Nada är en gemensam institution mellan
Kungliga Tekniska högskolan och Stockholms universitet.

Mathematical Modelling and Analysis of the Pyrosequencing Reaction System

ANNA SVANTESSON

Avhandling som med tillstånd av Kungliga Tekniska högskolan
framlägges till offentlig granskning för avläggande av teknologie licentiatexamen
torsdagen den 27 oktober 2005 kl 13.00
i sal 4523, Lindstedtsvägen 5,
Kungliga Tekniska högskolan, Stockholm.

ISBN 91-7178-177-3
TRITA-NA-0535
ISSN 0348-2952
ISRN KTH/NA/R--05/35--SE

Abstract

The Pyrosequencing^{TM} technology is a relatively new and easy-to-use technique for sequencing shorter DNA fragments. The method monitors pyrophosphate release following nucleotide incorporation into a DNA template using bioluminescence. The detection relies on the efficient cooperation of four different enzymes to monitor the DNA synthesis in real-time.

In this thesis, a new mathematical model of the Pyrosequencing reaction system is presented. The model is based on irreversible Michaelis-Menten kinetics and competitive inhibition. The modelling is aimed at bringing insights into the difficulties in the Pyrosequencing method and at increasing the understanding of the cooperation of the enzymes on a detailed level. Furthermore, the model can be used to optimise the chemistry and identify limiting factors with the ultimate goal of improving the read length of the method.

The simulation results show strong resemblance with experimental data. The Pyrosequencing model reproduces the dynamics of a light pulse resulting from single nucleotide incorporation with great accuracy, as well as the general features of a whole pyrogram^{TM}. Undesired effects, such as the plus- and minus-shifts, are successfully predicted by the model using two constant efficiency factors. The pulse broadening seen in experiments is partly explained by apyrase inhibition and dilution.

The inhibition of the nucleotide-degrading apyrase is further investigated in an experimental pre-study. The relative inhibition constants of all deoxynucleotides were measured and it was found that the analogue 2'-deoxyadenosine-5'-O-(1-thiotriphosphate) (dATPαS) and its di- and monophosphate counterparts have lower inhibition constants than the natural deoxynucleotides. Moreover, it is experimentally shown that dADPαS is much more difficult for the apyrase to degrade, and thus is responsible for the successive pulse broadening. Simulations support the theory that dADPαS has a very low catalytic constant.

An improved and biologically motivated model of the polymerase is introduced, and shows that the polymerase is an extremely efficient enzyme. However, in the presence of a nucleotide-degrading apyrase, the incorporation efficiency is, under the Pyrosequencing conditions, diminished to a level close to that of the minus-shift effect. The model also predicts that the polymerase is able to incorporate even very small amounts of nucleotides, thus offering a description of the plus-shift effect.

Keywords: apyrase, bioluminescence, DNA polymerase, DNA sequencing, enzyme kinetics, Klenow fragment, mathematical modelling, Michaelis-Menten, Pyrosequencing

ISBN 91-7178-177-3 • TRITA-NA-0535 • ISSN 0348-2952 • ISRN KTH/NA/R- -05/35- -SE

Sammanfattning

Pyrosequencing^{TM}-teknologin är en relativt ny och lättanvänd teknik för att sekvensera kortare DNA-fragment. Metoden övervakar med hjälp av bioluminiscens frigörandet av pyrofosfat, som är en följd av inbindningen av nukleotider i ett DNA-templat. Detektionen bygger på en effektiv samverkan mellan fyra olika enzymer för att övervaka syntetiseringen av DNA i realtid.

I denna avhandling presenteras en ny matematisk modell av reaktionssystemet i Pyrosequencing. Modellen baserar sig på irreversibel Michaelis-Mentenkinetik samt kompetitiv inhibering. Målet med modelleringen är att ge insikter i svårigheterna vid Pyrosequencing och att öka förståelsen av detaljerna vid samverkan mellan enzymerna. Vidare kan modellen användas för att optimera kemin och för att identifiera begränsande faktorer med målet att förbättra läslängden hos metoden.

Simuleringsresultaten uppvisar stora likheter med experimentella data. Pyrosequencing-modellen återger med stor exakthet dynamiken hos en ljuspuls, som skapas till följd av inkorporering av en nukleotid, liksom de allmäna dragen hos ett fullständigt pyrogram^{TM}. Oönskade effekter, såsom plus- och minusskift, förutsägs av modellen med hjälp av två konstanta effektivitetsfaktorer. Pulsbreddningen, som uppkommer i experiment, förklaras delvis av inhibering av apyraset och utspädning.

Inhiberingen av det nukleotidnedbrytande apyraset undersöks mer ingående i en experimentell förstudie. De relativa inhiberingskonstanterna för alla deoxynukleotider mäts. Resultatet är att analogen 2'-deoxyadenosine-5'-O-(1-thiotriphosphate) (dATPαS) samt dess di- och monofosfat motsvarigheter, har lägre inhiberingskonstant än de naturliga deoxynukleotiderna. Vidare visas det experimentellt att apyraset har mycket svårare att bryta ned dADPαS, och att denna därmed är ansvarig för den ökande pulsbreddningen. Simuleringar stödjer teorin att dADPαS har en mycket låg katalytisk konstant.

En förbättrad och biologiskt motiverad modell av polymeraset introduceras och visar att polymeraset är ett mycket effektivt enzym. I närvaro av ett nukleotidnedbrytande apyras och under Pyrosequencing-förhållanden minskar dock effektiviteten till en nivå liknande den för minusskiftseffekten. Modellen predikterar också att polymeraset dessutom kan inkorporera mycket små mängder av nukleotid, vilket därmed ger en beskrivning av plusskiftseffekten.

Nyckelord: apyras, bioluminiscens, DNA polymeras, DNA-sekvensering, enzymkinetik, Klenow-fragment, matematisk modellering, Michaelis-Menten, Pyrosequencing

ISBN 91-7178-177-3 • TRITA-NA-0535 • ISSN 0348-2952 • ISRN KTH/NA/R- -05/35- -SE

To my family

Preface

The work behind this licentiate thesis was financed by VINNOVA and the Parallel Scientific Computing Institute (PSCI) at the former Department of Numerical Analysis and Computer Science (NADA), at the Royal Institute of Technology (KTH); together with Biotage AB (former Pyrosequencing AB), Uppsala, Sweden.

Acknowledgements

First of all, I would like to express my sincere gratitude towards my supervisors Dr. Jeanette Hellgren Kotaleski (present), Prof. Anders Lansner (former) and Prof. Pål Nyrén for their scientific guidance and for involving me in this interesting project in the first place. A special thanks to Pål for proposing new ideas on where to proceed during the whole project time, and for his genuine curiosity, which inspired me.

A big thanks to all present and former people in the CBN/SANS group who have made the days richer both professionally and personally. I would especially like to thank Pål Westermark (nowadays Dr.) for help on the modelling part and for fruitful discussions on this project, as well as being a good travelling companion. Thanks also to Malin Sandström for being a good room mate and providing me with both interesting and relevant scientific news. And to Johannes Hjorth for nice conversations and collaboration on a couple of graduate courses. Thanks also to our administrator Harriett Johansson for excellent help on all practical matters.

Furthermore, I wish to thank Tommy Nordström at the Department of Biotechnology, KTH, for giving me a (relatively) painless introduction to working in the laboratory as well as interesting discussions on the Pyrosequencing method and life in general. I wish you the best of luck!

And last but not least, to my parents, Maria and Lennart, for their love, encouragement and support through the years. And to Danne for his friendship, humour and inspiration. Life wouldn't be the same without you!

Anna Svantesson

Stockholm, September 2005

Symbols and Abbreviations

ADP	adenosine diphosphate
dADPαS	alpha-thio-deoxyadenosine diphosphate
AMP	adenosine monophosphate
APS	adenosine 5'-phosphosulfate
apyr	apyrase, ATP-diphosphohydrolase
Asul	ATP sulfurylase
ATP	adenosine triphosphate
CCD	charge-coupled device
dATPαS	alpha-thio-deoxyadenosine triphosphate
ddNMP	dideoxynucleoside monophosphate
DNA	deoxyribonucleic acid
dNDP	deoxynucleoside diphosphate
dNMP	deoxynucleoside monophosphate
dNTP	deoxynucleoside triphosphate
E	enzyme or enzyme concentration
k	reaction rate
k_{cat}	catalytic constant, turnover number
K_d	dissociation constant
KF	Klenow fragment
K_i	inhibition constant
K_m	Michaelis constant
luc	luciferase
N	any one of adenosine (A), cytidine (C), guanosine (G) or thymidine (T)
NDP	nucleoside diphosphate
P_i	inorganic phosphate
PCR	polymerase chain reaction
pol	DNA polymerase
PP_i	inorganic pyrophosphate
S	substrate or substrate concentration
SNP	single nucleotide polymorphism
v	reaction rate or velocity
V_{max}	limiting rate, maximum reaction velocity

Part I.

Background

Contents

Chapter 1

INTRODUCTION

Deoxyribonucleic acid (DNA) sequencing is the procedure of determining the nucleotide order in the DNA molecule. The development of DNA sequencing techniques during the last thirty years, have made it possible to undertake ambitious projects such as the Human Genome Project (International Human Genome Sequencing Consortium, 2004), where the whole human genome was deciphered. Today, genomes of more than thousands organisms have been fully sequenced. However, DNA sequencing is so much more than sequencing whole genomes, since much information can be gained by sequencing specific target regions of the DNA. During the last fifteen years DNA sequencing has become an invaluable instrument in as varying fields as medicine, agriculture and forensic studies. For applications in these disciplines, the Pyrosequencing^{TM} technology has emerged as a strong alternative to the more conventional DNA sequencing techniques. It is a new and easy-to-use enzymatic method originally developed at the Royal Institute of Technology, Stockholm, Sweden, and the first automated systems were introduced on the market in the year 2000.

The aim of this thesis has been to explore the chemistry of the Pyrosequencing method using mathematical modelling. The motivation behind the work has been to get an increased understanding of the reaction system and the cooperativity of the participating enzymes, as well as to find explanations of various effects observed in sequencing experiments.

1.1. Why Use Mathematical Modelling?

Reality constitutes a set of complex systems. Nevertheless or perhaps because of that, we humans are eager to unravel the secrets or our surroundings and we continuously strive to find explanations of phenomena in our environment. When reality becomes too complicated to make exhaustive observations, a model of the system of interest can be very useful. Experimental recordings that would

otherwise be laborious, time-consuming and expensive, can instead be performed in model simulations, making only a minimal amount of experiments necessary. Furthermore, a model will give an indication of which experiments to perform and on the whole facilitates the analysis of a system. A model can also predict unforeseen events as well as be a means of controlling a complex technical process.

The modelling process is not a uni-directional route starting at point one and finishing with a definite model. Instead it should be looked upon as something dynamic, something that will likely take turns and iteratively find its way towards a model with as many desired properties as possible. The road to a pleasing model always begins with making observations, and based on these observations, a mental image is constructed. This image then has to be translated into something implementable and the language for doing this is mathematics. Important variables and constants have to be identified, and their relations have to be formulated into equations. Once a first model has been constructed, it can be used to make predictions. However, these predictions have to be thoroughly validated and tested against earlier observations and if necessary new experiments will follow. By this point, the model probably has to be revised, either by a small modification or by using a completely different approach. Repeated prediction/validation/measurement steps will refine the model and eventually lead to a satisfying result.

