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

Gene regulation models of viral genetic switches

MARIA WERNER

Avhandling som med tillstånd av Kungliga Tekniska högskolan
framlägges till offentlig granskning för avläggande av teknologie licentiatexamen
fredagen den 30 november 2007 kl 13.00
i sal RB35, Albanova,
Kungliga Tekniska högskolan, Stockholm.

ISBN 978-91-7178-806-1
TRITA-CSC-A 2007:20
ISSN 1653-5723
ISRN KTH/CSC/A- -07/20- -SE
© Maria Werner, november 2007

Abstract

The recent decades of research in molecular biology have resulted in breakthroughs concerning our knowledge of the genetic code, protein structures and functions of the different cellular components. With this new information follows an increased interest in constructing computational models of the biological systems. A computational model can range from a description of one specific protein to a complete cell or organism. The aim of a computational model is often to complement the experimental studies and help identify essential mechanisms of a system.

All processes taking place in our cells, from general metabolic processes to cell specific actions, originates from information encoded in our DNA. The first step in transferring the genetic information to a functional protein or RNA, is through the transcription of a gene. The transcription process is controlled by cellular proteins binding to DNA regions called promoters. The term "genetic switch", used in the title of this thesis, refers to a specific change in transcription activity, where one or several promoters get activated or silenced.

In this thesis, I present studies of the regulation mechanisms in two different genetic switches. The first is a switch between two central promoters in the Epstein-Barr virus. This human virus is mostly known for causing the 'kissing disease', but is also coupled to several cancer types. Infected cells can change between a resting and a proliferating phenotype, depending on which viral promoter is active. In order to understand what causes uncontrolled proliferation in tumors, it is important to understand the regulation of these viral promoters. The other switch is present in the phage λ, a bacterial virus. This virus has one specific promoter region, controlling expression of two proteins that determine if the phage will remain silent (lysogenic) in the host cell, or start producing new viral particles (go lytic).

For the Epstein-Barr virus we tested, and confirmed, the hypothesis that the regulation of the two central promoters can be obtained by only one viral and one human protein. Further, we studied the cooperative effects on one of the promoters, showing that steric hindrance at the promoter region results in a more effective switching than with only cooperative binding present. For the bacteriophage λ we studied the genetically altered λ-Lac mutants, presented by Little & Atsumi in 2006. We demonstrate that the experimental results cannot, in terms of its equilibria, be explained by the mechanisms generally believed to be in control of the lysogenic/lytic switch.

Keywords: genetic switch, gene regulation, thermodynamic model, phage λ, Epstein-Barr Virus

© Maria Werner 2007

ISBN 978-91-7178-806-1 • TRITA-CSC-A 2007:20 • ISSN 1653-5723 • ISRN KTH/CSC/A--07/20--SE

Sammanfattning

De senaste decennierna av forskning inom molekylärbiologi har resulterat i genombrott vad gäller vår kunskap om den genetiska koden, protein strukturer samt funktion hos de olika cellulära komponenter. Med all ny information följer ett intresse av att konstruera datamodeller för de biologiska systemen. En datamodell kan vara allt från en beskrivning av ett enda protein till en hel cell eller organism. Målet med en datamodell är ofta att komplettera de experimentella studierna och hjälpa till att identifiera viktiga mekanismer i systemet.

Alla processer i våra celler, från de generella metaboliska processerna till de cell specifika reaktionerna, härstammar från informationen i vårt DNA. Det första steget i att överföra den genetiska informationen till ett funktionellt protein, är genom transkriberingen av en gen. Transkriberingsprocessen kontrolleras av cellulära proteiner som binder till regioner på DNA som kallas promotorer. Termen "genetisk switch", som används i titeln på denna avhandling, refererar till en specifik ändring av transkriberingsaktiviteten, där en eller flera promoterer aktiveras eller inhiberas.

