & J. Shieh (Eds.), *Guide to Research Techniques in Neuroscience (Second Edition)* (pp. 253-271). Academic Press. https://doi.org/10.1016/B978-0-12-800511-8.00012-5

Case, B. C., & Hingorani, M. M. (2017). Polymerase. In *Reference Module in Life Sciences*. Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.06928-4

Chamberlain, J. S., Gibbs, R. A., Rainer, J. E., Nguyen, P. N., & Thomas, C. (1988). Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Research*, *16*(23), 11141-11156. https://doi.org/10.1093/nar/16.23.11141

Charlton, A., & Zachariou, M. (2008). Immobilized Metal Ion Affinity Chromatography of Histidine-Tagged Fusion Proteins. In M. Zachariou (Ed.), *Affinity Chromatography: Methods and Protocols* (pp. 137-150). Humana Press. https://doi.org/10.1007/978-1-59745-582-4_10

Cherepanov, A. V., & de Vries, S. (2003). Kinetics and thermodynamics of nick sealing by T4 DNA ligase. *European Journal of Biochemistry*, *270*(21), 4315-4325. https://doi.org/10.1046/j.1432-1033.2003.03824.x

Chien, A., Edgar, D. B., & Trela, J. M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *Journal of Bacteriology*, *127*(3), 1550-1557. https://doi.org/10.1128/jb.127.3.1550-1557.1976

Clark, D. P., Pazdernik, N. J., & McGehee, M. R. (2019). Chapter 6 - Polymerase Chain Reaction. In D. P. Clark, N. J. Pazdernik, & M. R. McGehee (Eds.), *Molecular Biology (Third Edition)* (pp. 168-198). Academic Cell. https://doi.org/10.1016/B978-0-12-813288-3.00006-9

Clark, E. D. B. (2001). Protein refolding for industrial processes. *Current Opinion in Biotechnology*, *12*(2), 202-207. https://doi.org/10.1016/S0958-1669(00)00200-7

Cline, J., Braman, J. C., & Hogrefe, H. H. (1996). PCR Fidelity of *Pfu* DNA Polymerase and Other Thermostable DNA Polymerases. *Nucleic Acids Research*, *24*(18), 3546-3551. https://doi.org/10.1093/nar/24.18.3546

Collin, M., & Kilian, M. (2014). Chapter 18 - Bacterial Modulation of Fc Effector Functions. In M. E. Ackerman & F. Nimmerjahn (Eds.), *Antibody Fc* (pp. 317-332). Academic Press. https://doi.org/10.1016/B978-0-12-394802-1.00018-2

Collin, M., & Olsén, A. (2001). Effect of SpeB and EndoS from *Streptococcus pyogenes* on human immunoglobulins. *Infection and Immunity*, *69*(11), 7187-7189. https://doi.org/10.1128/iai.69.11.7187-7189.2001

Collins, M., & Khalili, H. (2022). Soluble Papain to Digest Monoclonal Antibodies; Time and Cost-Effective Method to Obtain Fab Fragment. *Bioengineering*, *9*(5), 209. https://doi.org/10.3390/bioengineering9050209

Contesini, F. J., Melo, R. R. d., & Sato, H. H. (2018). An overview of *Bacillus* proteases: from production to application. *Critical Reviews in Biotechnology.*, *38*(3), 321-334. https://doi.org/10.1080/07388551.2017.1354354

Cooper, G. M. (2000). *The Cell: A Molecular Approach. 2nd edition*. Sinauer Associates 2000. http://lib.ugent.be/catalog/ebk01:3450000000002155

Coticone, S. R., & Bloch, W. (2014). *Methods for the reduction of stutter in microsatellite amplification*. United States Patent and Trademark Office. https://patents.google.com/patent/US8785126B2

Coulther, T. A., Stern, H. R., & Beuning, P. J. (2019). Engineering Polymerases for New Functions. *Trends in Biotechnology*, *37*(10), 1091-1103. https://doi.org/10.1016/j.tibtech.2019.03.011

Courtney, N. A., Encell, L. P., McLaren, R. A., Mokry, D. Z., Zimmerman, K., Schnell, K., Killoran, M., Knoche, K. K., Slater, M. R., Wood, M. G., & Brockman, M. (2024). *Engineered DNA polymerase with reduced artifact formation*. United States Patent and Trademark Office. https://patents.google.com/patent/US20240344039A1/

Culp, E., & Wright, G. D. (2017). Bacterial proteases, untapped antimicrobial drug targets. *The Journal of Antibiotics*, *70*(4), 366-377. https://doi.org/10.1038/ja.2016.138

Da Silva, N. A., & Srikrishnan, S. (2012). Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. *FEMS Yeast Research*, *12*(2), 197-214. https://doi.org/10.1111/j.1567-1364.2011.00769.x

Datta, K., & LiCata, V. J. (2003). Thermodynamics of the binding of *Thermus aquaticus* DNA polymerase to primed-template DNA. *Nucleic Acids Research*, *31*(19), 5590-5597. https://doi.org/10.1093/nar/gkg774

Dau, T., Bartolomucci, G., & Rappsilber, J. (2020). Proteomics Using Protease Alternatives to Trypsin Benefits from Sequential Digestion with Trypsin. *Analytical Chemistry*, *92*(14), 9523-9527. https://doi.org/10.1021/acs.analchem.0c00478

Davey, H. M., & Hexley, P. (2011). Red but not dead? Membranes of stressed *Saccharomyces cerevisiae* are permeable to propidium iodide. *Environmental Microbiology*, *13*(1), 163-171. https://doi.org/10.1111/j.1462-2920.2010.02317.x

Davidson, J. F., Fox, R., Harris, D. D., Lyons-Abbott, S., & Loeb, L. A. (2003). Insertion of the T3 DNA polymerase thioredoxin binding domain enhances the processivity and fidelity of *Taq* DNA polymerase. *Nucleic Acids Research*, *31*(16), 4702-4709. https://doi.org/10.1093/nar/gkg667

De Brabander, P., Uitterhaegen, E., Delmulle, T., De Winter, K., & Soetaert, W. (2023). Challenges and progress towards industrial recombinant protein production in yeasts: A review. *Biotechnology Advances*, *64*, 108121. https://doi.org/10.1016/j.biotechadv.2023.108121

de Knijff, P. (2019). From next generation sequencing to now generation sequencing in forensics. *Forensic Science International: Genetics*, *38*, 175-180. https://doi.org/10.1016/j.fsigen.2018.10.017

del Prado, A., Franco-Echevarría, E., González, B., Blanco, L., Salas, M., & de Vega, M. (2018). Noncatalytic aspartate at the exonuclease domain of proofreading DNA polymerases regulates both degradative and synthetic activities. *Proceedings of the National Academy of Sciences*, *115*(13), E2921-E2929. https://doi.org/10.1073/pnas.1718787115

del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M., & Díaz-Orejas, R. (1998). Replication and control of circular bacterial plasmids. *Microbiology and Molecular Biology Reviews*, *62*(2), 434-464. https://doi.org/10.1128/mmbr.62.2.434-464.1998

Denamur, E., & Matic, I. (2006). Evolution of mutation rates in bacteria. *Molecular Microbiology*, *60*(4), 820-827. https://doi.org/10.1111/j.1365-2958.2006.05150.x

Devesse, L., Davenport, L., Borsuk, L., Gettings, K., Mason-Buck, G., Vallone, P. M., Syndercombe Court, D., & Ballard, D. (2020). Classification of STR allelic variation using massively parallel sequencing and assessment of flanking region power. *Forensic Science International: Genetics*, *48*, 102356. https://doi.org/10.1016/j.fsigen.2020.102356

Dewangan, N., Sahu, K., & Bhardwaj, H. (2020). Revolutionizing Molecular Biology: The Evolution of PCR through Troubleshooting and Optimization. *NewBioWorld*, *2*, 1-17. https://doi.org/10.52228/NBW-JAAB.2020-2-1-3

Di Cera, E. (2009). Serine proteases. *IUBMB Life*, *61*(5), 510-515. https://doi.org/10.1002/iub.186

Dierig, L., Kunz, S. N., & Wiegand, P. (2024). Comparison of massively parallel sequencing to capillary electrophoresis for short tandem repeat genotyping of trace DNA. *Electrophoresis*, *45*(5-6), 451-462. https://doi.org/10.1002/elps.202300145

Dos Santos, A. L. (2011). Protease expression by microorganisms and its relevance to crucial physiological/pathological events. *World Journal of Biological Chemistry*, *2*(3), 48-58. https://doi.org/10.4331/wjbc.v2.i3.48

Dyer, R. P., & Weiss, G. A. (2022). Making the cut with protease engineering. *Cell Chemical Biology*, *29*(2), 177-190. https://doi.org/10.1016/j.chembiol.2021.12.001

Dymond, J. S. (2013). Chapter Twenty Three - Explanatory Chapter: Quantitative PCR. In J. Lorsch (Ed.), *Methods in Enzymology* (Vol. 529, pp. 279-289). Academic Press. https://doi.org/10.1016/B978-0-12-418687-3.00023-9

Ebeling, W., Hennrich, N., Klockow, M., Metz, H., Orth, H. D., & Lang, H. (1974). Proteinase K from *Tritirachium album* Limber. *European Journal of Biochemistry*, *47*(1), 91-97. https://doi.org/10.1111/j.1432-1033.1974.tb03671.x

Eckert, K. A., & Kunkel, T. A. (1990). High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Research*, *18*(13), 3739-3744. https://doi.org/10.1093/nar/18.13.3739

Eckert, K. A., & Kunkel, T. A. (1991). DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods and Applications*, *1*(1), 17-24. https://doi.org/10.1101/gr.1.1.17

Eun, H.-M. (1996). 6 - DNA Polymerases. In H.-M. Eun (Ed.), *Enzymology Primer for Recombinant DNA Technology* (pp. 345-489). Academic Press. https://doi.org/10.1016/B978-012243740-3/50009-0

Farooq, H., Ahmed, U., Fatima, Z., & Tabish, M. (2024). A Comprehensive Review of *Pfu* DNA Polymerase: Extraction, Production and Applications. *International Journal of Science and Research Archive*. https://doi.org/10.30574/ijsra.2024.13.2.2203

Farooqui, A. K., Ahmad, H., Rehmani, M. U., & Husain, A. (2023). Quick and easy method for extraction and purification of *Pfu*-*S*so7d, a high processivity DNA polymerase. *Protein Expression and Purification*, *208-209*, 106276. https://doi.org/10.1016/j.pep.2023.106276