Mathematical modelling has successfully been applied to a wide range of biological systems. Applications include mathematical modelling in the drug discovery process, where for instance physiologically based pharmacokinetic (PBPK) models are used to simulate and predict drug absorption (Theil *et al.*, 2003). Other examples are mathematical modelling of signalling networks on different levels of the nervous system from subcellular (Bhalla, 2003) to network level (De Schutter *et al.*, 2005). Whole organs like the human heart (Noble, 2002) as well as the spread of infectious diseases (Matthews and Woolhouse, 2005) have also been studied through modelling. With this thesis, a mathematical model of the Pyrosequencing reaction system can be added to that list.

1.2. Contributions

- A mathematical model of the Pyrosequencing reaction system has been developed. The model captures the dynamics of a light pulse, following single nucleotide incorporation, with great accuracy.
- The proposed model can also replicate the overall characteristics of a complete sequencing output, using two constant efficiency factors for the plus- and minus-shift effects.
- A short review of the Pyrosequencing technology is given in the thesis. Factors affecting the real length of the method are highlighted and the causes are discussed on the basis of the mathematical model.
- An experimental pre-study investigating the causes for the inhibition of the apyrase enzyme with pulse broadening as a result.

- A model of the experimental setup used in the apyrase inhibition study, which offers support for determined relations in the obtained kinetic parameters.
- An improved model of the polymerase step including the processivity property of the polymerase.
- The polymerase model suggests an explanation of the minus-shift effect, as well as an indication of the origin of the plus-shift.

1.3. Articles

The contents of this licentiate thesis are based on the following papers, which are found in the second part of the thesis. They are referred to in the text by their Roman numbers.

Paper I. Anna Svantesson, Pål O. Westermark, Jeanette Hellgren Kotaleski, Baback Gharizadeh, Anders Lansner, Pål Nyrén, A Mathematical Model of the Pyrosequencing Reaction System, *Biophys. Chem.* 110:129-145, 2004

Paper II. Anna Svantesson, Tommy Nordström, Jeanette Hellgren Kotaleski, Pål Nyrén, An Investigation of the Apyrase Step in the Pyrosequencing Model: Combining Experiments and Simulations, TRITA-NA-P0509, Royal Institute of Technology, School of Computer Science and Communication, 2005

Tommy Nordström and Pål Nyrén contributed to the design of the experiments.

Paper III. Anna Svantesson, Pål O. Westermark, Pål Nyrén, Jeanette Hellgren Kotaleski, Towards an Improved Polymerase Description in the Pyrosequencing Model, TRITA-NA-P0510, Royal Institute of Technology, School of Computer Science and Communication, 2005

1.4. Structure of the Thesis

The organisation of this thesis is as follows: In the next section (Chapter 2), a short review over the Pyrosequencing technology is presented. Understanding details in the chemistry of the Pyrosequencing reaction system has been the main motivation for the work leading up to this thesis. In Chapter 3, an introduction to enzyme kinetics, the mathematical tool for modelling chemical reactions, is given. This is followed by a description of the models used to model the steps in the Pyrosequencing system, and a summary of some of the more important results presented in **Papers I – III**. The thesis ends with concluding remarks and a discussion of future prospects for the further development and utilisation of the Pyrosequencing model.

Chapter 2

THE PYROSEQUENCING TECHNOLOGY

The discovery of the structure of the DNA molecule (Watson and Crick, 1953) has led to a dramatic development in biotechnology. DNA sequencing, i.e. the process of determining the exact sequence of the bases that constitute a DNA strand, is today an important and widely used tool in most areas of biochemistry. The first two methods for sequencing DNA were introduced in 1977 and include the Sanger method (Sanger *et al.*, 1977), also known as the dideoxy method, and the Maxam-Gilbert method (Maxam and Gilbert, 1977). These methods are based on electrophoretic and autoradiographic recordings of the results. The Pyrosequencing method (Nyrén, 2001; Ronaghi *et al.*, 1998), which is the object of interest of this thesis, is a rather new DNA sequencing technique that is based on a so-called sequencing-by-synthesis scheme. In this method, nucleotide incorporation into DNA is monitored in real-time using bioluminescence.

In this chapter an introduction to the Pyrosequencing technology is given. First an overview is presented, followed by a more detailed description of the enzymes and reactions involved. Then, a few words is said about the design of the Pyrosequencing instrument. Ever since the introduction of the method several chemical improvements have been suggested and the most important are reviewed here. The chapter ends with some examples of applications of the method and the main remaining problems, some of which are the focus of the thesis.

2.1. Enzymes in Pyrosequencing

In 1987, Pål Nyrén described a method for monitoring DNA polymerase activity (Nyrén, 1987) using bioluminescence. This was followed by a description of a completely new way of sequencing DNA (Hyman, 1988) involving the first three enzymes used in what was later to become the Pyrosequencing method. However, the suggested method utilised a series of connected capillary columns,

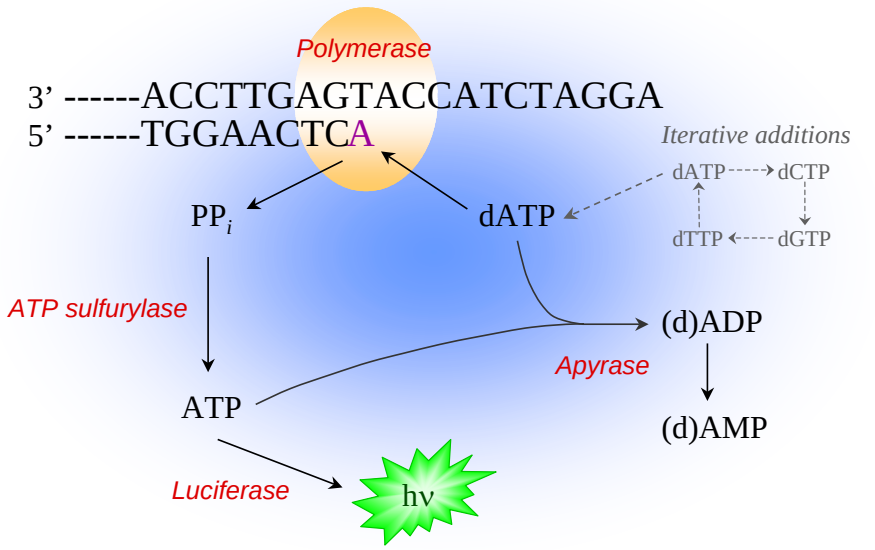

Figure 2.1. *The principle of Pyrosequencing. To a mixture of essentially four enzymes and DNA, nucleotides dNTP are added. The DNA polymerase, which is bound to the primer-template pair, catalyses the incorporation of the added nucleotide (here dATP), which produces PP_i. The PP_i is transformed into ATP by the ATP sulfurylase. The luciferase detects the ATP in a reaction where light is formed. The apyrase removes remaining ATP and unincorporated nucletides. The procedure is then repeated with the next nucleotide in order.*

which was cumbersome and rather impractical. The solution came with the introduction of a fourth enzyme, the nucleotide-degrading apyrase (Nyrén, 1994, 2001). The process time was much shortened and the system could be made highly automated. The method of Pyrosequencing was born.

An overview of the Pyrosequencing principle is given in Figure 2.1. The reaction system consists of the four enzymes DNA polymerase, ATP sulfurylase, luciferase and apyrase. Single-stranded and primed DNA fragments are mixed with necessary substrates and the four enzymes. Then, deoxynucleoside triphosphates, dATP, dCTP, dGTP or dTTP (collectively known as dNTP), are added one at a time in an iterative manner. In cases where the added nucleotide matches the DNA template, the nucleotide is incorporated by the polymerase. In this reaction, PPi is released, which in turn is converted by the ATP sulfurylase into adenosine triphosphate (ATP). The produced ATP drives a light-generating reaction catalysed by the luciferase. The amount of light (peak height) is proportional to the number of incorporated nucleotides. If an added nucleotide does not match, no PPi is formed and consequently no light is produced. The apyrase removes unincorporated nucleotides and remaining ATP between the nucleotide

additions. As the process proceeds, the complementary DNA strand grows and the nucleotide sequence is determined from the generated light pulses, which collectively form a so-called pyrogram™. Next, a more detailed description of the enzymes and the reactions is given.

2.1.1. Polymerase

DNA polymerases belong to a group of enzymes that is present in all living cells (Joyce and Steitz, 1994; Albà, 2001). They are catalysts for DNA polymerisation in the important process of replication. There are different types of DNA polymerases, but all possess 5' to 3' DNA synthesis capabilities. They also have some kind of 5' to 3' exonuclease activity and some polymerases exhibit a 3' to 5' exonuclease activity as well. The different types have different tasks in the cell, some synthesise new DNA and perform proofreading, whereas others are involved in DNA repair.

The DNA polymerase used in Pyrosequencing is the large subdomain of the *E. coli* DNA polymerase I (E.C. 2.7.7.7). This domain is known as the Klenow Fragment (KF) (Klenow *et al.*, 1971) and in Pyrosequencing a modified variant is used, excluding the enzyme's 3' to 5' exonuclease activity (exo$^-$) (Derbyshire *et al.*, 1988). The polymerase binds to the 3'-end of the DNA primer and the following synthesis proceeds in the 5' to 3' direction. This is achieved by incorporating a deoxynucleoside triphosphate (dNTP) onto the primer end. As a product, inorganic pyrophosphate PPi is released.

$$\text{DNA}_n + \text{dNTP} \longrightarrow \text{DNA}_{n+1} + \text{PP}_i$$

Here N stands for the four nucleosides: adenosine (A), guanosine (G), cytidine (C) and thymidine (T), and n denotes the length of the primer DNA. Polymerases are highly efficient enzymes and in the presence of matching nucleotides, the enzyme switches to processive synthesis. This signifies that the polymerase extends several nucleotides in a row before it dissociates from the DNA molecule. The DNA polymerase I has a reported processivity of approximately 20 (Bambara *et al.*, 1978) nucleotides, but there are polymerases that extend thousands of nucleotides (Tabor *et al.*, 1987) under the right conditions.

2.1.2. ATP Sulfurylase

The second enzyme in the Pyrosequencing system is the ATP sulfurylase (E.C. 2.7.7.4), which catalyses the reaction between pyrophosphate and adenosine 5' phosphosulfate (APS), thereby producing ATP. The reaction can be summarised as:

$$\text{PP}_i + \text{APS} \longrightarrow \text{ATP} + \text{SO}_4^{2-}$$

In vivo, the enzyme drives the reaction in the reverse direction even though that is kinetically unfavourable. This is achieved with the help of additional enzymes – the APS kinase that continuously removes the produced APS and the pyrophosphatase that hydrolysis PPi. This enzymatic system is thus responsible

for the indirect management of sulphur-containing compounds in the cell (Segel *et al.*, 1987). The ATP sulfurylase used in Pyrosequencing is taken from baker's yeast, *Saccharomyces cerevisiae* (Ullrich *et al.*, 2001), but the enzyme has also been found in many other types of organisms, such as spinach leafs (Li *et al.*, 1995) and rats (Brandan and Hirschberg, 1988).