I denna avhandling presenterar jag studier av genreglerings-mekanismer i två olika genetiska switchar. Denna första är en switch mellan två centrala promotorer i Epstein-Barr viruset. Detta humana virus är mest känt för att orsaka körtelfeber men är också kopplat till flera cancer typer. Infekterade celler kan byta mellan en vilande och en prolifererande fenotyp, beroende på vilken promotor som är aktiv. För att förstå vad som orsakar okontrollerad proliferering och ev. tumörbildning är det viktigt att förstå regleringen av dessa virala promotorer. Den andra switchen som studerats tillhör λ-fagen, ett bakteriellt virus. Detta virus har en specifik promotorregion som kontrollerar uttrycket av två proteiner som tillsammans avgör huruvida fagen ska ligga vilande (lysogent) i cellen, eller börja producera nya viruspartiklar (gå i lys).

För Epstein-Barr viruset testade vi, och bekräftade, hypotesen att regleringen av de två centrala promotorerna kan förklaras med endast ett viralt och ett humant protein. Därutöver studerade vi den kooperativa effekten hos en av dessa promotorer, och visade att kompetitiv blockering i promoterregionen resulterar i mer effektiv switching än enbart kooperativa bindningar. För bakteriofagen λ studerade vi de genetisk konstruerade λ-Lac mutanterna, skapade av Little & Atsumi år 2006. Vi visar att de experimentella resultaten inte enkelt kan förklaras med enbart den mekanism som generellt anses kontrollera switchen mellan lysogent och lytiskt tillstånd.

Keywords: genetisk switch, genreglering, termodynamisk model, λ-fagen, Epstein-Barr viruset

© Maria Werner 2007

ISBN 978-91-7178-806-1 • TRITA-CSC-A 2007:20 • ISSN 1653-5723 • ISRN KTH/CSC/A- -07/20- -SE

Tillägnad mina mor-och farföräldrar

Acknowledgement

When I finished my undergraduate studies I was relatively certain that I wanted to continue doing a PhD in a biology related field, but preferably without having to spend time at the lab bench. During my months applying for jobs from a sunny Rome, I was fortunate enough to find available openings at KTH, for the new professor Erik Aurell, one of which I got. So I would like to start this acknowledgement with thanking Erik for taking me on as his student. And of course also professor Ingemar Ernberg for welcoming me into his group at MTC, KI.

The thank you list continues with my colleagues at KTH and SU, with whom I have shared my ups and downs during these years. First of all Aymeric, my roommate and personal computer support, for being as pessimistic and cynical as me when needed, and still supportive. Thank you also Jakob, Martin and Marios for being friends and teachers at the same time. And of course Gemma, Elin, Johannes, Smörgåsklubben and all other friendly faces around me in the new group at CB. I also want to thank professor Kim Sneppen and everyone else at CMOL, NBI, for adopting me as an extra student for a while. Especially thank you to Sandeep and Namiko for the personal tuition in physics, and to my life-loving roommate Mille who helped me go from really awful table soccer player to almost ok.

So, having thanked the scientific people there are of course many other important people in my life, who has helped me stay somewhat sane during this time. So, big thank you hugs to Louise, Tina, Hanna, Ulrika, Anna, Anna, Helena, Karin, Åsa and everyone else, for putting up with a slightly work-obsessed friend for a long time. I am also very grateful for my family who believes in me and always supports me. But the biggest thank you goes to my Andreas, who makes me a better and happier person!

Contents

Chapter 1

INTRODUCTION

The recent decades of research in molecular biology have resulted in breakthroughs concerning our knowledge of the genetic code, protein structures, cellular components and much more. With the vast amount of new information arising it is natural for new research areas to develop, with focus on explaining the data given and identifying the mechanisms behind. Hence we have seen a marked increase in the interest of constructing mathematical and physical models of various biological systems, as well as the development of computerized tools for analyzing and storing data. In this way, biology is studied with well established methods that formerly have been applied in physics or computer science. If you come across phrases like "systems biology", "theoretical biology" or "biological physics" they most likely refer to projects where biological data is used to construct an *in silico* model or studied analytically, in search for a way to understand, predict and maybe control some biological system.