Fasim, A., More, V. S., & More, S. S. (2021). Large-scale production of enzymes for biotechnology uses. *Current Opinion in Biotechnology*, *69*, 68-76. https://doi.org/10.1016/j.copbio.2020.12.002

Fazekas, A. J., Royce, S., & and Newmaster, S. G. (2010). Improving Sequencing Quality from PCR Products Containing Long Mononucleotide Repeats. *Biotechniques*, *48*(4), 277-285. https://doi.org/10.2144/000113369

FDA. (1998). *Guidance for Human Somatic Cell Therapy and Gene Therapy*. https://www.fda.gov/cber/gdlns/somgene.pdf

Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E., & Villaverde, A. (2009). Microbial factories for recombinant pharmaceuticals. *Microbial Cell Factories*, *8*(1), 17. https://doi.org/10.1186/1475-2859-8-17

Fiala, G., & Stetter, K. O. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Archives of Microbiology*, *145*(1), 56-61. https://doi.org/10.1007/BF00413027

Fikes, J. D., Barkocy-Gallagher, G. A., Klapper, D. G., & Bassford, P. J. (1990). Maturation of *Escherichia coli* maltose-binding protein by signal peptidase I in vivo. Sequence requirements for efficient processing and demonstration of an alternate cleavage site. *Journal of Biological Chemistry*, *265*(6), 3417-3423. https://doi.org/10.1016/S0021-9258(19)39783-2

Filges, S., Yamada, E., Ståhlberg, A., & Godfrey, T. E. (2019). Impact of Polymerase Fidelity on Background Error Rates in Next-Generation Sequencing with Unique Molecular Identifiers/Barcodes. *Scientific Reports*, *9*(1), 3503. https://doi.org/10.1038/s41598-019-39762-6

Fungtammasan, A., Ananda, G., Hile, S. E., Su, M. S., Sun, C., Harris, R., Medvedev, P., Eckert, K., & Makova, K. D. (2015). Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Research*, *25*(5), 736-749. https://doi.org/10.1101/gr.185892.114

Garcia-Diaz, M., & Bebenek, K. (2007). Multiple functions of DNA polymerases. *Critical Reviews in Plant Sciences*, *26*(2), 105-122. https://doi.org/10.1080/07352680701252817

Gautam, A. (2022). DNA Isolation by Lysozyme and Proteinase K. In *DNA and RNA Isolation Techniques for Non-Experts* (pp. 85-88). Springer International Publishing. https://doi.org/10.1007/978-3-030-94230-4_11

Gawlitzek, M., Wagner, R., Conradt, H. S., Villers, C., & Verbert, A. (1994). Changes in the glycosylation pattern of recombinant proteins effected by defined culture conditions of BHK-21 cells. In R. E. Spier, J. B. Griffiths, & W. Berthold (Eds.), *Animal Cell Technology* (pp. 649-651). Butterworth-Heinemann. https://doi.org/10.1016/B978-0-7506-1845-8.50147-2

Giansanti, P., Tsiatsiani, L., Low, T. Y., & Heck, A. J. R. (2016). Six alternative proteases for mass spectrometry–based proteomics beyond trypsin. *Nature Protocols*, *11*(5), 993-1006. https://doi.org/10.1038/nprot.2016.057

Gnügge, R., & Rudolf, F. (2017). *Saccharomyces cerevisiae* Shuttle vectors. *Yeast*, *34*(5), 205-221. https://doi.org/10.1002/yea.3228

Goldberg, A. (2004). Protein Degradation and Protection against Misfolded or Damaged Proteins. *Nature*, *426*, 895-899. https://doi.org/10.1038/nature02263

Goldenberger, D., Perschil, I., Ritzler, M., & Altwegg, M. (1995). A simple" universal" DNA extraction procedure using SDS and proteinase K is compatible with direct PCR amplification. *Genome Research*, *4*(6), 368-370.

Goldstein, M. A., & Doi, R. H. (1995). Prokaryotic promoters in biotechnology. In M. R. El-Gewely (Ed.), *Biotechnology Annual Review* (Vol. 1, pp. 105-128). Elsevier. https://doi.org/10.1016/S1387-2656(08)70049-8

González-Páez, G. E., & Wolan, D. W. (2012). Ultrahigh and high resolution structures and mutational analysis of monomeric *Streptococcus pyogenes* SpeB reveal a functional role for the glycine-rich C-terminal loop. *Journal of Biological Chemistry*, *287*(29), 24412-24426. https://doi.org/10.1074/jbc.M112.361576

Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B. M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., dhe-Paganon, S., Park, H.-w., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R.,…Complexes, S. (2008). Protein production and purification. *Nature Methods*, *5*(2), 135-146. https://doi.org/10.1038/nmeth.f.202

Greci, M. D., Dooher, J. D., & Bell, S. D. (2022). The combined DNA and RNA synthetic capabilities of archaeal DNA primase facilitate primer hand-off to the replicative DNA polymerase. *Nature Communications*, *13*(1), 433. https://doi.org/10.1038/s41467-022-28093-2

Grewal, P. S., Samson, J. A., Baker, J. J., Choi, B., & Dueber, J. E. (2021). Peroxisome compartmentalization of a toxic enzyme improves alkaloid production. *Nature Chemical Biology*, *17*(1), 96-103. https://doi.org/10.1038/s41589-020-00668-4

Gubba, S., Low, D. E., & Musser, J. M. (1998). Expression and characterization of group A *Streptococcus* extracellular cysteine protease recombinant mutant proteins and documentation of seroconversion during human invasive disease episodes. *Infection and Immunity*, *66*(2), 765-770. https://doi.org/10.1128/iai.66.2.765-770.1998

Gurung, N., Ray, S., Bose, S., & Rai, V. (2013). A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond. *BioMed Research International*, *2013*(1), 329121. https://doi.org/10.1155/2013/329121

Gymrek, M. (2017). A genomic view of short tandem repeats. *Current Opinion in Genetics & Development*, *44*, 9-16. https://doi.org/10.1016/j.gde.2017.01.012

Haber, J. E. (2018). DNA Repair: The Search for Homology. *Bioessays*, *40*(5), e1700229. https://doi.org/10.1002/bies.201700229

Halbach, A., Rucktäschel, R., Rottensteiner, H., & Erdmann, R. (2009). The N-domain of Pex22p Can Functionally Replace the Pex3p N-domain in Targeting and Peroxisome Formation. *Journal of Biological Chemistry*, *284*(6), 3906-3916. https://doi.org/10.1074/jbc.M806950200

Hamdan, S. M., & van Oijen, A. M. (2010). Timing, Coordination, and Rhythm: Acrobatics at the DNA Replication Fork. *Journal of Biological Chemistry*, *285*(25), 18979-18983. https://doi.org/10.1074/jbc.R109.022939

Hamilton, S. C., Farchaus, J. W., & Davis, M. C. (2001). DNA Polymerases as Engines for Biotechnology. *Biotechniques*, *31*(2), 370-383. https://doi.org/10.2144/01312rv01

Han, J. Y., Seo, S. H., Song, J. M., Lee, H., & Choi, E.-S. (2018). High-level recombinant production of squalene using selected *Saccharomyces cerevisiae* strains. *Journal of Industrial Microbiology and Biotechnology*, *45*(4), 239-251. https://doi.org/10.1007/s10295-018-2018-4

Happonen, L., & Collin, M. (2024). Immunomodulating enzymes from *Streptococcus pyogenes*—In pathogenesis, as biotechnological tools, and as biological drugs. *Microorganisms*, *12*(1), 200. https://doi.org/10.3390/microorganisms12010200

Harwood, C. R., & Kikuchi, Y. (2022). The ins and outs of *Bacillus* proteases: activities, functions and commercial significance. *FEMS Microbiology Reviews*, *46*(1). https://doi.org/10.1093/femsre/fuab046

Hauge, X. Y., & Litt, M. (1993). A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. *Human Molecular Genetics*, *2*(4), 411-415. https://doi.org/10.1093/hmg/2.4.411

Hecht, K. A., O'Donnell, A. F., & Brodsky, J. L. (2014). The proteolytic landscape of the yeast vacuole. *Cellular Logistics*, *4*(1), e28023. https://doi.org/10.4161/cl.28023

Hedstrom, L. (2002). Serine Protease Mechanism and Specificity. *Chemical Reviews*, *102*(12), 4501-4524. https://doi.org/10.1021/cr000033x

Heid, C. A., Stevens, J., Livak, K. J., & Williams, P. M. (1996). Real time quantitative PCR. *Genome Research*, *6*(10), 986-994. https://doi.org/10.1101/gr.6.10.986

Henegariu, O., Heerema, N. A., Dlouhy, S. R., Vance, G. H., & Vogt, P. H. (1997). Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques*, *23*(3), 504-511. https://doi.org/10.2144/97233rr01

Hengge, R., & Turgay, K. (2009). Proteolysis in prokaryotes – from molecular machines to a systems perspective. *Research in Microbiology*, *160*(9), 615-617. https://doi.org/10.1016/j.resmic.2009.09.007

Herencias, C., Salgado-Briegas, S., Prieto, M. A., & Nogales, J. (2020). Providing new insights on the biphasic lifestyle of the predatory bacterium *Bdellovibrio bacteriovorus* through genome-scale metabolic modeling. *PLOS Computational Biology*, *16*(9), e1007646. https://doi.org/10.1371/journal.pcbi.1007646

Hernández-Rollán, C., Ehrmann, A. K., Vlassis, A., Kandasamy, V., & Nørholm, M. H. H. (2024). *Neq*2X7: a multi-purpose and open-source fusion DNA polymerase for advanced DNA engineering and diagnostics PCR. *BMC Biotechnology*, *24*(1), 17. https://doi.org/10.1186/s12896-024-00844-7

Heyde, S. A. H., & Nørholm, M. H. H. (2021). Tailoring the evolution of BL21(DE3) uncovers a key role for RNA stability in gene expression toxicity. *Communications Biology*, *4*(1), 963. https://doi.org/10.1038/s42003-021-02493-4

Hill, C. R., Butler, J. M., & Vallone, P. M. (2009). A 26plex autosomal STR assay to aid human identity testing. *Journal of Forensic Sciences*, *54*(5), 1008-1015. https://doi.org/10.1111/j.1556-4029.2009.01110.x