2.1.3. Luciferase

The luciferases (E.C. 1.13.12.7) constitute a group of enzymes that all have in common that they can produce light in bioluminescence. They are found in three main groups of light generating organisms, namely fireflies (*Lampyridae*), railroad worms (*Phengodidae*) and click beetles (*Elateridae*) (Viviani, 2002). The firefly luciferases have been the most intensively studied and in particular the luciferase from the North American *Photinus pyralis* firefly (DeLuca, 1976). The kinetics of the firefly luciferase has been well established (DeLuca and McElroy, 1974; Brovko *et al.*, 1994) and the reaction scheme can be summarised in the following two reaction steps

$$\text{luciferase} + \text{D-luciferin} + \text{ATP} \longrightarrow \text{luciferase} \cdot \text{luciferin} \cdot \text{AMP} + \text{PP}_i$$

$$\text{luciferase} \cdot \text{luciferin} \cdot \text{AMP} + \text{O}_2 \longrightarrow \text{luciferase} + \text{oxyluciferin} + \text{AMP} + \text{CO}_2 + h\nu$$

In the first step, the luciferase activates the substrate D-luciferin in the presence of Mg^{2+}. In the second step, the enzyme catalyses the formation of an energy-rich intermediate, which spontaneously decomposes, thereby emitting a photon of visible light. The light produced is in the yellow–green spectrum with $\lambda = 550 - 590$ nm (Sala-Newby and Campbell, 1994), and the activity of the enzyme can be monitored with extreme sensitivity. In the automated Pyrosequencing system, a charged-coupled device (CCD) camera is used for the light detection.

The luciferase can also produce light from dATP (Lee *et al.*, 1970). As a consequence, in Pyrosequencing, the natural dATP has been replaced by the 2'-deoxyadenosine-5'-O-(1-thiotriphosphate) (dATPαS) analogue, which is not a substrate to luciferase (Ronaghi *et al.*, 1996).

2.1.4. Apyrase

Apyrase (E.C. 3.6.1.5) is a nucleotide degrading enzyme that can hydrolyse triphosphates and diphosphates (Komoszynski and Wojtczak, 1996)

$$\text{dNTP} \longrightarrow \text{dNDP} + \text{P}_i$$

$$\text{dNDP} \longrightarrow \text{dNMP} + \text{P}_i$$

Commercially available apyrases have been purified from potato tubers *Solanum tuberosum*, where the occurrence of the enzyme is rich. This enzyme exists in several isoforms depending on the clonal variety. The two most common and

well-studied are the Pimpernel type (or apyrase A) and the Desirée type (or apyrase B) (Traverso-Cori *et al.*, 1970), where the former is the apyrase currently used in Pyrosequencing. The two varieties exhibit different kinetic properties (Kettlun *et al.*, 2005). The most distinguishing is the relative rates of the hydrolysis of ATP and ADP. This is the so-called ATPase/ADPase ratio, which is 10 for the Pimpernel type and 1 for the Desirée type.

The apyrase has a relatively low substrate specificity, which makes it very suitable for use in the Pyrosequencing reaction system. Unincorporated nucleotides, as well as remaining ATP, are all degraded by the apyrase, making a subsequent nucleotide addition possible without the need of a separate washing phase.

2.2. Performing Pyrosequencing

2.2.1. DNA Preparation Steps

Before the actual sequencing can begin, there are several steps of DNA preparation. Although variations of the preparation process exist, the main steps are

- **Isolation of DNA.** The DNA to be sequenced has to be extracted from the cell and purified in order to get rid of potential inhibitors and other substances that can disturb the reactions. A high degree of purification is essential for the succeeding steps.
- **DNA Amplification.** In order to obtain sufficient amounts of the isolated DNA region, the DNA has to be amplified. The most extensively used method for copying DNA fragments is the Polymerase Chain Reaction (PCR) (Mullis *et al.*, 1986).
- **Washing.** All excess amounts of PCR primers, nucleotides and PPi must be removed prior to the sequencing since these residues would otherwise interfere with the Pyrosequencing reactions.
- **Addition of Sequencing Primer.** The polymerase requires a 3'-end to begin the synthesis and a sequencing primer is therefore annealed to the DNA fragment.

2.2.2. The Pyrosequencing Instrument

The Pyrosequencing system is today marketed by the company Biotage AB, Uppsala, Sweden (`www.biotage.se`). A couple of different instruments are available together with an optimised kit of the required reagents. In the experimental part of this thesis, the PSQ^{TM} 96 instrument was used and the principle of this machine is described here, although other instruments have similar design.

The reaction mixture containing the prepared DNA is placed on a 96-well plate. Nucleotides, enzymes and additional substrates are loaded into a cartridge that is inserted into a dispenser. The sequencing process is initialised by the automatic dispensation of the enzymes and substrates into all wells. Nucleotides are then dispensed iteratively every 60 or 65 s. An integrated mixer

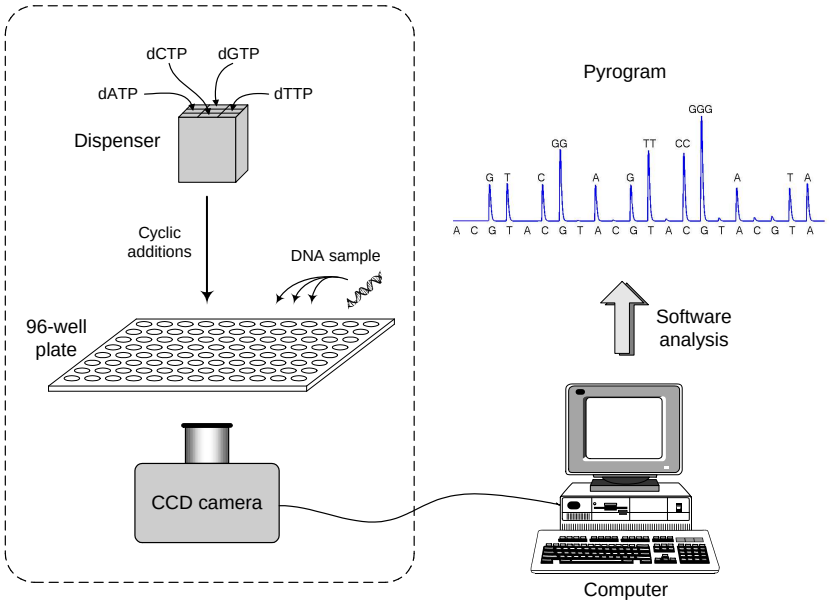

Figure 2.2. *The principle of the PSQ 96 instrument. The DNA sample is placed on a 96-well plate. The four nucleotides are added iteratively and the light response is detected by a CCD camera. A computer software deciphers the DNA sequencing from the resulting pyrogram.*

table ensures sufficient mixing of the reaction substrates during the whole process.

The overall principle of the Pyrosequencing instrument is shown in Figure 2.2. The detection system constitutes a charged-coupled device (CCD) camera that is mounted beneath the reaction wells, taking light recordings once every second. An array of lenses helps focus the emitted light onto the camera. The CCD camera sends the detected signals to a computer where the data is displayed as a pyrogram. A pattern recognition algorithm resolves the pyrogram and outputs the determined sequence.

2.3. Improvements in Pyrosequencing

Since the advent of the Pyrosequencing technology, several improvements of the chemistry of the method have been suggested. The more important refinements are summarised below.

Nucleotide Analogues

The nucleotides used in Pyrosequencing are the natural deoxytriphosphates except for the A nucleotide. Due to interference with the luciferase reaction, the

dATP was replaced by the dATPαS analogue (Ronaghi *et al.*, 1996). However, the read length of the method was still limited and the incorporation was particularly inefficient at homopolymeric T-regions. The originally used dATPαS solution contained a mixture of the two isomers Sp and Rp. Experiments with the pure Sp-isomer (Gharizadeh *et al.*, 2002) revealed that the Rp-isomer is not a substrate for Klenow and that the mixture of isomers inhibits the apyrase to a larger extent than the pure Sp-isomer. Switching to pure Sp-isomer dATPαS has increased the read length, and is the currently used A nucleotide.

However, the Sp-isomer dATPαS still has an inhibiting impact on the apyrase. As an alternative, another A nucleotide analogue, 7-deaza-2'-deoxyadenosine-5'-triphosphate (c^7dATP), was tested for use in Pyrosequencing (Eriksson *et al.*, 2004b). The results showed a significant decrease in apyrase inhibition when using the c^7dATP analogue.

Stabilising Proteins

Problems related to the formation of secondary structures of the DNA are of great concern and is for many sequences probably *the* most aggravating factor. DNA self-complementation, mispriming and unspecific hybridisation all disturb the polymerase and generate false signals in the pyrogram. A decrease in unspecific hybridisation can be obtained by the addition of single-stranded DNA binding proteins (SSBs). This was implemented in Pyrosequencing (Ronaghi, 2000; Ehn *et al.*, 2002) with successful results.

The 3'-end of a single-stranded DNA can loop back onto itself. To prevent this in Pyrosequencing, the 3'-end was locked using dideoxy nucleoside monophosphates (ddNMP) or modified oligonucleotides (Utting *et al.*, 2004). This reduced the occurrences of background signals. These are just two examples of stabilising strategies in an increasing list of suggestions.

Increased Temperature

Another way to decrease the formation of secondary DNA structures is by increasing the temperature. However, this affects the activity of the enzymes. Most enzymes have a temperature optimum and for Pyrosequencing, the limiting enzyme is the luciferase that loses most of its activity at temperatures above 30°C (Ford and Leach, 1998). However, in the presence of glycine betaine the luciferase stability was assured even at higher temperatures. This enabled the performance of the Pyrosequencing reaction at 37°C (Eriksson *et al.*, 2004a, 2003), with decreased false signals due to secondary structures, which otherwise form at lower temperatures, as a result.

A Different Polymerase

The choice of the enzymes in Pyrosequencing is crucial for the performance of the method. In one study, a different polymerase, the exonuclease deficient T7 DNA polymerase, also known as Sequenase, was tested (Gharizadeh *et al.*, 2004a). The Sequenase is a more efficient polymerase than Klenow and has a

higher processivity. The sequencing results were pleasing. Using Sequenase, the method could read through longer homopolymeric T-regions, and moreover, false signals due to secondary structures were reduced.

2.4. Some Example Applications

The strength of the Pyrosequencing method lies in its ability to rapidly sequence shorter DNA fragments of lengths up to 20–50 bases, both cost-efficiently and with great accuracy. Here, some examples of different applications for the Pyrosequencing method are given.

- **Detection of Single-Nucleotide Polymorphisms.** The main application for the Pyrosequencing technology is in the analysis of Single-Nucleotide Polymorphisms (SNPs) (Alderborn *et al.*, 2000; Ahmadian *et al.*, 2000; Fakhrai-Rad *et al.*, 2002). A SNP is a sequence of the genome that differs by a single nucleotide between one portion of a population and another. Variations in the human DNA can affect how humans develop diseases and respond to drug treatment. Analysing SNPs is therefore of great importance for biomedical research such as in development of new drugs and in evolutionary relationship studies. Using the Pyrosequencing method for SNP analysis involves sequencing several bases right before and after the actual SNP, which makes the method robust.
- **Microbial Typing.** Infectious diseases are caused by different kinds of microorganisms such as bacteria, fungi and viruses, and these organisms exist in different types. In finding an appropriate treatment, it is important to rapidly identify the different species and strains, which are the source of the infection. The Pyrosequencing technology are used for fungal (Gharizadeh *et al.*, 2004b), bacterial (Ronaghi *et al.*, 1999; Hjalmarsson *et al.*, 2004) and viral (Gharizadeh *et al.*, 2001; Elahi *et al.*, 2003) typing.
- **Clone Checking.** For applications involving the manipulation of DNA, such as gene cloning and site-directed mutagenesis, it is essential to check the resulting DNA for correctness. Clone checking by Pyrosequencing has been performed by employing a pre-programmed nucleotide dispensation strategy that corresponds to the desired DNA (Nourizad *et al.*, 2003).
- **Forensic Studies.** The Pyrosequencing technology has also been applied to sequencing of mitochondrial DNA in forensic analyses (Andreasson *et al.*, 2002). Mitochondrial DNA contains highly variable regions, which are utilised in the identification process.
- **Genome Sequencing by Miniaturisation** A modified version of the Pyrosequencing chemistry has been implemented by 454 Life Sciences (`www.454.com`) to perform hundreds of thousands of reactions in parallel on a single chip. The technology is run in picolitre-scale volumes and has shown to sequence a whole bacterial genome in one day (Margulies *et al.*, 2005).