There are many alternative ways to model a biological system. Depending on the system itself, whether it is a population, an organism, a cell or an even smaller part, one has to chose the appropriate model. Regarding cellular components one can chose to look at an atomistic level, while for whole cellular processes, involving many components, this can easily become too computationally demanding. One then generally has to model at the macroscopic level. Although the choice of microscopic or macroscopic level might come quite naturally, a more difficult part can be to define the range of the system itself. Often the amount of data available about the components in the system sets the limit on how detailed the model can be. For most theoreticians working with constructing the model, it is of importance to have a good understanding of the system itself, and preferably have a collaboration with experimentalists that can provide data. The better understanding of the system, the easier it is to see what components are relevant or can be left out in order to simplify the model.

Different methods of modeling address somewhat different questions, and hence no one method is necessarily better than the other. They instead com-

plement each other in the often complicated process of fully understanding a specific system. When it comes to molecular biology, new promising techniques are being developed and many are already available today. This includes efficient production of genetic constructs, high-throughput measurements of proteins or genetic materials, and more recently also the techniques for single cell studies enabling detailed measurements of reaction rates (Cookson *et al.*, 2005; Maerkl and Quake, 2007; Elf *et al.*, 2007). All these new techniques will hopefully, together with theoretical models, lead to better understanding of all the remarkable processes taking place in our cells.

1.1. Scope of the thesis

The projects presented in this thesis concern models of gene regulation in two different viral systems. The model technique has been the same for both systems, using equilibrium statistical mechanics, while the goals with the three projects varied. For the first study of the Epstein-Barr virus, (**paper I**), I tested a hypothesis concerning which transcription factors were involved in the regulation of two central viral promoters. The second study, (**paper II**), was a follow up looking closer at the cooperative effects on one of the viral promoters. The third project, (**paper III**), concerned the bacteriophage lambda, where I modeled two genetically altered lambda circuits, experimentally studied by Little & Atsumi in 2006. The aim was to see whether the experimental results could be understood and explained from a theoretical analysis.

1.2. List of papers

Paper I. Maria Werner, Ingemar Ernberg, JieZhi Zou, Jenny Almqvist and Erik Aurell. Epstein-Barr virus latency switch in human B-cells: a physico-chemical model. *BMC Systems Biology*, 1:40, 2007 (Provisional PDF online at time of printing).
Paper II. Maria Werner, LiZhe Zhu and Erik Aurell. Cooperative action in eukaryotic gene regulation: physical properties of a viral example. Accepted for publication in PRE, October 30th 2007.
Paper III. Maria Werner and Erik Aurell. A computational systems biology study of the λ-lac mutants. Manuscript in preparation.

1.3. Contribution to papers

Paper I. I did the background literature search to find available biological data on which to base the model. I constructed the statistical model, which I implemented. The results were analysed together with Erik Aurell and Ingemar Ernberg and I wrote the major part of the paper.
Paper II I planned the project together with Erik Aurell, and we coordinated the project during the implementation phase. LiZhe Zou wrote the computer code

and performed most of the simulations. Erik Aurell and I did the analysis and wrote the paper together.
Paper III. I did the background literature search to find all parameters, or if not available, the means of estimating them from available data. I constructed the model, implemented it and performed all simulations. I analyzed the results together with Erik Aurell and we wrote the manuscript together.

Chapter 2

GENE REGULATION

All processes taking place in our cells, from general metabolic processes to cell specific actions, originates from information encoded in our DNA. In order to control all this genetic information and translate it into a coordinated cell with specific properties, it takes a very complex regulatory machinery. This chapter aims to give a brief insight into one area of this machinery, the gene transcription process. Transcription of a gene is the first step in transferring genetic information to a functional protein or RNA, and its regulation is in general a very complex non-linear function of DNA binding proteins and mechanisms. This chapter does not cover all details of the process, but rather gives an introduction to the fundamental mechanisms.

2.1. The transcription mechanism

Transcription is the process where a specific gene in the DNA is copied, by the RNA polymerase enzyme, into a single stranded nucleotide molecule, RNA. The RNA can then be used as a template for the ribosome that translate the nucleotide code into an amino acid sequence, eventually forming the functional protein. In eukaryotes there are also various post-transcriptional alterations of the mRNA before it can be correctly translated. Although all steps in the chain from DNA to the complete functional protein can be modified, the most fundamental control lies at the transcriptional level. The transcriptional process can be very tightly regulated, to what degree depending on the gene and the complexity of the organism.