Hilz, H., Wiegers, U., & Adamietz, P. (1975). Stimulation of Proteinase K Action by Denaturing Agents: Application to the Isolation of Nucleic Acids and the Degradation of 'Masked' Proteins. *European Journal of Biochemistry*, *56*(1), 103-108. https://doi.org/10.1111/j.1432-1033.1975.tb02211.x

Hinnebusch, A. G., & Lorsch, J. R. (2012). The mechanism of eukaryotic translation initiation: new insights and challenges. *Cold Spring Harbor Perspectives in Biology*, *4*(10). https://doi.org/10.1101/cshperspect.a011544

Hou, J., Tyo, K. E. J., Liu, Z., Petranovic, D., & Nielsen, J. (2012). Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae*. *FEMS Yeast Research*, *12*(5), 491-510. https://doi.org/10.1111/j.1567-1364.2012.00810.x

Hu, K. K. Y., Suri, A., Dumsday, G., & Haritos, V. S. (2024). Cross-feeding promotes heterogeneity within yeast cell populations. *Nature Communications*, *15*(1), 418. https://doi.org/10.1038/s41467-023-44623-y

Huang, C.-J., Lin, H., & Yang, X. (2012). Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *Journal of Industrial Microbiology and Biotechnology*, *39*(3), 383-399. https://doi.org/10.1007/s10295-011-1082-9

Huang, E., Maldonado, A. Q., Kjellman, C., & Jordan, S. C. (2022). Imlifidase for the treatment of anti-HLA antibody-mediated processes in kidney transplantation. *American Journal of Transplantation*, *22*(3), 691-697. https://doi.org/10.1111/ajt.16828

Im, H., Dwidar, M., & Mitchell, R. J. (2018). Bdellovibrio bacteriovorus HD100, a predator of Gram-negative bacteria, benefits energetically from Staphylococcus aureus biofilms without predation. *The ISME Journal*, *12*(8), 2090-2095. https://doi.org/10.1038/s41396-018-0154-5

İncir, İ., & Kaplan, Ö. (2024). *Escherichia coli* as a versatile cell factory: Advances and challenges in recombinant protein production. *Protein Expression and Purification*, *219*, 106463. https://doi.org/10.1016/j.pep.2024.106463

Ishino, S., & Ishino, Y. (2014). DNA polymerases as useful reagents for biotechnology - the history of developmental research in the field. *Frontiers in Microbiology*, *5*, 465. https://doi.org/10.3389/fmicb.2014.00465

Ishino, Y., Iwasaki, H., Kato, I., & Shinagawa, H. (1994). Amino acid sequence motifs essential to 3'-->5' exonuclease activity of *Escherichia coli* DNA polymerase II. *Journal of Biological Chemistry*, *269*(20), 14655-14660. https://doi.org/10.1016/S0021-9258(17)36675-9

Iyer, P. V., & Ananthanarayan, L. (2008). Enzyme stability and stabilization—Aqueous and non-aqueous environment. *Process Biochemistry*, *43*(10), 1019-1032. https://doi.org/10.1016/j.procbio.2008.06.004

Jacobsen, H., Klenow, H., & Overgaard-Hansen, K. (1974). The N-terminal amino-acid sequences of DNA polymerase I from *Escherichia coli* and of the large and the small fragments obtained by a limited proteolysis. *European Journal of Biochemistry*, *45*(2), 623-627. https://doi.org/10.1111/j.1432-1033.1974.tb03588.x

Jensen, N. B., Strucko, T., Kildegaard, K. R., David, F., Maury, J., Mortensen, U. H., Forster, J., Nielsen, J., & Borodina, I. (2014). EasyClone: method for iterative chromosomal integration of multiple genes *Saccharomyces cerevisiae*. *FEMS Yeast Research*, *14*(2), 238-248. https://doi.org/10.1111/1567-1364.12118

Jessop-Fabre, M. M., Jakociunas, T., Stovicek, V., Dai, Z. J., Jensen, M. K., Keasling, J. D., & Borodina, I. (2016). EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes into *Saccharomyces cerevisiae* via CRISPR-Cas9. *Biotechnology Journal*, *11*(8), 1110-1117. https://doi.org/10.1002/biot.201600147

Jobling, M. A., & Gill, P. (2004). Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics*, *5*(10), 739-751. https://doi.org/10.1038/nrg1455

Johansson, B. P., Shannon, O., & Björck, L. (2008). IdeS: a bacterial proteolytic enzyme with therapeutic potential. *PLoS One*, *3*(2), e1692. https://doi.org/10.1371/journal.pone.0001692

Johnson, E. A., & Echavarri-Erasun, C. (2011). Chapter 3 - Yeast Biotechnology. In C. P. Kurtzman, J. W. Fell, & T. Boekhout (Eds.), *The Yeasts (Fifth Edition)* (pp. 21-44). Elsevier. https://doi.org/10.1016/B978-0-444-52149-1.00003-3

Jolly, L. (2001). The commercial viability of fuel ethanol from sugar cane. *Sugar Cane International*, 6–14.

Jones, E. W. (1991). Three proteolytic systems in the yeast *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, *266*(13), 7963-7966. https://doi.org/10.1016/S0021-9258(18)92922-4

Jourdan, A., Sangha, B., Kim, E., Nawaz, S., Malik, V., Vij, R., & Sekhsaria, S. (2020). Antibiotic hypersensitivity and adverse reactions: management and implications in clinical practice. *Allergy, Asthma & Clinical Immunology*, *16*, 6. https://doi.org/10.1186/s13223-020-0402-x

Joyce, C. M., & Steitz, T. A. (1994). Function and structure relationships in DNA polymerases. *Annual Review of Biochemistry*, *63*, 777-822. https://doi.org/10.1146/annurev.bi.63.070194.004021

Joyce, C. M., & Steitz, T. A. (1995). Polymerase structures and function: variations on a theme? *Journal of Bacteriology*, *177*(22), 6321-6329. https://doi.org/10.1128/jb.177.22.6321-6329.1995

Kahne, D., & Still, W. C. (1988). Hydrolysis of a peptide bond in neutral water. *Journal of the American Chemical Society*, *110*(22), 7529-7534. https://doi.org/10.1021/ja00230a041

Kallazhi, A., Rahman, A., Römling, U., & Jonas, K. (2025). Direct and indirect pathways linking the Lon protease to motility behaviors in the pathogen *Pseudomonas aeruginosa*. *PLOS Pathogens*, *21*(6), e1013288. https://doi.org/10.1371/journal.ppat.1013288

Kaltenboeck, B., & Wang, C. (2005). Advances in real-time PCR: application to clinical laboratory diagnostics. *Advances in Clinical Chemistry*, *40*, 219-259. https://doi.org/10.1016/s0065-2423(05)40006-2

Kapp, K., Schrempf, S., Lemberg, M., & Dobberstein, B. (2009). Post-Targeting Functions of Signal Peptides. https://www.ncbi.nlm.nih.gov/books/NBK6322/

Karbalaei, M., Rezaee, S. A., & Farsiani, H. (2020). *Pichia pastoris*: A highly successful expression system for optimal synthesis of heterologous proteins. *Journal of Cellular Physiology*, *235*(9), 5867-5881. https://doi.org/10.1002/jcp.29583

Karlsson, A. O., Holmlund, G., Egeland, T., & Mostad, P. (2007). DNA-testing for immigration cases: The risk of erroneous conclusions. *Forensic Science International*, *172*(2), 144-149. https://doi.org/10.1016/j.forsciint.2006.12.015

Kastberg, L. B., Ard, R., Jensen, M. K., & Workman, C. T. (2022). Burden Imposed by Heterologous Protein Production in Two Major Industrial Yeast Cell Factories: Identifying Sources and Mitigation Strategies. *Frontiers in Fungal Biology*, *3*, 827704. https://doi.org/10.3389/ffunb.2022.827704

Kaunitz, J. D. (2015). The Discovery of PCR: ProCuRement of Divine Power. *Digestive Diseases and Sciences*, *60*(8), 2230-2231. https://doi.org/10.1007/s10620-015-3747-0

Kebschull, J. M., & Zador, A. M. (2015). Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Research*, *43*(21), e143. https://doi.org/10.1093/nar/gkv717

Kellogg, D. E., Rybalkin, I., Chen, S., Mukhamedova, N., Vlasik, T., Siebert, P. D., & Chenchik, A. (1994). *Taq*Start Antibody: "hot start" PCR facilitated by a neutralizing monoclonal antibody directed against *Taq* DNA polymerase. *Biotechniques*, *16*(6), 1134-1137.

Kim, H., Yoo, S. J., & Kang, H. A. (2015). Yeast synthetic biology for the production of recombinant therapeutic proteins. *FEMS Yeast Research*, *15*(1), 1-16. https://doi.org/10.1111/1567-1364.12195

Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W., & Vogelstein, B. (2011). Detection and quantification of rare mutations with massively parallel sequencing. *Proceedings of the National Academy of Sciences*, *108*(23), 9530-9535. https://doi.org/10.1073/pnas.1105422108

Kirk, O., Borchert, T. V., & Fuglsang, C. C. (2002). Industrial enzyme applications. *Current Opinion in Biotechnology*, *13*(4), 345-351. https://doi.org/10.1016/S0958-1669(02)00328-2

Klenow, H., & Henningsen, I. (1970). Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *Escherichia coli* B by limited proteolysis. *Proceedings of the National Academy of Sciences of the United States of America*, *65*(1), 168-175. https://doi.org/10.1073/pnas.65.1.168

Kohli, U., Nigam, P., Singh, D., & Chaudhary, K. (2001). Thermostable, alkalophilic and cellulase free xylanase production by *Thermoactinomyces thalophilus* subgroup C. *Enzyme and Microbial Technology*, *28*(7), 606-610. https://doi.org/10.1016/S0141-0229(01)00320-9

Komai, T., Ishikawa, Y., Yagi, R., Suzuki-Sunagawa, H., Nishigaki, T., & Handa, H. (1997). Development of HIV-1 protease expression methods using the T7 phage promoter system. *Applied Microbiology and Biotechnology*, *47*(3), 241-245. https://doi.org/10.1007/s002530050920

Kozak, M. (1984). Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature*, *308*(5956), 241-246. https://doi.org/10.1038/308241a0

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research*, *15*(20), 8125-8148. https://doi.org/10.1093/nar/15.20.8125

Krezel, A., Mylonas, M., Kopera, E., & Bal, W. (2006). Sequence-specific Ni(II)-dependent peptide bond hydrolysis in a peptide containing threonine and histidine residues. *Acta Biochimica Polonica*, *53*(4), 721-727. https://doi.org/10.18388/abp.2006_3299

Kühne, W. (1877). Über das Verhalten verschiedener organisirter und sog. ungeformter Fermente. In *Verhandlungen des Naturhistorisch-medizinischen Vereines zu Heidelberg* (Vol. 1, pp. 190).