2.5. Aspects of Pyrosequencing

In this section some remaining problems and interesting aspects regarding the Pyrosequencing reaction system are presented. A majority of these aspects have been the motivations behind the development of a mathematical model of the system. Further discussion around them is given in Chapter 4.

Pulse Broadening

Examining experimental pyrograms closely, it can be seen that the light pulses are slightly wider at later dispensations compared to those produced early in the sequencing process. This is known as a pulse broadening effect and is primarily attributed to a successive inhibition of the apyrase (Gharizadeh *et al.*, 2002). This inhibition is more pronounced at A dispensation and is therefore connected to the dATPαS nucleotide and its hydrolysed counterparts.

An additional factor influencing the broadening is the effect of dilution. At each dispensation, 0.2 μl is added to the reaction mixture, which has an initial volume of 50 μl. Dilution decreases the activity of all the enzymes in general and this is modelled in **Paper I**.

The causes of the pulse broadening effect is investigated using both simulations and experimental techniques in **Paper II**.

Minus-shift Effect

The performance of the Pyrosequencing method is dependent on a highly efficient polymerase that performs synthesis with great fidelity. However, the polymerase shows a slight inefficiency during Pyrosequencing, meaning not all template DNA is elongated at each nucleotide addition. This effect is known as the minus-shift effect. The DNA that has not been elongated is substrate to a polymerising reaction the next time a matching nucleotide is dispensed. Incomplete incorporation is therefore seen as small background signals appearing after the real pulse in the pyrogram. In **Paper III** a kinetic explanation of the minus-shift is investigated.

Plus-shift Effect

There is also a plus-shift effect in Pyrosequencing experiments, where background signals appear before the real signal. This effect is attributed to an inefficiency of the apyrase, possibly due to inhibition. Unincorporated nucleotides are not completely degraded, with the result that these can function as substrates whenever matching DNA is present. This effect is highlighted in **Paper III**.

A-effect

Peak heights at A dispensations are approximately 5–15% higher than peaks appearing after the incorporation of the other natural nucleotides (Eriksson *et al.*,

2004b). This is referred to as an A-effect. The A-effect is concentration dependent and is more pronounced when higher dATPαS concentrations are used. The Klenow polymerase incorporates this modified nucleotide at a somewhat slower rate compared to the natural nucleotides. To compensate this, a higher amount of dATPαS is used since otherwise the incorporation of this nucleotide into homopolymeric T-regions is inefficient.

Impurities

Reagents are seldom completely pure, and impurities such as enzymes and substrates can have a negative impact on the sequencing procedure. PPi is a common contaminant that directly interferes with the Pyrosequencing system since it is a substrate for the ATP sulfurylase. High amounts of PPi can consume a great deal of the APS before the actual sequencing begins, which may result in premature termination of the signals in the pyrogram. Sources of contaminating PPi are insufficient washing after the PCR step and the nucleotide solutions since nucleotides naturally degrade by time. PPi can be degraded by the enzymes pyrophosphatase (PPase) or alkaline phosphatase.

Another serious contaminant is the enzyme NDP-kinase. This enzyme catalyses the reaction between ATP/dNTP and dNDP forming the products ADP/dNDP and dNTP. The dNTP can then be used by the polymerase in the DNA synthesis with unsynchronised extension as the result (Nordström, 2003; Karamohamed *et al.*, 1999). Hence, using enzymes with high purity is essential.

Chapter 3

Introduction to Enzyme Kinetics

Enzyme kinetics is the study of the rate of enzyme-catalysed chemical reactions and the impact that various factors have on it. As a foundation lies a simple mathematical framework describing the way in which the reaction velocity is altered by changes in the concentrations of the participating reactants.

In this chapter a short introduction to the subject of enzyme kinetics is given. The focus is put on Michaelis-Menten kinetics since this is the type that was used in the modelling of the Pyrosequencing process. The treatment follows the recommendations on symbols and terminology by the Nomenclature Committee of the International Union of Biochemistry (NC-IUB, 1983). For further reading on the subject; a comprehensive and well-written book is Athel Cornish-Bowden's *Fundamentals of Enzyme Kinetics* (Cornish-Bowden, 2004).

3.1. Basic Definitions

Living systems depend on chemical reactions, which, on their own, normally occur at extremely slow rates. Enzymes are proteins (with a few exceptions) that can catalyse chemical reactions. By catalysing, it is meant that the rate of a reaction is increased by lowering the activation energy that has to be gained before the reaction can occur. Enzymes themselves are never consumed in a reaction but form intermediate complexes with the participating substrates and guide the formation of the resulting products.

The kinetics of reactions is described using mathematical equations and these are usually expressed in terms of concentrations. To indicate concentration the $[\cdot]$ notation is generally used, that is to say the concentration of a reactant X is denoted by $[X]$. When there is no risk of misconception the brackets are sometimes omitted. Now, consider a reaction where a substrate S is transformed into a product P under the influence of an enzyme E. This enzyme-catalysed reaction can be written in the form

$$S \rightleftharpoons P \tag{3.1}$$

The rate of the consumption of the substrate of this reaction is equal to the derivative of the substrate concentration with respect to time, by definition. When there is a one-to-one stoichiometric ratio between the substrate and the product, like in reaction (3.1), the rate of formation of the product is identical to that of the consumption of the substrate apart from a change in sign. This must hold since nothing is allowed to leave the system. The rate v of this first-order reaction is

$$v = -\frac{d[S]}{dt} = \frac{d[P]}{dt} \tag{3.2}$$

By introducing a rate constant k_1 for the forward direction of the reaction and a corresponding backward rate constant k_{-1}, the reaction rate can also be expressed as

$$v = -k_1[S] = k_{-1}[P] \tag{3.3}$$

At equilibrium, the rate of the forward reaction is equal to the rate of the reverse reaction. This leads to an expression for the *equilibrium constant* K_{eq} of the reaction.

$$K_{eq} = \frac{[P]}{[S]} = \frac{k_1}{k_{-1}} \tag{3.4}$$

For reactions involving more than one substrate and one product, the equilibrium constant is expressed as the product of all product concentrations divided by the product of all substrate concentrations. In a more general context, it is often spoken of a *dissociation* equilibrium constant K_d. It is defined for the dissociation of a reactant from an enzyme: $EX \rightleftharpoons E + X$. Likewise, an *association* equilibrium constant can be defined for the reverse reaction, that is the association of the reactant from the enzyme.

The above treatment has only concerned so-called *elementary* reactions, which are reactions where no reaction intermediates form (i.e. intermediates have not been detected or are not relevant for the description of the chemistry). Most reactions, however, are not this simple. On the other hand, virtually all chemical reaction can be broken down into several steps of individual elementary reactions, and this is substantially used in enzyme kinetics.

3.2. Michaelis-Menten Kinetics

In 1913, Leonor Michaelis and Maud Menten described a simple and, what has since then shown to be, highly applicable model of an enzymatic reaction (Michaelis and Menten, 1913). Their analysis was a further development of the work of the chemist Victor Henri (Henri, 1903). They proposed the following reaction mechanism, including a reaction intermediate ES, for the conversion of a substrate S into a product P:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \tag{3.5}$$

where, k_1, k_{-1} and k_2 are the rate constants for each of the elementary reaction steps. The first step is reversible and the second is irreversible, making the overall scheme irreversible. Next, the famous Michaelis-Menten equation is derived using the treatment of Briggs and Haldane (Briggs and Haldane, 1925). This will provide as an example of how the mathematics in enzyme kinetics works.

3.2.1. Irreversible Reactions

The starting point for Briggs and Haldane derivation is the irreversible Michaelis-Menten mechanism (3.5). The species of interest is the enzyme-substrate complex. The rate of change of this intermediate complex can be written as

$$\frac{d[ES]}{dt} = k_1\big([E]_0 - [ES]\big) \cdot [S] - k_{-1}[ES] - k_2[ES] \tag{3.6}$$

where $[E]_0$ is the initial concentration of free enzyme. Now, the system is assumed to have reached steady-state (this is known as the Briggs-Haldane assumption), meaning the concentration of the intermediate complex no longer changes. Thus, $d[ES]/dt = 0$ and the $[ES]$ can be solved for in expression (3.6). This gives the relation

$$[ES] = \frac{k_1[E]_0[S]}{k_{-1} + k_2 + k_1[S]} \tag{3.7}$$

The rate of reaction (3.5) is always $v = d[P]/dt = k_2[ES]$ and at steady-state it can be expressed as

$$v = \frac{k_1 k_2[E]_0[S]}{k_{-1} + k_2 + k_1[S]} = \frac{k_2[E]_0[S]}{\frac{k_{-1}+k_2}{k_1} + [S]} \tag{3.8}$$

This is the Michaelis-Menten expression for the rate of a chemical reaction. Usually it is written in the following form

$$v = \frac{V_{max}[S]}{K_m + [S]} \tag{3.9}$$

where

$$V_{max} = k_2 \cdot [E]_0 \qquad \text{and} \qquad K_m = \frac{k_{-1} + k_2}{k_1} \tag{3.10}$$

The two new parameters that were introduced are the *limiting rate* V_{max} and the *Michaelis constant* K_m. Equation (3.9) is not just a convenient way of expressing the reaction rate. In fact, these parameters can be interpreted in terms of kinetic properties. In Figure 3.1, the rate has been plotted as a function of the substrate concentration. The limiting rate, also known as the maximum velocity, is indicated by a dashed horizontal line, revealing the origin of the parameter's name. V_{max} is the upper limit on the reaction rate as the substrate concentration gets very large.

Most often, the rate constant k_2 is written as k_{cat}, leaving $V_{max} = k_{cat} \cdot [E]_0$. The rate constant is the *catalytic constant* but is also referred to as the *turnover number* since it reflects the number of substrate molecules that are converted by one enzyme molecule per second. Hence, both V_{max} and k_{cat} determine the velocity of the reaction.

The other parameter in the Michaelis-Menten equation is the Michaelis constant K_m. From Figure 3.1 it is seen that it represents the substrate concentration where the rate reaches half of the limiting rate. K_m reflects the strength of the bond between the enzyme and the substrate. A low K_m value corresponds to a tight bond.

The irreversible description of a reaction is often a very good approximation and the modelling performed in this thesis deals with these kinds of rate equations. This is motivated in **Paper I**.