The transcriptional control mechanisms differ between prokaryotic and eukaryotic cells, where the latter cells are more complex. The fundamentals are however the same. First, the RNA polymerase enzyme has to get attached close to the transcription start site of the gene. This site, called core promoter site, is located upstream of the gene transcription start site. In prokaryotes the polymerase can attach directly to the core promoter, while in eukaryotes it is first

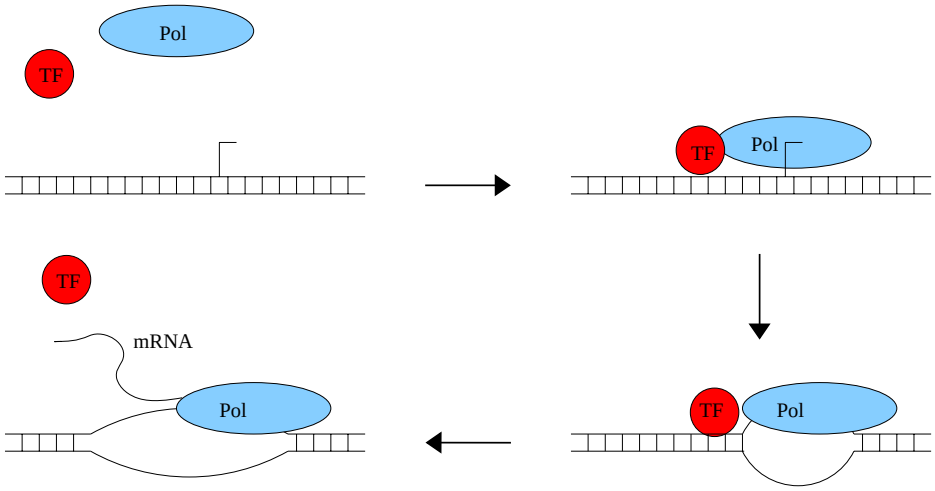


Figure 2.1. *Illustration of the transcriptional mechanism. The polymerase is helped by transcription factors (TFs) to bind to the core promoter site. Transcription is then initiated by unwinding of the DNA double helix into two single strands, the coding strand and the template strand. The polymerase can thereafter travel along the template strand and copy each nucleotide into a single stranded messenger RNA strand, mRNA.*

identified by general transcription factors that help the RNA polymerase to bind (Orphanides *et al.*, 1996). Thereafter, transcription is initiated through unwinding of the two DNA strands into a bubble so that the template strand becomes accessible. The next step is the elongation, where the RNA polymerase travels along the template strand, reading each nucleotide and copies it to form the mRNA. Eventually, when the whole gene has been transcribed, the elongation terminates. In prokaryotes there is a specific termination sequence located after each gene, while in eukaryotes the termination signals are not completely understood and transcription often continue past the protein coding region (Ogbourne and Antalis, 1998). Figure 2.1 illustrates the different steps in the transcriptional process.

2.2. Promoter structure

The core promoter is the sequence on the DNA that is recognized by the general transcription factors in eukaryotes, or directly by the RNA polymerase complex in prokaryotes. In prokaryotes this part usually constitutes the total promoter, while in eukaryotes the promoter is often much more complex and spans over longer segments of the DNA. The sequence of the core promoter differs, but a common motif in eukaryotes is the TATA sequence, located 25-30 bp upstream of the transcription start site (Orphanides *et al.*, 1996). Also prokaryotes have conserved sequences about 10-30 bp upstream of the initiation site (Serfling

et al., 1985).

For prokaryotes, transcription activity is largely determined by the rate of elongation by the polymerase, and not dependent on activating transcription factors (Struhl, 1999). Binding of the polymerase itself can therefore be enough to yield strong transcription. In eukaryotes, the core promoter is sometimes referred to as the basal promoter, since it can only enable a basal level of gene expression, *i.e.* it initiates transcription at a low rate. In order to induce a higher transcriptional efficiency, and specificity, eukaryotic promoters also include various types of binding sites for gene specific transcription factors. Unlike the general transcription factors that always are present in the cell, the level of these gene specific transcription factors vary between cell types and with environmental conditions and the cell's development. This permit gene transcription to be regulated much more precisely in time and quantity.