Kwon, K., Hasseman, J., Latham, S., Grose, C., Do, Y., Fleischmann, R. D., Pieper, R., & Peterson, S. N. (2011). Recombinant expression and functional analysis of proteases from *Streptococcus pneumoniae, Bacillus anthracis*, and *Yersinia pestis*. *BMC Biochemistry*, *12*(1), 17. https://doi.org/10.1186/1471-2091-12-17

Laatri, S., El Khayari, S., & Qriouet, Z. (2024). Exploring the molecular aspect and updating evolutionary approaches to the DNA polymerase enzymes for biotechnological needs: A comprehensive review. *International Journal of Biological Macromolecules*, *276*, 133924. https://doi.org/10.1016/j.ijbiomac.2024.133924

Lane, M. D., & Seelig, B. (2016). Highly efficient recombinant production and purification of streptococcal cysteine protease streptopain with increased enzymatic activity. *Protein Expression and Purification*, *121*, 66-72. https://doi.org/10.1016/j.pep.2016.01.002

Lewis, M. (2005). The lac repressor. *Comptes Rendus Biologies*, *328*(6), 521-548. https://doi.org/10.1016/j.crvi.2005.04.004

Li, A., Parsania, C., Tan, K., Todd, R. B., & Wong, K. H. (2021). Co-option of an extracellular protease for transcriptional control of nutrient degradation in the fungus *Aspergillus nidulans*. *Communications Biology*, *4*(1), 1409. https://doi.org/10.1038/s42003-021-02925-1

Li, J., Liang, Q., Song, W., & Marchisio, M. A. (2017). Nucleotides upstream of the Kozak sequence strongly influence gene expression in the yeast *S. cerevisiae*. *Journal of Biological Engineering*, *11*(1), 25. https://doi.org/10.1186/s13036-017-0068-1

Li, J., Wang, S., VanDusen, W. J., Schultz, L. D., George, H. A., Herber, W. K., Chae, H. J., Bentley, W. E., & Rao, G. (2000). Green fluorescent protein in *Saccharomyces cerevisiae*: Real-time studies of the GAL1 promoter. *Biotechnology and Bioengineering*, *70*(2), 187-196. https://doi.org/10.1002/1097-0290(20001020)70:2<187::AID-BIT8>3.0.CO;2-H

Liu, B., Zhang, J., Li, B., Liao, X., Du, G., & Chen, J. (2013). Expression and characterization of extreme alkaline, oxidation-resistant keratinase from *Bacillus licheniformis* in recombinant *Bacillus subtilis* WB600 expression system and its application in wool fiber processing. *World Journal of Microbiology and Biotechnology*, *29*(5), 825-832. https://doi.org/10.1007/s11274-012-1237-5

Liu, X., & Kokare, C. (2017). Chapter 11 - Microbial Enzymes of Use in Industry. In G. Brahmachari (Ed.), *Biotechnology of Microbial Enzymes* (pp. 267-298). Academic Press. https://doi.org/10.1016/B978-0-12-803725-6.00011-X

Liu, Z. Y., & Yu, X. Z. (2025). Engineering *Bacillus subtilis* for high-value bioproduction: recent advances and applications. *Microbial Cell Factories*, *24*(1), 182. https://doi.org/10.1186/s12934-025-02818-6

Loh, C. A., Shields, D. A., Schwing, A., & Evrony, G. D. (2024). High-fidelity, large-scale targeted profiling of microsatellites. *Genome Research*, *34*(7), 1008-1026. https://doi.org/10.1101/gr.278785.123

Lohr, D., Venkov, P., & Zlatanova, J. (1995). Transcriptional regulation in the yeast GAL gene family: a complex genetic network. *The FASEB Journal*, *9*(9), 777-787. https://doi.org/https://doi.org/10.1096/fasebj.9.9.7601342

Lonze, B. E., Tatapudi, V. S., Weldon, E. P., Min, E. S., Ali, N. M., Deterville, C. L., Gelb, B. E., Benstein, J. A., Dagher, N. N., Wu, M., & Montgomery, R. A. (2018). IdeS (Imlifidase): A Novel Agent That Cleaves Human IgG and Permits Successful Kidney Transplantation Across High-strength Donor-specific Antibody. *Annals of Surgery*, *268*(3), 488-496. https://doi.org/10.1097/sla.0000000000002924

López-Otín, C., & Bond, J. S. (2008). Proteases: multifunctional enzymes in life and disease. *Journal of Biological Chemistry*, *283*(45), 30433-30437. https://doi.org/10.1074/jbc.R800035200

Lopez, P. J., Guillerez, J., Sousa, R., & Dreyfus, M. (1998). On the mechanism of inhibition of phage T7 RNA polymerase by lac repressor. *Journal of Molecular Biology*, *276*(5), 861-875. https://doi.org/10.1006/jmbi.1997.1576

Lorenz, T. C. (2012). Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *Journal of Visualized Experiments.* (63), e3998. https://doi.org/10.3791/3998

Lovett, S. T. (2007). Polymerase Switching in DNA Replication. *Molecular Cell*, *27*(4), 523-526. https://doi.org/10.1016/j.molcel.2007.08.003

Lozano Terol, G., Gallego-Jara, J., Sola Martínez, R. A., Martínez Vivancos, A., Cánovas Díaz, M., & de Diego Puente, T. (2021). Impact of the Expression System on Recombinant Protein Production in *Escherichia coli* BL21. *Frontiers in Microbiology*, *12*, 682001. https://doi.org/10.3389/fmicb.2021.682001

Lu, C., & Erickson, H. P. (1997). Expression in *Escherichia coli* of the thermostable DNA polymerase from *Pyrococcus furiosus*. *Protein Expression and Purification*, *11*(2), 179-184. https://doi.org/10.1006/prep.1997.0775

Lundberg, K. S., Shoemaker, D. D., Adams, M. W. W., Short, J. M., Sorge, J. A., & Mathur, E. J. (1991). High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene*, *108*(1), 1-6. https://doi.org/10.1016/0378-1119(91)90480-Y

Lyons, B., Kwan, A. H., & Truscott, R. J. W. (2016). Spontaneous cleavage of proteins at serine and threonine is facilitated by zinc. *Aging Cell*, *15*(2), 237-244. https://doi.org/10.1111/acel.12428

Maeda, H., & Yamamoto, T. (1996). Pathogenic mechanisms induced by microbial proteases in microbial infections. *Biological Chemistry Hoppe-Seyler*, *377*(4), 217-226. https://doi.org/10.1515/bchm3.1996.377.4.217

Mahmoud, S. A., & Chien, P. (2018). Regulated Proteolysis in Bacteria. *Annual Review of Biochemistry*, *87*(Volume 87, 2018), 677-696. https://doi.org/10.1146/annurev-biochem-062917-012848

Makowski, Ł., Trojanowski, D., Till, R., Lambert, C., Lowry, R., Sockett, R. E., & Zakrzewska-Czerwińska, J. (2019). Dynamics of Chromosome Replication and Its Relationship to Predatory Attack Lifestyles in *Bdellovibrio bacteriovorus*. *Applied and Environmental Microbiology*, *85*(14), e00730-00719. https://doi.org/10.1128/AEM.00730-19

Mao, F., Leung, W.-Y., & Xin, X. (2007). Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC Biotechnology*, *7*(1), 76. https://doi.org/10.1186/1472-6750-7-76

Martín, P., de Simón, L. F., Luque, G., Farfán, M. J., & Alonso, A. (2014). Improving DNA data exchange: Validation studies on a single 6 dye STR kit with 24 loci. *Forensic Science International: Genetics*, *13*, 68-78. https://doi.org/10.1016/j.fsigen.2014.07.002

Martinez, J. L., & Baquero, F. (2000). Mutation frequencies and antibiotic resistance. *Antimicrobial Agents and Chemotherapy*, *44*(7), 1771-1777. https://doi.org/10.1128/aac.44.7.1771-1777.2000

Martinusen, S. G., Nelson, S. E., Slaton, E. W., Long, L. F., Pho, R., Ajayebi, S., & Denard, C. A. (2025). Protease engineering: Approaches, tools, and emerging trends. *Biotechnology Advances*, *82*, 108602. https://doi.org/10.1016/j.biotechadv.2025.108602

McDonald, C., Taylor, D., & Linacre, A. (2024). PCR in Forensic Science: A Critical Review. *Genes*, *15*(4), 438. https://doi.org/10.3390/genes15040438

Meldgaard, M., & Morling, N. (1997). Detection and quantitative characterization of artificial extra peaks following polymerase chain reaction amplification of 14 short tandem repeat systems used in forensic investigations. *Electrophoresis*, *18*(11), 1928-1935. https://doi.org/10.1002/elps.1150181107

Mikkelsen, M. D., Buron, L. D., Salomonsen, B., Olsen, C. E., Hansen, B. G., Mortensen, U. H., & Halkier, B. A. (2012). Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metabolic Engineering*, *14*(2), 104-111. https://doi.org/10.1016/j.ymben.2012.01.006

Mingeot-Leclercq, M.-P., & Tulkens, P. M. (1999). Aminoglycosides: Nephrotoxicity. *Antimicrobial Agents and Chemotherapy*, *43*(5), 1003-1012. https://doi.org/10.1128/aac.43.5.1003

Mitidieri, S., Souza Martinelli, A. H., Schrank, A., & Vainstein, M. H. (2006). Enzymatic detergent formulation containing amylase from *Aspergillus niger*: A comparative study with commercial detergent formulations. *Bioresource Technology*, *97*(10), 1217-1224. https://doi.org/10.1016/j.biortech.2005.05.022

Morihara, K., & Tsuzuki, H. (1975). Specificity of Proteinase K from *Tritirachium album* Limber for Synthetic Peptides. *Agricultural and Biological Chemistry*, *39*(7), 1489-1492. https://doi.org/10.1271/bbb1961.39.1489

Mullis, K. B. (1990). The unusual origin of the polymerase chain reaction. *Scientific American*, *262*(4), 56-61, 64-55. https://doi.org/10.1038/scientificamerican0490-56

Mullis, K. B., & Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In *Methods in Enzymology* (Vol. 155, pp. 335-350). Academic Press. https://doi.org/10.1016/0076-6879(87)55023-6