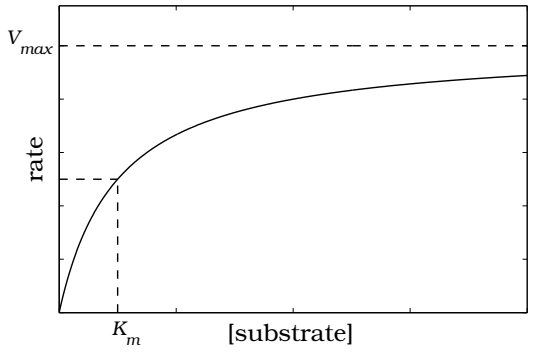

Figure 3.1. *Dependence of the initial rate on the substrate concentration for a reaction following the Michaelis-Menten equation (3.9). The Michaelis constant K_m is defined as the substrate concentration where the rate is $V_{max}/2$, and V_{max} is the maximum velocity.*

3.2.2. Reversible Reactions

Most reactions, however, are not strictly irreversible and they exhibit at least a small backward rate. The irreversible Michaelis-Menten expression is for many cases still a good approximation, but there are also numerous reactions that do not fit into that model. Some of these reactions can be modelled using the reversible Michaelis-Menten scheme instead:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} EP \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E + P \tag{3.11}$$

The mechanism has been expanded by an enzyme-product intermediate. Making the same Briggs-Haldane assumption, the reversible rate expression can instead be written as (Cornish-Bowden, 2004, section 2.7)

$$v = \frac{[E]_0}{1 + \frac{[S]}{K_{mS}} + \frac{[P]}{K_{mP}}} \left([S] \frac{k_{catS}}{K_{mS}} - [P] \frac{k_{catP}}{K_{mP}} \right) \tag{3.12}$$

Here, the K_{mS} and K_{mP} are the Michaelis- constants of the forward and reverse reactions, respectively and equivalently for the catalytic constants.

3.2.3. Reactions of More Than One Substrate

A majority of all reactions is not just a simple conversion of one substrate into one product, but involve several substrates and several products. The far most common situation is represented by the following scheme

$$A + B \rightleftharpoons P + Q \tag{3.13}$$

where two substrates A and B react to form two products P and Q. A Michaelis-Menten type of rate expression for this reaction is (Cornish-Bowden, 2004, section 7.4)

$$v = \frac{V_{max}[A][B]}{K_{iA}K_{mB} + K_{mB}[A] + K_{mA}[B] + [A][B]} \tag{3.14}$$

where product terms have been omitted, thus representing the irreversible case. The parameter K_{iA} is the inhibition constant of substrate A (see section 3.3). Adding more substrate and products is straight-forward, even though the number of terms quickly gets rather large.

3.3. Inhibition

The rate of an enzyme-catalysed reaction can be affected by the presence of other compounds. In case where the compound binds to the enzyme and thereby decreases the rate of the reaction, the compound is called an *inhibitor*. The inhibition of an enzyme is divided into two categories, reversible and irreversible.

In irreversible inhibition, the inhibitor effectively blocks the enzyme thereby decreasing the maximum number of possible enzyme-substrate complexes. Reversible inhibitors are usually characterised as belonging to one of three types, namely *competitive inhibition*, *uncompetitive inhibition* or *non-competitive inhibition*.

3.3.1. Competitive Inhibition

Competitive inhibition reflects the situation where the inhibitor has almost the same molecular structure as one of the substrates. Hence, it has the ability to bind to the same active site on the enzyme as the substrate, and the inhibitor and the substrate are competing against each other. This mechanism is important in many cellular systems and is frequently utilised in drug discovery.

Taking inhibition into account does not involve a major alteration of the mathematical analysis for the rate expression. In the presence of a competitive inhibitor I, the modified Michaelis-Menten expression for an irreversible reaction takes on the form (Cornish-Bowden, 2004, section 5.2.1)

$$v = \frac{V_{max}[S]}{K_m(1 + \frac{[I]}{K_i}) + [S]} \tag{3.15}$$

Here, the K_i is known as the *inhibition constant* for the inhibitor I, defined as K_i = $[E][I]/[EI]$. A large inhibition constant signifies a small impact on the rate of the reaction. It is often informative to write the two expressions (3.15) and (3.9) on the same form and consider the *apparent* V_{max} and K_m values. These are the changed parameters in the presents of influencing factors. In competitive inhibition, the V_{max} of the enzyme is unchanged, whereas the K_m is increased making the substrate binding more difficult.

Competitive inhibition is the type of inhibition that was used to model the apyrase inhibition in the Pyrosequencing model, see **Papers I** and **II**. Rate expressions on the form (3.15) also apply to reactions subjected to *product inhibition*, that is the negative impact a product can exert on its own production. The inhibitor I is simply replaced by the product P.

3.3.2. Uncompetitive Inhibition

Uncompetitive inhibition occurs when the inhibitor binds to the enzyme-substrate complex ES but not the free enzyme E. One way of looking at this is that when the substrate binds to the enzyme, the enzyme undergoes a conformational change that facilitates the binding of the inhibitor. The resulting ESI complex is unable to form the product and the only alternatives are to remain in the ternary complex or return to the ES form. The reaction has reached a dead-end.

The rate expression for uncompetitive inhibition is (Cornish-Bowden, 2004, section 5.2.3)

$$v = \frac{V_{max}[S]}{K_m + [S](1 + \frac{[I]}{K_i})} = \frac{\left(\frac{V_{max}}{1+[I]/K_i}\right)[S]}{\left(\frac{K_m}{1+[I]/K_i}\right) + [S]} \tag{3.16}$$

Note that, the inhibition constant K_i here is not the same as the inhibition constant for competitive inhibition since the inhibitor now binds to the ES complex. In the uncompetitive situation, the inhibitor can be said to affect both V_{max} and K_m of the reaction, by assigning them lower values than compared to the real reaction.

3.3.3. Non-competitive Inhibition

Non-competitive inhibition is a combination of the competitive and uncompetitive inhibition types in the sense that the inhibitor can bind to both free enzyme and the enzyme-substrate complex. This is more strictly known as *mixed inhibition*, and (pure) non-competitive inhibition is the special case where the dissociation constants of the inhibitor are the same for both E and ES. Assuming that ESI cannot form product and that I binds reversibly to E, the rate can be written as (Cornish-Bowden, 2004, section 5.2.2)

$$v = \frac{\left(\frac{V_{max}}{1+[I]/K_i}\right)[S]}{K_m + [S]} \tag{3.17}$$

In non-competitive inhibition, the K_m is left unaltered, whereas the V_{max} is decreased in proportion to the inhibitor concentration. This means that the inhibitor affects the catalytic properties of the enzyme, but does not influence the substrate's binding to the enzyme. This is only plausible for very small inhibitors such as protons, metal ions and small anions. Hence, non-competitive inhibition is rarely encountered in nature but it is included here for the sake of completeness.

Chapter 4

EXPLORING THE PYROSEQUENCING REACTION SYSTEM

This chapter summarises the work in the present investigation. To begin with, a kinetic model of the Pyrosequencing reaction system is introduced. The model description is presented in full in **Paper I**. In the following two papers, refinements of the model in the form of alternative model descriptions of the apyrase and polymerase are investigated. In **Paper II**, an experimental pre-study of the inhibition of the apyrase is described and a probable explanation for the pulse broadening is given. In **Paper III**, an improved kinetic model of the polymerase is suggested together with possible consequences for the interpretation of the minus-and plus-shift effects.

4.1. A First Model of the Pyrosequencing System (Paper I)

In **Paper I**, a first kinetic model of the Pyrosequencing reaction system was described. Two different models, constructed using the same set of enzymatic reactions, were implemented. The first model simulates single nucleotide incorporation and the other, which is a further development of the former one, is used for simulating sequencing, which involves repeated nucleotide incorporations. Both models are based on an *irreversible* Michaelis-Menten kinetics description for all four enzymes.

4.1.1. One-pulse-model

The purpose of the so-called one-pulse-model was to capture the dynamics of the light signal that is produced following single nucleotide incorporation. Apart from functioning as a starting-point for further model developments, such a model can give valuable information about the different factors influencing the

shape of a light pulse. This is instructive since in order to minimise the time between nucleotide additions during sequencing, signal properties like high amplitude and rapid decay are desirable. Thus understanding the underlying principles that determine the light pulse dynamics is important.

Using parameter optimisation by inspection, the output from the one-pulse-model was fit to an experimental pulse with excellent agreement. The result is pictured in Fig. 4 in **Paper I**. The final parameter setup followed reported literature values rather well (Table 3 in **Paper I**), except in the case of the catalytic constant of the apyrase. The $k_{cat,apyr}$ had to be decreased by a factor of ten for the pulse shape to match experimental data. However, literature values are rare and those reported were recorded under other experimental conditions and show large variations. The kinetic parameters for the apyrase were further discussed in **Paper II**.

The one-pulse-model was used to investigate the effect of the enzyme properties on the system output. This was achieved by varying the activity of the four enzymes and observing the impacts on pulse characteristics such as peak height, rise time and decay time. The main observations are summarised here. The luciferase showed little influence on the pulse properties apart from a natural connection to the pulse height. The polymerase, on the other hand, gave rise to an optimal pulse shape for a certain range of enzyme concentrations, and the dynamics was sensitive to changes in the polymerase velocity. This is partly elucidated by the fact that the nucleotide concentrations are near K_m. An increased ATP sulfurylase activity resulted in a more optimal pulse shape, although the activity above a certain value reached a saturation level where no further improvement was gained. This level lies close to the concentration used in Pyrosequencing experiments. Finally, the rate of the apyrase (for the hydrolysis of ATP) has a large impact on the peak height and decay time of a pulse, which is in accordance with experimental observations.

4.1.2. Reproducing a Pyrogram

New aspects on the reaction system emerge as the model is expanded to handle the sequencing scenario, where nucleotides are iteratively added to the reaction mixture. Aspects such as the minus- and plus-shift were phenomenologically modelled using two constant efficiency factors. These factors were introduced as a simple starting-point but proved to capture the essence of the two effects remarkably well.

Figure 4.1 displays a simulated pyrogram using the first Pyrosequencing model and the corresponding experimental reference case (Gharizadeh *et al.*, 2001). The simulation was carried out with an incorporation efficiency of 99% and a plus-shift effect equal to 3%. These values were chosen since they produced quite realistic pyrograms. Comparisons between the simulated and experimental pyrograms reveal a rather strong resemblance between the two, with all the larger pulses placed in the right positions and of proportionally accurate heights.

The pyrograms also display smaller pulses in between the real light signals.