2.2.1. Transcription factors

Transcription factors (TFs) are proteins that bind DNA promoter regions and regulate the transcriptional efficiency. They all have a DNA binding domain, although the structure and function of this domain vary (Wray *et al.*, 2003). Different subfamilies share the same DNA binding domain that can be strictly repressing or activating or have both functions. There are also cases in which a transcription factor has two DNA binding domains, one repressing and the other activating, and hence can function in two ways (Wray *et al.*, 2003). To enable transcriptional control, most transcription factors also have a protein binding domain. This domain may directly interact with the basal transcriptional machinery at the core promoter, or with co-factors, other proteins that mediate the regulatory interactions.

2.2.2. Binding sites and regulatory regions

The DNA sequence recognized by the transcription factor can vary in length, but usually they span 5-20 bp (Wray *et al.*, 2003). Interestingly, the binding sites for eukaryotic factors appear to be shorter than their analogues in prokaryotes (Bilu and Barkai, 2005). The preferred binding site for a transcription factor is often referred to as the consensus site (Stormo, 2000). The factors' affinity is strongest for the consensus site, but minor sequence variations are usually tolerated, at the cost of lower affinity but without loss of function. Co-factors or ligands bound to the transcription factor bound may alter the binding affinity, through modulation of the DNA binding domain structure.

Since the binding site of a specific transcription factor is relatively short, there is a significant probability that there will be many sites on the genome to which the transcription factor can bind. Therefore binding sites are often found grouped together, enabling cooperative actions between neighboring bound transcription factors.

In eukaryotes, the specific transcription factor binding sites can be located both upstream and downstream of the core promoter, and positioned as far as 1000 bp away (Serfling *et al.*, 1985; Ogbourne and Antalis, 1998). Therefore it

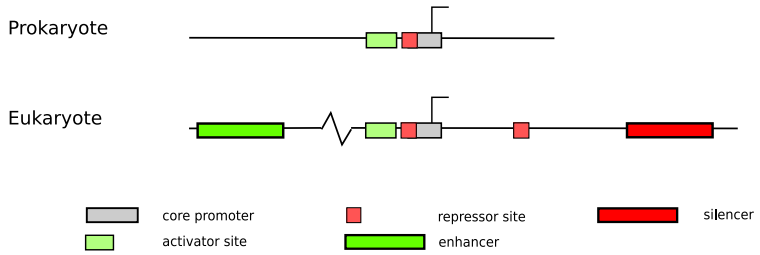


Figure 2.2. *Schematic illustration over typical prokaryotic and eukaryotic promoter structures. Prokaryotes can have binding sites for both repressors and activators, although it is not always necessary. When existing, they are often placed on close proximity to the core promoter site. In eukaryotes, transcription activation is strongly dependent on activation, and therefore the promoter architecture is typically more complex. Various numbers of activating and repressing sites can be included, as well as larger regulatory blocks, enhancers or silencers.*

might be difficult to identify all regulatory binding sites involved in controlling a gene. The number of involved transcription factors also varies greatly, depending on how tightly the specific gene expression must be controlled. Longer regions of binding sites, which activate transcription independent of position and orientation with regard to the transcription start site, are often referred to as enhancers (Arnosti and Kulkarni, 2005). Enhancers with the same structure as in eukaryotes have been identified in several different viruses (Serfling *et al.*, 1985). Just like enhancers, position- and orientation independent regions exist that acts as repressors on transcription. These are commonly referred to as silencers (Ogbourne and Antalis, 1998). Recently, insulators, segments able to block enhancer action have been identified (Brasset and Vaury, 2005).

In prokaryotes, activation of transcription is not always necessary, but may still occur if the promoter is weak (Struhl, 1999). Most often the regulatory mechanisms are simpler, with binding sites for transcription factors in close proximity to, or overlapping with, the polymerase binding site. Figure 2.2 illustrates what a typical prokaryotic and eukaryotic promoter region may look like.