Muszewska, A., Stepniewska-Dziubinska, M. M., Steczkiewicz, K., Pawlowska, J., Dziedzic, A., & Ginalski, K. (2017). Fungal lifestyle reflected in serine protease repertoire. *Scientific Reports*, *7*(1), 9147. https://doi.org/10.1038/s41598-017-09644-w

Muteeb, G., Rehman, M. T., Shahwan, M., & Aatif, M. (2023). Origin of Antibiotics and Antibiotic Resistance, and Their Impacts on Drug Development: A Narrative Review. *Pharmaceuticals*, *16*(11). https://doi.org/10.3390/ph16111615

Myburgh, M. W., Schwerdtfeger, K. S., Cripwell, R. A., van Zyl, W. H., & Viljoen-Bloom, M. (2023). CHAPTER ONE - Promoters and introns as key drivers for enhanced gene expression in *Saccharomyces cerevisiae*. In G. M. Gadd & S. Sariaslani (Eds.), *Advances in Applied Microbiology*. (Vol. 124, pp. 1-29). Academic Press. https://doi.org/10.1016/bs.aambs.2023.07.002

Narayanan, M., Ali, S. S., & El-Sheekh, M. (2023). A comprehensive review on the potential of microbial enzymes in multipollutant bioremediation: Mechanisms, challenges, and future prospects. *Journal of Environmental Management*, *334*, 117532. https://doi.org/10.1016/j.jenvman.2023.117532

Navarro-López, B., Baeta, M., Granizo-Rodríguez, E., Moreno-López, O., Kleinbielen, T., Ferragut, J. F., Picornell, A., & de Pancorbo, M. M. (2022). Development and validation of a new multiplex for upgrading Y-STRs population databases from 12 to 23 markers and its forensic casework application. *Scientific Reports*, *12*(1), 21734. https://doi.org/10.1038/s41598-022-25785-z

Navarro, E., Serrano-Heras, G., Castaño, M. J., & Solera, J. (2015). Real-time PCR detection chemistry. *Clinica Chimica Acta*, *439*, 231-250. https://doi.org/10.1016/j.cca.2014.10.017

Nelson, D. C., Garbe, J., & Collin, M. (2011). Cysteine proteinase SpeB from *Streptococcus pyogenes* – a potent modifier of immunologically important host and bacterial proteins. *Biological Chemistry*, *392*(12), 1077-1088. https://doi.org/10.1515/BC.2011.208

Nielsen, T. L., Holmberg, S., & Petersen, J. G. L. (1990). Regulated overproduction and secretion of yeast carboxypeptidase Y. *Applied Microbiology and Biotechnology*, *33*(3), 307-312. https://doi.org/10.1007/BF00164527

Nigam, P., & Singh, D. (1995). Enzyme and microbial systems involved in starch processing. *Enzyme and Microbial Technology*, *17*(9), 770-778. https://doi.org/10.1016/0141-0229(94)00003-A

Nigam, P. S. (2013). Microbial Enzymes with Special Characteristics for Biotechnological Applications. *Biomolecules*, *3*(3), 597-611. https://doi.org/10.3390/biom3030597

O'Donnell, M., Langston, L., & Stillman, B. (2013). Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harbor Perspectives in Biology*, *5*(7). https://doi.org/10.1101/cshperspect.a010108

Oda, K. (2011). New families of carboxyl peptidases: serine-carboxyl peptidases and glutamic peptidases. *Journal of Biochemistry*, *151*(1), 13-25. https://doi.org/10.1093/jb/mvr129

Offei, B., Braun-Galleani, S., Venkatesh, A., Casey, W. T., O'Connor, K. E., Byrne, K. P., & Wolfe, K. H. (2022). Identification of genetic variants of the industrial yeast *Komagataella phaffii* (*Pichia pastoris*) that contribute to increased yields of secreted heterologous proteins. *PLoS Biology*, *20*(12), e3001877. https://doi.org/10.1371/journal.pbio.3001877

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., & Sugino, A. (1968). Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proceedings of the National Academy of Sciences of the United States of America*, *59*(2), 598-605. https://doi.org/10.1073/pnas.59.2.598

Olsen, J. V., Ong, S.-E., & Mann, M. (2004). Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Molecular & Cellular Proteomics*, *3*(6), 608-614. https://doi.org/10.1074/mcp.T400003-MCP200

Olszewski, M., Śpibida, M., Bilek, M., & Krawczyk, B. (2017). Fusion of *Taq* DNA polymerase with single-stranded DNA binding-like protein of *Nanoarchaeum equitans*-Expression and characterization. *PLoS One*, *12*(9), e0184162. https://doi.org/10.1371/journal.pone.0184162

Omnus, D. J., Fink, M. J., Kallazhi, A., Xandri Zaragoza, M., Leppert, A., Landreh, M., & Jonas, K. (2023). The heat shock protein LarA activates the Lon protease in response to proteotoxic stress. *Nature Communications*, *14*(1), 7636. https://doi.org/10.1038/s41467-023-43385-x

Overton, T. W. (2014). Recombinant protein production in bacterial hosts. *Drug Discovery Today*, *19*(5), 590-601. https://doi.org/10.1016/j.drudis.2013.11.008

Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A., & Ghasemi, Y. (2018). A comprehensive review of signal peptides: Structure, roles, and applications. *European Journal of Cell Biology*, *97*(6), 422-441. https://doi.org/10.1016/j.ejcb.2018.06.003

Palomares, L. A., Estrada-Moncada, S., & Ramírez, O. T. (2004). Production of Recombinant Proteins. In P. Balbás & A. Lorence (Eds.), *Recombinant Gene Expression: Reviews and Protocols* (pp. 15-51). Humana Press. https://doi.org/10.1385/1-59259-774-2:015

Pandey, A., Nigam, P., Soccol, C. R., Soccol, V. T., Singh, D., & Mohan, R. (2000). Advances in microbial amylases. *Biotechnology and applied biochemistry*, *31*(2), 135-152. https://doi.org/10.1042/ba19990073

Panja, A. S., Maiti, S., & Bandyopadhyay, B. (2020). Protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges. *Scientific Reports*, *10*(1), 1822. https://doi.org/10.1038/s41598-020-58825-7

Paul, N., Shum, J., & Le, T. (2010). Hot Start PCR. In N. King (Ed.), *RT-PCR Protocols: Second Edition* (pp. 301-318). Humana Press. https://doi.org/10.1007/978-1-60761-629-0_19

Pedroza Matute, S., & Iyavoo, S. (2025). Implementation of NGS and SNP microarrays in routine forensic practice: opportunities and barriers. *BMC Genomics*, *26*(1), 541. https://doi.org/10.1186/s12864-025-11723-6

Pelt-Verkuil, E., Belkum, A., & Hays, J. P. (2008). *Principles and Technical Aspects of PCR Amplification* (1 ed.). Springer Dordrecht. https://doi.org/10.1007/978-1-4020-6241-4

Persson, H., Vindebro, R., & Pawel-Rammingen, U. v. (2013). The Streptococcal cysteine protease SpeB is not a natural immunoglobulin-cleaving enzyme. *Infection and Immunity*, *81*(6), 2236-2241. https://doi.org/10.1128/IAI.00168-13

Petsch, D., Deckwer, W. D., & Anspach, F. B. (1998). Proteinase K Digestion of Proteins Improves Detection of Bacterial Endotoxins by the *Limulus* Amebocyte Lysate Assay: Application for Endotoxin Removal from Cationic Proteins. *Analytical Biochemistry*, *259*(1), 42-47. https://doi.org/10.1006/abio.1998.2655

Pienaar, E., Theron, M., Nelson, M., & Viljoen, H. J. (2006). A quantitative model of error accumulation during PCR amplification. *Computational Biology and Chemistry*, *30*(2), 102-111. https://doi.org/10.1016/j.compbiolchem.2005.11.002

Polizeli, M. L. T. M., Rizzatti, A. C. S., Monti, R., Terenzi, H. F., Jorge, J. A., & Amorim, D. S. (2005). Xylanases from fungi: properties and industrial applications. *Applied Microbiology and Biotechnology*, *67*(5), 577-591. https://doi.org/10.1007/s00253-005-1904-7

Pontrelli, S., Chiu, T.-Y., Lan, E. I., Chen, F. Y. H., Chang, P., & Liao, J. C. (2018). *Escherichia coli* as a host for metabolic engineering. *Metabolic Engineering*, *50*, 16-46. https://doi.org/10.1016/j.ymben.2018.04.008

Porath, J. (1992). Immobilized metal ion affinity chromatography. *Protein Expression and Purification*, *3*(4), 263-281. https://doi.org/10.1016/1046-5928(92)90001-D

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

Pronk, J. T. (2002). Auxotrophic yeast strains in fundamental and applied research. *Applied and Environmental Microbiology*, *68*(5), 2095-2100. https://doi.org/10.1128/AEM.68.5.2095-2100.2002

Radzicka, A., & Wolfenden, R. (1996). Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *Journal of the American Chemical Society*, *118*(26), 6105-6109. https://doi.org/10.1021/ja954077c

Rahban, M., Zolghadri, S., Salehi, N., Ahmad, F., Haertlé, T., Rezaei-Ghaleh, N., Sawyer, L., & Saboury, A. A. (2022). Thermal stability enhancement: Fundamental concepts of protein engineering strategies to manipulate the flexible structure. *International Journal of Biological Macromolecules*, *214*, 642-654. https://doi.org/10.1016/j.ijbiomac.2022.06.154

Rainha, J., Rodrigues, J. L., & Rodrigues, L. R. (2021). CRISPR-Cas9: A Powerful Tool to Efficiently Engineer *Saccharomyces cerevisiae*. *Life*, *11*(1), 13. https://doi.org/10.3390/life11010013

Rao, M. B., Tanksale, A. M., Ghatge, M. S., & Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteases. *Microbiology and Molecular Biology Reviews*, *62*(3), 597-635. https://doi.org/10.1128/mmbr.62.3.597-635.1998

Rao, N. S., Ermann Lundberg, L., Tomasson, J., Tullberg, C., Brink, D. P., Palmkron, S. B., van Niel, E. W. J., Håkansson, S., & Carlquist, M. (2023). Non-inhibitory levels of oxygen during cultivation increase freeze-drying stress tolerance in *Limosilactobacillus reuter*i DSM 17938. *Frontiers in Microbiology*, *Volume 14 - 2023*. https://doi.org/10.3389/fmicb.2023.1152389