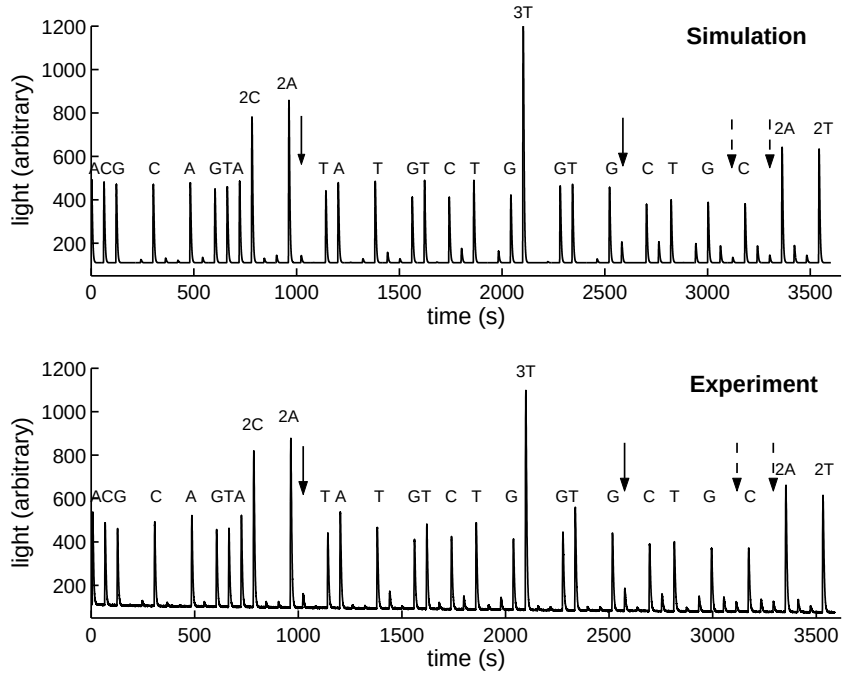

Figure 4.1. *Top: Pyrogram simulated with the Pyrosequencing model in* ***Paper I*** *with an incorporation efficiency of 99% and a plus-shift of 3%. The dispensation order is ACGT. Bottom: The corresponding experimental pyrogram (Gharizadeh* et al.*, 2001). Minus- and plus-shifts discussed in the text are indicated by solid and dashed arrows, respectively.*

These ghost, or false, pulses arise as a result of the above mentioned minus- and plus-shift effects. The minus-shifts appear due to inefficient nucleotide incorporation by the polymerase. In the model, this was replicated by only elongating 99% of the available DNA at each nucleotide addition. The remaining DNA was left unelongated and could then function as a substrate in a polymerising reaction the next time a matching nucleotide was added to the reaction mixture. Typical minus-shifts can for instance be observed in Figure 4.1 after the incorporation of the first double C:s. At the following C dispensation (at 1020 s), there is a small ghost pulse that appears due to the 1% of the DNA that was not elongated at the previous addition of dCTP. An additional minus-shift example appears at the *next* succeeding T dispensation (at 2580 s) following the 3T homopolymer (at 2100 s). The after-effects of this minus-shift are also seen in Figure 4.1 as an overall fall-off in peak height and an increased amount of ghost signals. The minus-shift effect is clearly more pronounced after the synthesis of homopolymers and the problem aggravates as the sequencing proceeds, eventually making the interpretation of the output impossible.

This problem is enhanced by the plus-shift effect. The plus-shifts are caused by incomplete nucleotide degradation by the apyrase. Undegraded dNTPs serve as substrates in the DNA synthesising reaction catalysed by the polymerase, and they allow some templates to be elongated in advance compared to the true template. This was likewise modelled by a constant factor, which forced 3% of all non-matching DNA to be elongated independently of the amount of nucleotides present and independent of dispensation order. The plus-shift effect could not be replicated without the introduction of this efficiency factor, since the apyrase's nucleotide degradation in the simulations was extremely efficient. Examples of plus-shifts can be seen before each of the two double pulses 2A and 2T that terminate the pyrogram in Figure 4.1. The small plus-shift pulses (at 3120 s and 3300 s, respectively) appear at the A and T dispensations preceding the corresponding real pulses, and they are a result of DNA templates that have run ahead of the true template. A plus-shift pulse can also be observed the second last dispensation (at 3480 s). This pulse originates from the triple A:s in the extra sequence GCAAA that follows the plotted sequence (not shown).

It was earlier noted that the causes of the plus-shift effect could not be explained by this first model of the Pyrosequencing system. The effect calls on the presence of remaining triphosphates to elongate a portion of the DNA templates in advance, but the apyrase was found to be extremely efficient. This was also the situation when it came to finding an explanation of the pulse broadening effect. As the sequencing proceeds, reaction products accumulate and seem to have an inhibiting impact on the apyrase, which results in broader pulses. However, the parameter setup used in **Paper I** did not produce any observable pulse broadening and the amount of accumulated diphosphates was very low.

Finally, the model was used to predict the effects of dilution. Each nucleotide addition, 0.2 μl is dispensed to an initial volume of 50 μl. Consequently, after one hundred dispensations, there has been a 40% increase in the reaction volume. This might, in experiments, to at least some extent be compensated by a continuous evaporation. Nonetheless, the simulations showed that dilution in general results in lower peak heights as well as longer rise and decay times.

4.2. Apyrase Inhibition (Paper II)

The origin of the pulse broadening effect was more closely investigated in **Paper II**. The decay time of a light pulse reflects the efficiency of the apyrase. Consequently, it has been suggested (Gharizadeh *et al.*, 2002) that the apyrase is inhibited as more reaction products accumulate. Moreover, this effect seems to be connected to the A nucleotide only. In the Pyrosequencing system, the ordinary dATP has been replaced by the modified nucleotide alpha-thio-deoxyadenosine triphosphate (dATPαS), since natural dATP interferes with the light generating step. The replacement of an oxygen for a sulphur presumably influences the apyrase's ability to degrade the alpha-thio nucleotide. More specifically, the diphosphate dADPαS is regarded as the candidate for the inhibiting effect, since if it were the dATPαS then there would be problems with background signals during sequencing. In addition, the monophosphates usually have very high

K_m or K_i values.

The Pyrosequencing model described in **Paper I** was used to simulate the pulse broadening effect. The broader pulses at later dispensations were obtained in the model by assigning lower values on the kinetic parameters for the modified A nucleotide, and specifically dADPαS. Consequently, a model prediction was that the dADPαS binds stronger to the apyrase (has a lower Michaelis constant K_m) and that the hydrolysis of this nucleotide takes a longer time (has a lower catalytic constant k_{cat}).

The aim of the work behind **Paper II** was to investigate the hypothesis concerning the dADPαS nucleotide's influence on the pulse broadening, by carrying out a series of experimental measurements on the apyrase. The experiments were designed with the purpose of using the Pyrosequencing instrument for the recordings, which quickly gave indicative results. The activity of the apyrase was measured in the presence of all possible tri-, di- and monophosphates available during Pyrosequencing, and compared to a reference activity where ATP was the only substrate. In light of this, the reported values of the inhibition constants K_i should be regarded as qualitative and the true contribution of **Paper II** is the relationships between the K_i values of all the nucleotides in the Pyrosequencing system.

The results from **Paper II** indicate that the modified diphosphate dADPαS has a lower K_m value than the other natural diphosphates. This implies that the alpha-thio diphosphate binds harder to the apyrase and that it therefore is harder for the apyrase to degrade. Equivalent measurements were also performed on the alternative adenosine analogue, c^{7}dADP, which has been shown to produce much less pulse broadening. However, this analogue also exhibited a low Michaelis constant and it was concluded that the difference between the two adenosine nucleotides lie in the values of their catalytic constants k_{cat}. The alpha-thio analogue should have a lower k_{cat} than the c^{7}dADP analogue, making its degradation much less effective with a successive accumulation as a result. The experiments described in **Paper II**, in which repeated ATP dispensations were performed, offer strong support for the conclusion that it is dADPαS that is mainly responsible for the pulse broadening due to its inhibitory effect of the apyrase.

The Pyrosequencing model from **Paper I** was used to verify the experimental observations. Simulations, where the Michaelis constant for dADPαS was decreased with an amount equal to the factor obtained from the relative K_i comparisons, were performed. The results were compatible with the experimental observations – merely decreasing the K_m value for dADPαS was not sufficient to explain the pulse broadening. However, by combining this with a decrease in the catalytic constant k_{cat}, the model output showed successive pulse broadening. An example from such a simulation is shown in Figure 4.2, where a part of a pyrogram (Eriksson *et al.*, 2004b) has been modelled with and without the inhibiting effect of dADPαS. The graph shows the 117$^{\text{th}}$ to 125$^{\text{th}}$ dispensations of the reference pyrogram, meaning there have been 29 A dispensations preceding the plotted dispensations. The broadening effect gets more pronounced as the sequencing proceeds.

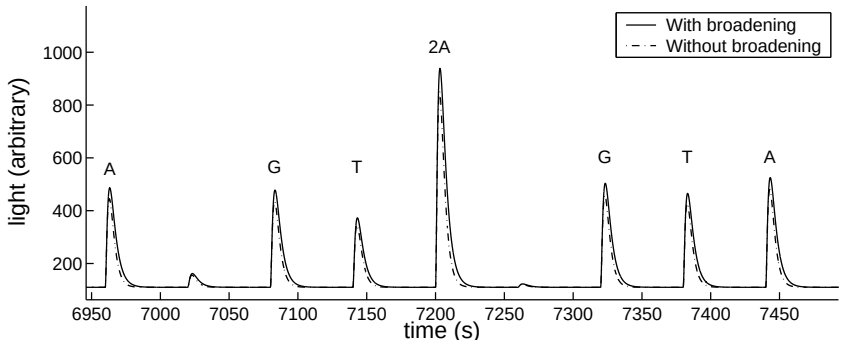

Figure 4.2. *Part of two simulated pyrograms (117^{th} to 125^{th} dispensations) showing the inhibitory effect on the apyrase with resulting pulse broadening. The dash-dotted curve corresponds to a simulation where the dADPαS was treated as the natural diphosphate, and the solid curve corresponds to a simulation where K_m and k_{cat} for dADPαS was decreased, thereby allowing inhibition of the apyrase.*

The simulations also captured the so-called A-effect. In experiments, the incorporation of dATPαS results in slightly higher light signals than for the other nucleotides. The origin of this effect is still unknown but in **Paper II** a new hypothesis is discussed. It was shown that the apyrase degrades the dATPαS nucleotide slightly less efficient than the natural triphosphates (lower K_m value for this particular nucleotide). This implies that more light is produced before the enzyme has removed the triphosphates, with a higher signal as a result.

4.3. Polymerase Aspects (Paper III)

The performance of the Pyrosequencing method is highly dependent on the the efficiency of the incorporation of the deoxynucleotides into DNA by the DNA polymerase. Incomplete incorporation results in a minus-shift effect where background signals complicate the interpretation of the resulting pyrogram. The minus-shift was successfully modelled in the first Pyrosequencing model using a constant inefficiency factor adjusted to 1%. However, a more realistic description would increase the understanding of the origins of this effect and would be more beneficial in future applications of the model. An improved polymerase model should thus account for the rate of the polymerase, time and amount of nucleotides and matching DNA present. In **Paper III**, this was investigated using a more biologically motivated model of the polymerase reaction.

The suggested polymerase model was a reduced form of the complete kinetic mechanism of the Klenow fragment proposed by the group of Stephen J. Benkovic (Dahlberg and Benkovic, 1991; Eger and Benkovic, 1992). The reduced polymerase description included the essential steps of the complete mechanism and was able to reproduce experimental data reasonably well. The use of a reduced model allowed for smoother parameter estimation since the number of adjustable parameters were decreased.

4.3.1. Minus-shift Effect

Once the reduced model had been formulated and the parameters had been adjusted, the model was investigated with the intention of using it to replicate the minus-shift effect. The model should be able to explain the 1% inefficiency in nucleotide incorporation, which was predicted by the first Pyrosequencing model. Simulations with the reduced polymerase model alone showed that the nucleotide incorporation was extremely efficient. Even the polymerisation of a homopolymer of length ten resulted in almost complete (99.94%) elongation after 60 s, which is the time between nucleotide additions in Pyrosequencing. Hence, if the enzyme was this efficient, no minus-shift would be observed.