Raran-Kurussi, S., Cherry, S., Zhang, D., & Waugh, D. S. (2017). Removal of Affinity Tags with TEV Protease. *Methods in Molecular Biology*, *1586*, 221-230. https://doi.org/10.1007/978-1-4939-6887-9_14

Rawat, A., Roy, M., Jyoti, A., Kaushik, S., Verma, K., & Srivastava, V. K. (2021). Cysteine proteases: Battling pathogenic parasitic protozoans with omnipresent enzymes. *Microbiological Research*, *249*, 126784. https://doi.org/10.1016/j.micres.2021.126784

Rawlings, N. D., & Barrett, A. J. (2013a). Chapter 1 - Introduction: Aspartic and Glutamic Peptidases and Their Clans. In N. D. Rawlings & G. Salvesen (Eds.), *Handbook of Proteolytic Enzymes (Third Edition)* (pp. 3-19). Academic Press. https://doi.org/10.1016/B978-0-12-382219-2.00001-6

Rawlings, N. D., & Barrett, A. J. (2013b). Chapter 77 - Introduction: Metallopeptidases and Their Clans. In N. D. Rawlings & G. Salvesen (Eds.), *Handbook of Proteolytic Enzymes (Third Edition)* (pp. 325-370). Academic Press. https://doi.org/10.1016/B978-0-12-382219-2.00077-6

Rawlings, N. D., Barrett, A. J., & Bateman, A. (2011). Asparagine peptide lyases: a seventh catalytic type of proteolytic enzymes. *Journal of Biological Chemistry*, *286*(44), 38321-38328. https://doi.org/10.1074/jbc.M111.260026

Rawlings, N. D., Waller, M., Barrett, A. J., & Bateman, A. (2014). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, *42*(Database issue), D503-509. https://doi.org/10.1093/nar/gkt953

Raz, O., Biezuner, T., Spiro, A., Amir, S., Milo, L., Titelman, A., Onn, A., Chapal-Ilani, N., Tao, L., Marx, T., Feige, U., & Shapiro, E. (2019). Short tandem repeat stutter model inferred from direct measurement of in vitro stutter noise. *Nucleic Acids Research*, *47*(5), 2436-2445. https://doi.org/10.1093/nar/gky1318

Ren, J.-W., Zhang, J.-P., Mei, Z.-L., Shao, J.-Y., Xu, G.-Q., Li, H., Gong, J.-S., Zhang, X.-M., Shi, J.-S., Zhang, X.-j., & Xu, Z.-h. (2025). Regulatory significance of terminator: A systematic approach for dissecting terminator-mediated enhancement of upstream mRNA stability. *Synthetic and Systems Biotechnology*, *10*(1), 326-335. https://doi.org/10.1016/j.synbio.2024.11.006

Rendulic, S., Jagtap, P., Rosinus, A., Eppinger, M., Baar, C., Lanz, C., Keller, H., Lambert, C., Evans, K. J., Goesmann, A., Meyer, F., Sockett, R. E., & Schuster, S. C. (2004). A Predator Unmasked: Life Cycle of *Bdellovibrio bacteriovorus* from a Genomic Perspective. *Science*, *303*(5658), 689-692. https://doi.org/10.1126/science.1093027

Rittié, L., & Perbal, B. (2008). Enzymes used in molecular biology: a useful guide. *Journal of Cell Communication and Signaling*, *2*(1-2), 25-45. https://doi.org/10.1007/s12079-008-0026-2

Roach, D., & Gehrke, C. W. (1970). The hydrolysis of proteins. *Journal of Chromatography A*, *52*, 393-404. https://doi.org/10.1016/S0021-9673(01)96589-6

Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2007). REBASE—enzymes and genes for DNA restriction and modification. *Nucleic Acids Research*, *35*(suppl_1), D269-D270. https://doi.org/10.1093/nar/gkl891

Roberts, R. J., Vincze, T., Posfai, J., & Macelis, D. (2009). REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Research*, *38*(suppl_1), D234-D236. https://doi.org/10.1093/nar/gkp874

Robinson, P. K. (2015). Enzymes: principles and biotechnological applications. *Essays in Biochemistry*, *59*, 1-41. https://doi.org/10.1042/bse0590001

Rogers, L. D., & Overall, C. M. (2013). Proteolytic post-translational modification of proteins: proteomic tools and methodology. *Molecular & Cellular Proteomics*, *12*(12), 3532-3542. https://doi.org/10.1074/mcp.M113.031310

Romei, M. G., Leonard, B., Kim, I., Kim, H. S., & Lazar, G. A. (2023). Antibody-guided proteases enable selective and catalytic degradation of challenging therapeutic targets. *Journal of Biological Chemistry*, *299*(5), 104685. https://doi.org/10.1016/j.jbc.2023.104685

Rosano, G. L., Morales, E. S., & Ceccarelli, E. A. (2019). New tools for recombinant protein production in *Escherichia coli*: A 5-year update. *Protein Science*, *28*(8), 1412-1422. https://doi.org/10.1002/pro.3668

Ryan, M. H., Petrone, D., Nemeth, J. F., Barnathan, E., Björck, L., & Jordan, R. E. (2008). Proteolysis of purified IgGs by human and bacterial enzymes in vitro and the detection of specific proteolytic fragments of endogenous IgG in rheumatoid synovial fluid. *Molecular Immunology*, *45*(7), 1837-1846. https://doi.org/10.1016/j.molimm.2007.10.043

Saenger, W. (2013). Chapter 714 - Proteinase K. In N. D. Rawlings & G. Salvesen (Eds.), *Handbook of Proteolytic Enzymes (Third Edition)* (pp. 3240-3242). Academic Press. https://doi.org/10.1016/B978-0-12-382219-2.00714-6

Saha, B., Bothast, R., & Jordan, D. (2009). Enzymes, industrial. *Encyclopedia of microbiology*, 281-294.

Sanchez, S., & Demain, A. L. (2017). Chapter 1 - Useful Microbial Enzymes—An Introduction. In G. Brahmachari (Ed.), *Biotechnology of Microbial Enzymes* (pp. 1-11). Academic Press. https://doi.org/10.1016/B978-0-12-803725-6.00001-7

Schenk, J., Balazs, K., Jungo, C., Urfer, J., Wegmann, C., Zocchi, A., Marison, I., & Stockar, U. (2008). Influence of specific productivity and glycosylation of a recombinant avidin produced by a *Pichia pastoris* Mut+ strain. *Biotechnology and Bioengineering*, *99*, 368-377. https://doi.org/10.1002/bit.21565

Schillinger, J., Koci, M., Bravo-Rodriguez, K., Heilmann, G., Kaschani, F., Kaiser, M., Beuck, C., Luecke, H., Huber, R., Hellerschmied, D., Burston, S. G., & Ehrmann, M. (2024). High resolution analysis of proteolytic substrate processing. *Journal of Biological Chemistry*, *300*(11), 107812. https://doi.org/10.1016/j.jbc.2024.107812

Schroeder, G. K., Lad, C., Wyman, P., Williams, N. H., & Wolfenden, R. (2006). The time required for water attack at the phosphorus atom of simple phosphodiesters and of DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(11), 4052-4055. https://doi.org/10.1073/pnas.0510879103

Schütz, A., Bernhard, F., Berrow, N., Buyel, J. F., Ferreira-da-Silva, F., Haustraete, J., van den Heuvel, J., Hoffmann, J. E., de Marco, A., Peleg, Y., Suppmann, S., Unger, T., Vanhoucke, M., Witt, S., & Remans, K. (2023). A concise guide to choosing suitable gene expression systems for recombinant protein production. *STAR Protocols*, *4*(4), 102572. https://doi.org/10.1016/j.xpro.2023.102572

Sellick, C. A., Campbell, R. N., & Reece, R. J. (2008). Chapter 3 Galactose Metabolism in Yeast—Structure and Regulation of the Leloir Pathway Enzymes and the Genes Encoding Them. In *International Review of Cell and Molecular Biology* (Vol. 269, pp. 111-150). Academic Press. https://doi.org/10.1016/S1937-6448(08)01003-4

Shakoori, A. R. (2017). Fluorescence *In Situ* Hybridization (FISH) and Its Applications. In T. A. Bhat & A. A. Wani (Eds.), *Chromosome Structure and Aberrations* (pp. 343-367). Springer India. https://doi.org/10.1007/978-81-322-3673-3_16

Shapiro, H. M. (2005). *Practical flow cytometry*. John Wiley & Sons.

Shi, Y., Niu, Y., Zhang, P., Luo, H., Liu, S., Zhang, S., Wang, J., Li, Y., Liu, X., Song, T., Xu, T., & He, S. (2023). Characterization of genome-wide STR variation in 6487 human genomes. *Nature Communications*, *14*(1), 2092. https://doi.org/10.1038/s41467-023-37690-8

Sibirny, A. A. (2016). Yeast peroxisomes: structure, functions and biotechnological opportunities. *FEMS Yeast Research*, *16*(4). https://doi.org/10.1093/femsyr/fow038

Sibley, C. D., Peirano, G., & Church, D. L. (2012). Molecular methods for pathogen and microbial community detection and characterization: current and potential application in diagnostic microbiology. *Infection, Genetics and Evolution*, *12*(3), 505-521. https://doi.org/10.1016/j.meegid.2012.01.011

Sidstedt, M., Gynnå, A. H., Kiesler, K. M., Jansson, L., Steffen, C. R., Håkansson, J., Johansson, G., Österlund, T., Bogestål, Y., Tillmar, A., Rådström, P., Ståhlberg, A., Vallone, P. M., & Hedman, J. (2024). Ultrasensitive sequencing of STR markers utilizing unique molecular identifiers and the SiMSen-Seq method. *Forensic Science International: Genetics*, *71*, 103047. https://doi.org/10.1016/j.fsigen.2024.103047

Sidstedt, M., Rådström, P., & Hedman, J. (2020). PCR inhibition in qPCR, dPCR and MPS-mechanisms and solutions. *Analytical and Bioanalytical Chemistry*, *412*(9), 2009-2023. https://doi.org/10.1007/s00216-020-02490-2

Sidstedt, M., Steffen, C. R., Kiesler, K. M., Vallone, P. M., Rådström, P., & Hedman, J. (2019). The impact of common PCR inhibitors on forensic MPS analysis. *Forensic Science International: Genetics*, *40*, 182-191. https://doi.org/10.1016/j.fsigen.2019.03.001

Simas, R. G., Pessoa Junior, A., & Long, P. F. (2023). Mechanistic aspects of IPTG (isopropylthio-β-galactoside) transport across the cytoplasmic membrane of *Escherichia coli*-a rate limiting step in the induction of recombinant protein expression. *Journal of Industrial Microbiology and Biotechnology*, *50*(1). https://doi.org/10.1093/jimb/kuad034

Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., & Panda, A. K. (2015). Protein recovery from inclusion bodies of *Escherichia coli* using mild solubilization process. *Microbial Cell Factories*, *14*(1), 41. https://doi.org/10.1186/s12934-015-0222-8

Sjögren, J., Andersson, L., Mejàre, M., & Olsson, F. (2017). Generating and purifying Fab fragments from human and mouse IgG using the bacterial enzymes IdeS, SpeB and Kgp. In P. Nordenfelt & M. Collin (Eds.), *Bacterial Pathogenesis: Methods and Protocols* (pp. 319-329). Springer New York. https://doi.org/10.1007/978-1-4939-6673-8_21

Sloan, D. B., Broz, A. K., Sharbrough, J., & Wu, Z. (2018). Detecting Rare Mutations and DNA Damage with Sequencing-Based Methods. *Trends in Biotechnology*, *36*(7), 729-740. https://doi.org/10.1016/j.tibtech.2018.02.009

Sockett, R. E. (2009). Predatory Lifestyle of Bdellovibrio bacteriovorus. *Annual Review of Microbiology*, *63*, 523-539. https://doi.org/10.1146/annurev.micro.091208.073346

Sonenberg, N., & Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, *136*(4), 731-745. https://doi.org/10.1016/j.cell.2009.01.042

Song, P., Zhang, X., Wang, S., Xu, W., Wang, F., Fu, R., & Wei, F. (2023). Microbial proteases and their applications [Review]. *Frontiers in Microbiology*, *Volume 14 - 2023*. https://doi.org/10.3389/fmicb.2023.1236368

Sørensen, H. P., & Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microbial Cell Factories*, *4*(1), 1. https://doi.org/10.1186/1475-2859-4-1

Sparkes, R., Kimpton, C., Watson, S., Oldroyd, N., Clayton, T., Barnett, L., Arnold, J., Thompson, C., Hale, R., Chapman, J., Urquhart, A., & Gill, P. (1996). The validation of a 7-locus multiplex STR test for use in forensic casework. (I). Mixtures, ageing, degradation and species studies. *International Journal of Legal Medicine*, *109*(4), 186-194. https://doi.org/10.1007/bf01225517

Spohner, S. C., Müller, H., Quitmann, H., & Czermak, P. (2015). Expression of enzymes for the usage in food and feed industry with *Pichia pastoris*. *Journal of Biotechnology*, *202*, 118-134. https://doi.org/10.1016/j.jbiotec.2015.01.027

Staadig, A., Hedman, J., & Tillmar, A. (2023). Applying Unique Molecular Indices with an Extensive All-in-One Forensic SNP Panel for Improved Genotype Accuracy and Sensitivity. *Genes*, *14*(4), 818. https://doi.org/10.3390/genes14040818

Staadig, A., & Tillmar, A. (2019). An overall limited effect on the weight-of-evidence when taking STR DNA sequence polymorphism into account in kinship analysis. *Forensic Science International: Genetics*, *39*, 44-49. https://doi.org/10.1016/j.fsigen.2018.11.020

Ståhlberg, A., Krzyzanowski, P. M., Jackson, J. B., Egyud, M., Stein, L., & Godfrey, T. E. (2016). Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing. *Nucleic Acids Research*, *44*(11), e105-e105. https://doi.org/10.1093/nar/gkw224

Stanbury, P. F., Whitaker, A., & Hall, S. J. (2017). Chapter 12 - The production of heterologous proteins. In P. F. Stanbury, A. Whitaker, & S. J. Hall (Eds.), *Principles of Fermentation Technology (Third Edition)* (pp. 725-775). Butterworth-Heinemann. https://doi.org/10.1016/B978-0-08-099953-1.00012-0

Steimann, T., Heite, Z., Germer, A., Blank, L. M., Büchs, J., Mann, M., & Magnus, J. B. (2024). Understanding exopolysaccharide byproduct formation in *Komagataella phaffii* fermentation processes for recombinant protein production. *Microbial Cell Factories*, *23*(1), 131. https://doi.org/10.1186/s12934-024-02403-3

Steitz, T. A. (1998). A mechanism for all polymerases. *Nature*, *391*(6664), 231-232. https://doi.org/10.1038/34542

Steitz, T. A. (1999). DNA Polymerases: Structural Diversity and Common Mechanisms. *Journal of Biological Chemistry*, *274*(25), 17395-17398. https://doi.org/10.1074/jbc.274.25.17395

Stephens, K. M., Barta, R., Fleming, K., Perez, J. C., Wu, S.-F., Snedecor, J., Holt, C. L., LaRue, B., & Budowle, B. (2023). Developmental validation of the ForenSeq MainstAY kit, MiSeq FGx sequencing system and ForenSeq Universal Analysis Software. *Forensic Science International: Genetics*, *64*, 102851. https://doi.org/10.1016/j.fsigen.2023.102851

Stockwell, S. R., Landry, C. R., & Rifkin, S. A. (2015). The yeast galactose network as a quantitative model for cellular memory. *Molecular BioSystems*, *11*(1), 28-37. https://doi.org/10.1039/c4mb00448e

Stockwell, S. R., & Rifkin, S. A. (2017). A living vector field reveals constraints on galactose network induction in yeast. *Molecular Systems Biology*, *13*(1), 908. https://doi.org/10.15252/msb.20167323

Stovicek, V., Holkenbrink, C., & Borodina, I. (2017). CRISPR/Cas system for yeast genome engineering: advances and applications. *FEMS Yeast Research*, *17*(5). https://doi.org/10.1093/femsyr/fox030

Studier, F. W. (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of Molecular Biology*, *219*(1), 37-44. https://doi.org/10.1016/0022-2836(91)90855-Z

Studier, F. W., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, *189*(1), 113-130. https://doi.org/10.1016/0022-2836(86)90385-2

Su, Y., Liu, C., Fang, H., & Zhang, D. (2020). *Bacillus subtilis*: a universal cell factory for industry, agriculture, biomaterials and medicine. *Microbial Cell Factories*, *19*(1), 173. https://doi.org/10.1186/s12934-020-01436-8

Subramani, S. (1992). Targeting of proteins into the peroxisomal matrix. *The Journal of Membrane Biology*, *125*(2), 99-106. https://doi.org/10.1007/BF00233350

Sudol, A. S. L., Butler, J., Ivory, D. P., Tews, I., & Crispin, M. (2022). Extensive substrate recognition by the Streptococcal antibody-degrading enzymes IdeS and EndoS. *Nature Communications*, *13*(1), 7801. https://doi.org/10.1038/s41467-022-35340-z

Sukharev, S. A., Pleshakova, O. V., & Sadovnikov, V. B. (1997). Role of proteases in activation of apoptosis. *Cell Death & Differentiation*, *4*(6), 457-462. https://doi.org/10.1038/sj.cdd.4400263

Szecsi, P. (1992). The aspartic proteases. *Scandinavian journal of clinical and laboratory investigation. Supplementum*, *210*, 5-22. https://doi.org/10.3109/00365519209104650

Tabor, S., & Richardson, C. C. (1985). A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(4), 1074-1078. https://doi.org/10.1073/pnas.82.4.1074

Teichert, U., Mechler, B., Müller, H., & Wolf, D. H. (1989). Lysosomal (Vacuolar) Proteinases of Yeast Are Essential Catalysts for Protein Degradation, Differentiation, and Cell Survival. *Journal of Biological Chemistry*, *264*(27), 16037-16045. https://doi.org/10.1016/S0021-9258(18)71584-6

The Royal Swedish Academy of Sciences. (1993). *The Nobel Prize in Chemistry 1993 - Press Release* https://www.nobelprize.org/prizes/chemistry/1993/press-release/. Accessed 2025-10-01

Tindall, K. R., & Kunkel, T. A. (1988). Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry*, *27*(16), 6008-6013. https://doi.org/10.1021/bi00416a027

Tominaga, M., Shima, Y., Nozaki, K., Ito, Y., Someda, M., Shoya, Y., Hashii, N., Obata, C., Matsumoto-Kitano, M., Suematsu, K., Matsukawa, T., Hosoya, K., Hashiba, N., Kondo, A., & Ishii, J. (2024). Designing strong inducible synthetic promoters in yeasts. *Nature Communications*, *15*(1), 10653. https://doi.org/10.1038/s41467-024-54865-z

Tomschy, A., Tessier, M., Wyss, M., Brugger, R., Broger, C., Schnoebelen, L., Van Loon, A. P. G. M., & Pasamontes, L. (2000). Optimization of the catalytic properties of *Aspergillus fumigatus* phytase based on the three-dimensional structure. *Protein Science*, *9*(7), 1304-1311. https://doi.org/10.1110/ps.9.7.1304

Travis, J., & Potempa, J. (2000). Bacterial proteinases as targets for the development of second-generation antibiotics. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, *1477*(1), 35-50. https://doi.org/10.1016/S0167-4838(99)00278-2

Trono, D. (2019). Chapter 13 - Recombinant Enzymes in the Food and Pharmaceutical Industries. In R. S. Singh, R. R. Singhania, A. Pandey, & C. Larroche (Eds.), *Advances in Enzyme Technology* (pp. 349-387). Elsevier. https://doi.org/10.1016/B978-0-444-64114-4.00013-3

Turk, B. (2006). Targeting proteases: successes, failures and future prospects. *Nature Reviews Drug Discovery*, *5*(9), 785-799. https://doi.org/10.1038/nrd2092

Turk, B., Turk, D., & Turk, V. (2012). Protease signalling: the cutting edge. *Embo Journal*, *31*(7), 1630-1643. https://doi.org/10.1038/emboj.2012.42