However, a similar simulation was carried out where the nucleotides simultaneously were degraded by the apyrase to more mimic the situation in the Pyrosequencing system. When the apyrase was included in the model, the nucleotide concentration quickly decreased to a level where the polymerase could no longer perform DNA synthesis with perfect efficiency. The actual value of the inefficiency was very sensitive to the nucleotide concentration, but the 1% was found to be in the range of the experimental concentrations in Pyrosequencing. For the synthesis of homopolymers, the problem with minus-shifts get increasingly serious as the incorporation efficiency declines with the number of identical bases. This was also predicted by the model and seemed to produce accurate estimates of the inefficiency.

The simulation results are summarised in Figure 4.3, where the percentage of elongated DNA (single nucleotide incorporation) is plotted as a function of the nucleotide concentration for: (1) the model where the polymerase acts alone (solid), and (2) the model where the apyrase is allowed to degrade nucleotides alongside the polymerase (dashed). Here, the efficiency of the polymerase is clearly displayed, as well as what happens when the apyrase is present. Typical nucleotide concentrations used in Pyrosequencing lie in the range 0.6 – 3.2 μM (Gharizadeh *et al.*, 2001), depending on nucleotide type. Thus, by assigning different parameter values for different nucleotides, the 1% minus-shift could be captured by the proposed polymerase model.

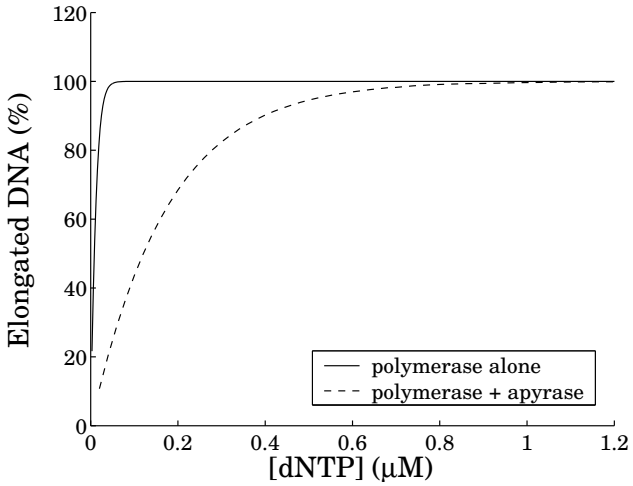

Figure 4.3. *Percentage completely elongated DNA as a function of the nucleotide concentration using the polymerase model. The dashed curve is the result in the presence of the nucleotide-degrading apyrase, and the solid is without apyrase. The concentrations used in Pyrosequencing lie in the range 0.6 – 3.2 μM.*

4.3.2. Plus-shift Effect

The efficiency of the polymerase was also found to, at least partly, explain the plus-shift effect. As the sequencing proceeds, the apyrase is inhibited by reaction products and the triphosphates are not completely degraded. Undegraded nucleotides are then incorporated by the polymerase whenever there is matching DNA. This results in non-specific signals in the pyrogram.

One question at issue has been how much nucleotides that need to be present in order to create plus-shift pulses. The reduced polymerase model provided an indication of the answer. The results of two simulations with the polymerase alone and the polymerase together with the apyrase, are shown in Figure 4.4. Again, the percentage completely elongated DNA is plotted as a function of the nucleotide concentration, but with the focus on *small* concentrations. In order to get results similar to the 3% factor, which was used for modelling the plus-shift effect in the first Pyrosequencing model, there have to be undegraded nucleotides in the order of a few nanomolars[1]. In the presence of the apyrase, the amount obviously has to be slightly larger. This result is supported by earlier experimental observations, where the dispensation of a nucleotide solution diluted by a factor 1000 still gave rise to visible incorporations (Nordström, 2003).

Concluding, using the reduced model description of the polymerase, the appearance of the plus-shift effect can be explained. However, in the existing model

[1]This is naturally dependent on the amount of DNA template.

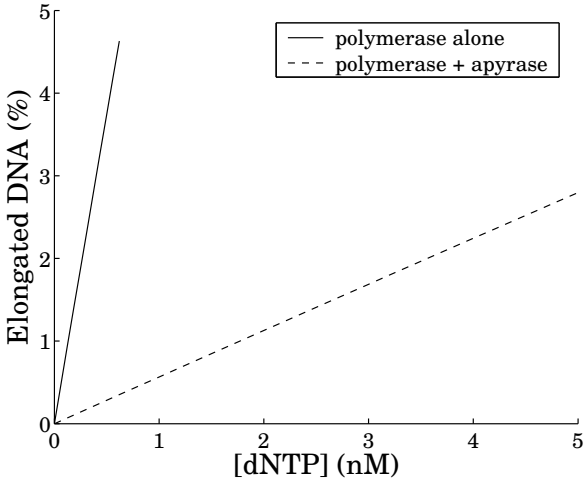

Figure 4.4. *Percentage completely elongated DNA as a function of the nucleotide concentration using the polymerase model. The dashed curve is the result in the presence of the nucleotide-degrading apyrase, and the solid is without apyrase.*

of the Pyrosequencing reaction system, the apyrase is still extremely efficient in degrading the triphosphates. As a consequence, the amount of undegraded nucleotides lies far below the required concentration in the nanomolars range. An alternative explanation to the plus-shift might therefore be the presence of contaminating NDP-kinase, which upgrades dNDP to dNTP. The experimental work that will follow **Paper II** will reveal whether the apyrase is this efficient or not.

Chapter 5

Conclusions and Future Work

The Pyrosequencing method is a DNA sequencing technique that utilises enzyme-coupled reactions and bioluminescence to monitor pyrophosphate release, following nucleotide incorporation, in real-time. In this thesis, a mathematical model of the reaction system employed by the Pyrosequencing method has been developed. The modelling was conducted using basic enzyme kinetics, such as irreversible Michaelis-Menten rate expressions combined with elements of competitive inhibition. The developed mathematical model reproduced a light pulse following single nucleotide incorporation with excellent accuracy. Furthermore, the model was able to capture the overall characteristics of an output pyrogram using a phenomenological approach for the description of the plus- and minus-shift effects. Two constant efficiency factors were sufficient to reproduce these aspects qualitatively. However, the accumulations of triphosphates were too small to account for the plus-shifts in reality. The effects of dilution and apyrase inhibition were also investigated.

An experimental pre-study was performed to establish the causes of the observed pulse broadening effect. This effect is a consequence of the continuous inhibition of the nucleotide-degrading apyrase. The experiments revealed that the inhibitory effect is mainly caused by the alpha-thio diphosphate dADPαS, which is harder for the apyrase to degrade compared to the other nucleotides. The difference lies in a lower Michaelis constant as well as a low catalytic constant for this particular diphosphate.

The mathematical description of the polymerase was improved by the introduction of a more biologically motivated model. The new polymerase model showed that nucleotide incorporation was naturally very efficient, but in combination with a nucleotide-degrading enzyme, such as the apyrase, the incorporation efficiency could be brought down to a level corresponding to the minus-shift factor in the original Pyrosequencing model. In addition, the new polymerase model predicted that for the plus-shift effect to appear in the model, undegraded nucleotides in the nanomolar order need to be remaining in the reaction mix-

ture.

For future analyses of the Pyrosequencing reaction system using both simulations and experiments, the following work is suggested:

- Continue the experimental work on the apyrase to establish the magnitude of the inhibition by dADPαS more quantitatively. This should be accompanied by further simulations with the Pyrosequencing model, investigating the relationship between kinetic parameters and amount of accumulated nucleotides.
- More detailed analysis of the minus-shift using the polymerase model proposed in **Paper III**. The effect should be quantified more carefully based on experimental data, and the parameters in the model should be adjusted to agree with values for each of the four nucleotides.
- Implementation of an expanded Pyrosequencing model where the polymerase model proposed in **Paper III** is incorporated as well as the kinetic parameters controlling apyrase inhibition. Simulations investigating all important effects such as plus- and minus-shift and broadening should follow. This will give a valuable tool for following the dynamics of the reactants in the system.
- Optimisation of the sequencing method using the expanded Pyrosequencing model with respect to enzyme and substrate concentrations.
- Apply time-dependent sensitivity analysis (Ingalls and Sauro, 2003; Schwacke and Voit, 2005) on the Pyrosequencing system in order to pin-point the controlling steps in the reaction scheme and what influences the plus- and minus-shifts the most. This will be very helpful in an optimisation task.
- A sensitivity analysis can also identify enzyme-specific properties, which are limiting to the performance, och thereby point at alternative enzymes that could improve the read-length, e.g. faster and more processive polymerases, more efficient apyrases etc.

Bibliography

AHMADIAN, A., GHARIZADEH, B., GUSTAFSSON, A., STERKY, F. AND NYRÉN, P., Singel-nucleotide polymorphism analysis by Pyrosequencing, *Anal. Biochem.*, 280:103–110, **2000**.

ALBÀ, M., Replicative DNA polymerases, *Genome Biol.*, 2:3002.1–3002.4, **2001**.

ALDERBORN, A., KRISTOFFERSON, A. AND HAMMERLING, U., Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing, *Genome Res.*, 10:1249–1258, **2000**.

ANDREASSON, H., ASP, A., ALDERBORN, A., GYLLENSTEN, U. AND ALLEN, M., Mitochondrial sequence analysis for forensic identification using Pyrosequencing technology, *Biotechniques*, 32:130–133, **2002**.

BAMBARA, R., UYEMURA, D. AND CHOI, T., On the processive mechanism of Escherichia coli DNA polymerase I, *J. Biol. Chem.*, 253:413–423, **1978**.

BHALLA, U., Understanding complex signaling networks through models and metaphors, *Prog. Biophys. Mol. Biol.*, 81:45–64, **2003**.

BRANDAN, E. AND HIRSCHBERG, C., Purification of rat liver N-heparan-sulfate sulfotransferase, *J. Biol. Chem.*, 263:2417–2422, **1988**.

BRIGGS, G. AND HALDANE, J., A note on the kinetics of enzyme action, *Biochem. J.*, 19:339–339, **1925**.

BROVKO, L., GANDELMAN, O., POLENOVA, T. AND UGAROVA, N., Kinetics of bioluminescence in the firefly luciferin-luciferase system, *Biochemistry (Moscow)*, 59:195–201, **1994**.

CORNISH-BOWDEN, A., *Fundamentals of Enzyme Kinetics*, Portland Press, third edition, **2004**.

DAHLBERG, M. AND BENKOVIC, S., Kinetics mechanism of DNA Polymerase I (Klenow Fragment): Identification of a second conformational change and evaluation of the internal equilibrium constant, *Biochemistry*, 30:4835–4843, **1991**.

DE SCHUTTER, E., EKEBERG, O., J., H. K., ACHARD, P. AND LANSNER, A., Biophysically detailed modelling of microcircuits and beyond, *Trends Neurosci.*, 28:562–569, **2005**.

DELUCA, M., Firefly luciferase, *Adv. Enzymol. Relat. Areas Mol. Biol.*, 44:37–68, **1976**.

DELUCA, M. AND MCELROY, W., Kinetics of firefly luciferase catalyzed reactions, *Biochemistry*, 13:921–925, **1974**.

DERBYSHIRE, V., FREEMONT, P., SANDERSON, M., BEESE, L., FRIEDMAN, J., JOYCE, C. AND STEITZ, T., Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I, *Science*, 240:199–201, **1988**.

EGER, B. AND BENKOVIC, S., Minimal kinetic mechanism for misincorporation by DNA polymerase I (Klenow fragment), *Biochemistry*, 31:9227–9236, **1992**.

EHN, M., AHMADIAN, A., NILSSON, P., LUNDEBERG, J. AND HOBER, S., Escherichia coli single-stranded DNA-binding protein, a molecular tool for improved sequence quality in Pyrosequencing, *Electrophoresis*, 23:3289–3299, **2002**.

ELAHI, E., POURMAND, N., CHAUNG, R., ROFOOGARAN, A., BOISVER, J., SAMIMI-RAD, K. AND DAVIS, R. R. M., Determination of hepatitis C virus genotype by Pyrosequencing, *J. Virol. Methods*, 109:171–176, **2003**.

ERIKSSON, J., GHARIZADEH, B., NORDSTRÖM, T. AND NYRÉN, P., Pyrosequencing technology at elevated temperature, *Electrophoresis*, 25:20–27, **2004**a.

ERIKSSON, J., GHARIZADEH, B., NOURIZAD, N. AND NYRÉN, P., 7-deaza-2'-deoxyadenosine-5'-triphosphate as an alternative nucleotide for the Pyrosequencing technology, *Nucleosides, Nucleotides & Nucleic Acids*, 23:1583–1594, **2004**b.

ERIKSSON, J., NORDSTRÖM, T. AND NYRÉN, P., Method enabling firefly luciferasebased bioluminometric assays at elevated temperature, *Anal. Biochem.*, 314:158–161, **2003**.

FAKHRAI-RAD, H., POURMAND, N. AND M., R., Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms, *Hum. Mutat.*, 19:479–485, **2002**.

FORD, S. AND LEACH, F., Improvements in the application of firefly luciferase assays, *Methods Mol. Biol.*, 102:3–20, **1998**.

GHARIZADEH, B., ERIKSSON, J., NOURIZAD, N., NORDSTRÖM, T. AND NYRÉN, P., Improvements in Pyrosequencing technology by employing Sequenase polymerase, *Anal. Biochem.*, 330:272–280, **2004**a.

GHARIZADEH, B., KALANTARI, M., GARCIA, C., JOHANSSON, B. AND NYRÉN, P., Typing of human papillomavirus by Pyrosequencing, *Lab. Invest.*, 81:673–679, **2001**.

GHARIZADEH, B., NORBERG, E., LÖFFLER, J., JALAL, S., TOLLEMAR, J., EINSELE, H., KLINGSPOR, J. AND NYRÉN, P., Identification of medically important fungi by the Pyrosequencing technology, *Mycoses*, 47:29–33, **2004**b.

GHARIZADEH, B., NORDSTRÖM, T., AHMADIAN, A., RONAGHI, M. AND NYRÉN, P., Long-read Pyrosequencing using pure 2'-deoxyadenosine-5'-O'-(1-thiotriphosphate) Sp-isomer, *Anal. Biochem.*, 301:82–90, **2002**.

HENRI, V., *Lois générales de l'action des diastases*, Hermann, Paris, **1903**.

HJALMARSSON, S., ALDERBORN, A., FOCK, C., MULDIN, I., KLING, H., UHLEN, M. AND ENGSTRAND, L., Rapid combined characterization of microorganism and host genotypes using a single technology, *Helicobacter.*, 9:138–145, **2004**.

HYMAN, E., A new method for sequencing DNA, *Anal. Biochem.*, 174:423–436, **1988**.

INGALLS, B. AND SAURO, H., Sensitivity analysis of stoichiometric networks: an extension of metabolic control analysis to non-steady state trajectories, *J. Theor. Biol.*, 222:23–36, **2003**.

International Human Genome Sequencing Consortium, Finishing the euchromatic sequencing of the human genome, *Nature*, 431:931–945, **2004**.

JOYCE, C. AND STEITZ, T., Function and structure relationships in DNA polymerases, *Annu. Rev. Biochem.*, 63:777–822, **1994**.

KARAMOHAMED, S., NORDSTRÖM, T. AND NYRÉN, P., Real-time bioluminometric method for detection of nucleoside diphosphate kinase activity, *Biotechniques*, 26:728–734, **1999**.

KETTLUN, A., ESPINOSA, V., GARCI, L. AND VALENZUELA, M., Potato tuber isoapyrases: substrate specificity, affinity labeling, and proteolytic susceptibility, *Phytochemistry*, 66:975–82, **2005**.

KLENOW, H., OVERGAARD-HANSEN, K. AND PATKAR, S., Proteolytic cleavage of native DNA polymerase into two different catalytic fragments, *Eur. J. Biochem.*, 22:371–381, **1971**.

KOMOSZYNSKI, M. AND WOJTCZAK, A., Apyrases (ATP diphosphohydrolases, EC 3.6.1.5): function and relationship to ATPases, *Biochim. Biophys. Acta*, 1310:233–241, **1996**.

LEE, R., DENBURG, J. AND MCELROY, W., Substrate-binding properties of firefly luciferase: II. ATP-binding site, *Arch. Biochem. Biophys.*, 141:38–52, **1970**.

LI, H., DEYRUP, A., JR., M. J., DOMOWICZ, M., KONSTANTINIDIS, A. AND SCHWARTZ, N., The isolation and characterization of cDNA encoding the mouse bifunctional ATP sulfurylase-adenosine 5'-phosphosulfate kinase, *J. Biol. Chem.*, 270:29453–29459, **1995**.

MARGULIES, M., EGHOLM, M., ALTMAN, W., ATTIYA, S., BADER, J., BEMBEN, L., BERKA, J., BRAVERMAN, M., CHEN, Y., CHEN, Z., DEWELL, S., DU, L., FIERRO, J., GOMES, X., GODWIN, B., HE, W., HELGESEN, S., HO, C., IRZYK, G., JANDO, S., ALENQUER, M., JARVIE, T., JIRAGE, K., KIM, J., KNIGHT, J., LANZA, J., LEAMON, J., LEFKOWITZ, S., LEI, M., LI, J., LOHMAN, K., LU, H., MAKHIJANI, V., MCDADE, K., MCKENNA, M., MYERS, E., NICKERSON, E., NOBILE, J., PLANT, R., PUC, B., RONAN, M., ROTH, G., SARKIS, G., SIMONS, J., SIMPSON, J., SRINIVASAN, M., TARTARO, K., TOMASZ, A., VOGT, K., VOLKMER, G., WANG, S., WANG, Y., WEINER, M., YU, P., BEGLEY, R. AND ROTHBERG, J., Genome sequencing in microfabricated high-density picolitre reactors, *Nature*, 437:376–380, **2005**.

MATTHEWS, L. AND WOOLHOUSE, M., New approaches to quantifying the spread of infection, *Nat. Rev. Microbiol.*, 3:529–536, **2005**.

MAXAM, A. AND GILBERT, W., A new method of sequencing DNA, *Proc. Natl. Acad. Sci. USA*, 74:560–564, **1977**.

MICHAELIS, L. AND MENTEN, M., Die Kinetik der Invertinwirkung, *Biochem. Z.*, 49:333–369, **1913**.

MULLIS, K., FALOONA, F., SCHARF, S., SAIKI, R., HORN, G. AND ERLICH, H., Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction, *Cold Spring Harbor Symp. Quant. Biol.*, 51:263–273, **1986**.

NC-IUB, Symbolism and terminology in enzyme kinetics: Recommendations 1981., Nomenclature Committee of the International Union of Biochemistry (NC-IUB), *Arch. Biochem. Biophys.*, 224:732–740, **1983**.

NOBLE, D., Modeling the heart - from genes to cells to the whole organ, *Science*, 295:1678–1682, **2002**.

NORDSTRÖM, T., *Pyrosequencing: automation, optimization and tempate preparation*, Ph.D. thesis, Royal Institute of Technology, **2003**.

NOURIZAD, N., GHARIZADEH, B. AND NYRÉN, P., Method for clone checking, *Electrophoresis*, 24:1712–1715, **2003**.

NYRÉN, P., Enzymatic method for continuous monitoring of DNA polymerase activity, *Anal. Biochem.*, 167:235–238, **1987**.

NYRÉN, P., Apyrase immobilized on paramagnetic beads used to improve detection limits in bioluminometric ATP monitoring, *J. Biolumin. Chemilumin.*, 9:29–34, **1994**.

NYRÉN, P., Method of sequencing DNA based on detection of the release of pyrophosphate and enzymatic nucleotide degradation, *In U.S. Patent 6,258,568B1*, **2001**.

RONAGHI, M., Improved performance of Pyrosequencing using single-stranded DNA-binding protein, *Anal. Biochem.*, 286:282–288, **2000**.

RONAGHI, M., KARAMOHAMED, S., PETTERSSON, B., UHLEN, M. AND NYREN, P., Real-time DNA sequencing using detection of pyrophosphate release, *Anal. Biochem.*, 242:84–89, **1996**.

RONAGHI, M., NYGREN, M., LUNDEBERG, J. AND NYRÉN, P., Analyses of secondary structures in DNA by Pyrosequencing, *Anal. Biochem.*, 267:65–71, **1999**.

RONAGHI, M., UHLÉN, M. AND NYRÉN, P., Real-time pyrophosphate detection for DNA sequencing, *Science*, 281:363–365, **1998**.

SALA-NEWBY, G. AND CAMPBELL, A., Stepwise removal of the C-terminal 12 amino acids of firefly luciferase results in graded loss of activity, *Biochim. Biophys. Acta*, 1206:155–160, **1994**.

SANGER, F., NICKLEN, S. AND COULSON, A., DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. USA*, 74:5463–5467, **1977**.

SCHWACKE, J. AND VOIT, E., Computation and analysis of time-dependent sensitivities in Generalized Mass Action systems, *J. Theor. Biol.*, 236:21–38, **2005**.

SEGEL, I., RENOSTO, F. AND SEUBERT, P., Sulfate-activating enzymes, *Methods Enzymol.*, 143:334–349, **1987**.

TABOR, S., HUBER, H. AND RICHARDSON, C., Escherichia coli thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7, *J. Biol. Chem.*, 262:16212–16223, **1987**.

THEIL, F.-P., GUENTERT, T., HADDAD, S. AND POULIN, P., Utility of physiologically based pharmacokinetic models to drug development and rational drug discovery candidate selection, *Toxicol. Lett.*, 138:29–49, **2003**.

TRAVERSO-CORI, A., TRAVERSO, S. AND REYES, H., Different molecular forms of potato apyrase, *Arch. Biochem. Biophys.*, 137:133–142, **1970**.

ULLRICH, T., BLAESSE, M. AND HUBER, R., Crystal structure of ATP sulfurylase from Saccharomyces cerevisiae, a key enzyme in sulfate activation, *EMBO J.*, 20:316–329, **2001**.

UTTING, M., HAMPE, J., PLATZER, M. AND HUSE, T., Locking of 3' ends of single-stranded DNA templates for improved Pyrosequencing performance, *Biotechniques*, 37:66–7,70–3, **2004**.

VIVIANI, V., The origin, diversity and structure function relationships of insect luciferases, *Cell. Mol. Life Sci.*, 59:1833–1850, **2002**.

WATSON, J. AND CRICK, F., Molecular structure of nucleic acids, *Nature*, 4356:737–738, **1953**.

Part II.

Papers