Uchii, K., Doi, H., Okahashi, T., Katano, I., Yamanaka, H., Sakata, M. K., & Minamoto, T. (2019). Comparison of inhibition resistance among PCR reagents for detection and quantification of environmental DNA. *Environmental DNA*, *1*(4), 359-367. https://doi.org/10.1002/edn3.37

van den Hazel, H. B., Kielland-Brandt, M. C., & Winther, J. R. (1996). Review: Biosynthesis and function of yeast vacuolar proteases. *Yeast*, *12*(1), 1-16. https://doi.org/10.1002/(SICI)1097-0061(199601)12:1<1::AID-YEA902>3.0.CO;2-N

van Rijswijck, D. M. H., Bondt, A., de Kat, N., Lood, R., & Heck, A. J. R. (2024). Direct comparison of the hinge-cleaving proteases IgdE and BdpK for LC-MS-based IgG1 clonal profiling. *Analytical Chemistry*, *96*(1), 23-27. https://doi.org/10.1021/acs.analchem.3c03712

Vargas, B., dos Santos, J., Pereira, G., & de Mello, F. (2023). An atlas of rational genetic engineering strategies for improved xylose metabolism in *Saccharomyces cerevisiae*. *PeerJ*, *11*, e16340. https://doi.org/10.7717/peerj.16340

Verma, S., Dixit, R., & Pandey, K. C. (2016). Cysteine Proteases: Modes of Activation and Future Prospects as Pharmacological Targets. *Frontiers in Pharmacology*, *Volume 7 - 2016*. https://doi.org/10.3389/fphar.2016.00107

Viguera, E., Canceill, D., & Ehrlich, S. D. (2001). Replication slippage involves DNA polymerase pausing and dissociation. *The EMBO Journal*, *20*(10), 2587-2595. https://doi.org/10.1093/emboj/20.10.2587

Vijayakumar, V. E., & Venkataraman, K. (2024). A Systematic Review of the Potential of *Pichia pastoris* (*Komagataella phaffii*) as an Alternative Host for Biologics Production. *Molecular Biotechnology*, *66*(7), 1621-1639. https://doi.org/10.1007/s12033-023-00803-1

Vincents, B., von Pawel-Rammingen, U., Björck, L., & Abrahamson, M. (2004). Enzymatic characterization of the Streptococcal endopeptidase, IdeS, reveals that it is a cysteine protease with strict specificity for IgG cleavage due to exosite binding. *Biochemistry*, *43*(49), 15540-15549. https://doi.org/10.1021/bi048284d

Vlasak, J., & Ionescu, R. (2011). Fragmentation of monoclonal antibodies. *MAbs*, *3*(3), 253-263. https://doi.org/10.4161/mabs.3.3.15608

Vollenbroich, V., Meyer, J., Engels, R., Cardinali, G., Menezes, R. A., & Hollenberg, C. P. (1999). Galactose induction in yeast involves association of Gal80p with Gal1p or Gal3p. *Molecular and General Genetics MGG*, *261*(3), 495-507. https://doi.org/10.1007/s004380050993

von Pawel-Rammingen, U., Johansson, B. P., & Björck, L. (2002). IdeS, a novel Streptococcal cysteine proteinase with unique specificity for immunoglobulin G. *EMBO Journal*, *21*(7), 1607-1615. https://doi.org/10.1093/emboj/21.7.1607

Wall, D. (2009). Recombinant DNA, Basic Procedures. In M. Schaechter (Ed.), *Encyclopedia of Microbiology (Third Edition)* (pp. 271-280). Academic Press. https://doi.org/10.1016/B978-012373944-5.00031-6

Walsh, P. S., Fildes, N. J., & Reynolds, R. (1996). Sequence Analysis and Characterization of Stutter Products at the Tetranucleotide Repeat Locus VWA. *Nucleic Acids Research*, *24*(14), 2807-2812. https://doi.org/10.1093/nar/24.14.2807

Wang, A. C., & Wang, I. Y. (1977). Cleavage sites of human IgGl immunoglobulin by papain. *Immunochemistry*, *14*(3), 197-200. https://doi.org/10.1016/0019-2791(77)90194-X

Wang, G., Huang, M., & Nielsen, J. (2017). Exploring the potential of *Saccharomyces cerevisiae* for biopharmaceutical protein production. *Current Opinion in Biotechnology*, *48*, 77-84. https://doi.org/10.1016/j.copbio.2017.03.017

Wang, W., Chen, K., & Xu, C. (2006). DNA quantification using EvaGreen and a real-time PCR instrument. *Analytical Biochemistry*, *356*(2), 303-305. https://doi.org/10.1016/j.ab.2006.05.027

Wang, Y., Prosen, D. E., Mei, L., Sullivan, J. C., Finney, M., & Vander Horn, P. B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Research*, *32*(3), 1197-1207. https://doi.org/10.1093/nar/gkh271

Ward, O. P. (2011). 3.49 - Proteases. In M. Moo-Young (Ed.), *Comprehensive Biotechnology (Second Edition)* (pp. 571-582). Academic Press. https://doi.org/10.1016/B978-0-08-088504-9.00222-1

Weber, J. L., & Wong, C. (1993). Mutation of human short tandem repeats. *Human Molecular Genetics*, *2*(8), 1123-1128. https://doi.org/10.1093/hmg/2.8.1123

Wei, L., Wang, Z., Zhang, G., & Ye, B. (2017). Characterization of Terminators in *Saccharomyces cerevisiae* and an Exploration of Factors Affecting Their Strength. *ChemBioChem*, *18*(24), 2422-2427. https://doi.org/10.1002/cbic.201700516

Weis, B. L., Schleiff, E., & Zerges, W. (2013). Protein targeting to subcellular organelles via mRNA localization. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1833*(2), 260-273. https://doi.org/10.1016/j.bbamcr.2012.04.004

Wenig, K., Chatwell, L., von Pawel-Rammingen, U., Björck, L., Huber, R., & Sondermann, P. (2004). Structure of the Streptococcal endopeptidase IdeS, a cysteine proteinase with strict specificity for IgG. *Proceedings of the National Academy of Sciences*, *101*(50), 17371-17376. https://doi.org/10.1073/pnas.0407965101

Wetmore, D. R., Wong, S.-L., & Roche, R. S. (1992). The role of the pro-sequence in the processing and secretion of the thermolysin-like neutral protease from *Bacillus cereus*. *Molecular Microbiology*, *6*(12), 1593-1604. https://doi.org/10.1111/j.1365-2958.1992.tb00884.x

Willey, J. M., Sherwood, L. M., & Woolverton, C. J. (2019). *Prescott's Microbiology* (11 ed.). McGraw-Hill Education.

Wintrode, P. L., Miyazaki, K., & Arnold, F. H. (2000). Cold Adaptation of a Mesophilic Subtilisin-like Protease by Laboratory Evolution. *Journal of Biological Chemistry*, *275*(41), 31635-31640. https://doi.org/10.1074/jbc.M004503200

Witte, A. K., Sickha, R., Mester, P., Fister, S., Schoder, D., & Rossmanith, P. (2018). Essential role of polymerases for assay performance – Impact of polymerase replacement in a well-established assay. *Biomolecular Detection and Quantification*, *16*, 12-20. https://doi.org/10.1016/j.bdq.2018.10.002

Woerner, A. E., Mandape, S., King, J. L., Muenzler, M., Crysup, B., & Budowle, B. (2021). Reducing noise and stutter in short tandem repeat loci with unique molecular identifiers. *Forensic Science International: Genetics*, *51*, 102459. https://doi.org/10.1016/j.fsigen.2020.102459

Woessmann, J., Petrosius, V., Üresin, N., Kotol, D., Aragon-Fernandez, P., Hober, A., auf dem Keller, U., Edfors, F., & Schoof, E. M. (2023). Assessing the Role of Trypsin in Quantitative Plasma and Single-Cell Proteomics toward Clinical Application. *Analytical Chemistry*, *95*(36), 13649-13658. https://doi.org/10.1021/acs.analchem.3c02543

Xie, J., Zhuang, Z., Gou, S., Zhang, Q., Wang, X., Lan, T., Lian, M., Li, N., Liang, Y., Ouyang, Z., Ye, Y., Wu, H., Lai, L., & Wang, K. (2023). Precise genome editing of the Kozak sequence enables bidirectional and quantitative modulation of protein translation to anticipated levels without affecting transcription. *Nucleic Acids Research*, *51*(18), 10075-10093. https://doi.org/10.1093/nar/gkad687

Xie, Y., Han, X., & Miao, Y. (2018). An effective recombinant protein expression and purification system in *Saccharomyces cerevisiae*. *Current Protocols in Molecular Biology*, *123*(1), e62. https://doi.org/10.1002/cpmb.62

Yamanoi, E., Sakurada, M., & Ueno, Y. (2021). Low stutter ratio by SuperFi polymerase. *Forensic Science International: Reports*, *3*, 100201. https://doi.org/10.1016/j.fsir.2021.100201

Yang, N., Matthew, M. A., & Yao, C. (2023). Roles of Cysteine Proteases in Biology and Pathogenesis of Parasites. *Microorganisms*, *11*(6). https://doi.org/10.3390/microorganisms11061397

Yin, M.-Q., Xu, K., Luan, T., Kang, X.-L., Yang, X.-Y., Li, H.-X., Hou, Y.-H., Zhao, J.-Z., & Bao, X.-M. (2024). Metabolic engineering for compartmentalized biosynthesis of the valuable compounds in *Saccharomyces cerevisiae*. *Microbiological Research*, *286*, 127815. https://doi.org/10.1016/j.micres.2024.127815

Yip, K. S. P., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R., & Rice, D. W. (1995). The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure*, *3*(11), 1147-1158. https://doi.org/10.1016/S0969-2126(01)00251-9

Yuan, J.-j., Geng, S.-l., & Wang, T.-y. (2025). Research Progress Fusion Tags for Recombinant Protein Production. *Biotechnology and applied biochemistry*, e2749. https://doi.org/10.1002/bab.2749

Zhang, Y.-Z., Zhang, W.-X., & Chen, X.-L. (2020). Mechanisms for Induction of Microbial Extracellular Proteases in Response to Exterior Proteins. *Applied and Environmental Microbiology*, *86*(19), e01036-01020. https://doi.org/10.1128/AEM.01036-20

Zhao, H., Li, H., Jia, Y., Wen, X., Guo, H., Xu, H., & Wang, Y. (2020). Building a Robust Chromatin Immunoprecipitation Method with Substantially Improved Efficiency. *Plant Physiology*, *183*(3), 1026-1034. https://doi.org/10.1104/pp.20.00392

Zhao, J., Fang, H., & Zhang, D. (2020). Expanding application of CRISPR-Cas9 system in microorganisms. *Synthetic and Systems Biotechnology*, *5*(4), 269-276. https://doi.org/10.1016/j.synbio.2020.08.001