2.2.3. Regulatory actions by transcription factors

A eukaryotic promoter can include both different enhancers, silencers and other transcription factor binding sites in various combinations. In addition, one specific transcription factor may both repress or activate, depending on the position of the binding site and if it binds alone or with modifying cofactors. This complex situation means that simply locating a binding sequence, or region, close to a transcription start site, does not tell you what function it plays in the regulation. The only way to fully understand the regulation of a promoter is to conduct *in vivo* experiments with mutated or deleted promoter regions and analyze the transcription patterns.

There are many different ways in which transcription factors alter the transcriptional efficiency. The most straightforward method for repression is to place binding sites overlapping the polymerase attachment sequence (Serfling *et al.*, 1985). This ensures a steric hinder for the polymerase to attach to DNA, but the method's efficiency is dependent on the transcription factor's affinity to the sites. The reversed mechanism can take place if sites for activators are placed close to the core promoter. Activators can then directly interact with the polymerase to increases its binding affinity for a weak core promoter. If the affinity is strong, a direct interaction with the polymerase, might in fact lead to a decreased transcription if the polymerase then are kept too tightly at the core and cannot start the elongation. From these simplest cases, almost an endless number of repressing and activating actions may be combined. A physical distance between the enhancer and the core promoter can be overcome by DNA looping, a phenomena observed both in eukaryotes, prokaryotes and viruses (Revet *et al.*, 1999; Frappier and O'Donnell, 1991a).

What has not been discussed here are certain other types of regulatory mechanisms. The eukaryotic genome is organized in higher order structures, with the DNA wrapped around histone proteins, forming a more condensed structure called chromatin. The histone structures can change, depending on if they become methylated or acetylated, and thereby the chromatin structure becomes more or less condensed, hence affecting the possibility for polymerase to bind (Turner, 2000; Jones and Takai, 2001). Thus, a transcriptional regulation at higher order takes place when transcription factors that are still able to bind to the condensed chromatin, recruit chromatin modifying enzymes (Struhl, 1999). Silencing of genes through chromatin methylation can be inherited to daughter cells at cell division. This type of heritable control in gene expression, without DNA alterations, is referred to as epigenetics (Wolffe and Matzke, 1999; Goldberg *et al.*, 2007).

2.3. Genetic switches

If you look up the word switch in the dictionary, it can be explained as a turning, shifting or changing. In biology, the term "genetic switch" usually refers to a specific change in promoter activity. It can be that one promoter switches its transcription from off to on, or vice versa, or there can be a switch between two different promoters being active. Of course, promoters get turned off and on constantly in the cell, while not all of these events are referred to as switches. In general, one talks about a genetic switch when there is a significant change in promoter activity over a short period of time, or due to a small change in the cellular environment. A cell may for example switch between two important phenotypes depending on the environmental conditions, or a virus switches between entering the latent or lytic pathway.

Since all cellular activities depend ultimately on gene expression, an observed phenotypic change might be caused by a specific genetic switch. In medicine it is vital to identify and understand changes in cellular phenotypes that are characteristic for different diseases. If one then can identify the spe-

cific genetic alterations, causing an observed phenotypic change, it can help in the development of treatments. Especially within tumuor biology, much focus lies on identifying the genetic alterations, and sometimes switches, causing the growths of tumor cells.

In this thesis, two different viral genetic switches will be presented. One in a human virus, the Epstein-barr virus, important for cell proliferation onset, and possbily coupled with tumor-induction. The other switch is in a bacterial virus, central for the decision between entering a resting state in the cell or inducing production of new viral particles.

Chapter 3

PHYSICS IN BIOLOGY

This chapter will present one way of modeling gene expression with statistical mechanics as a basis. In order to understand these models, some background in biochemistry, thermodynamics and statistical mechanics is necessary, and will therefore be presented here.

3.1. Molecular reactions and basic kinetics

Proteins and other molecules in our cells are involved in many types of different processes. Most often these processes involve two or more molecules forming a complex. The binding reactions can be fast or slow, and the complexes may be stable or unstable. The field of kinetics in biochemistry is concerned with understanding molecular reactions and how to describe them with common notations and definitions.

Them most fundamental reaction is when two molecules, A and B, encounter each other and form a complex, AB:

$$A + B \rightleftharpoons AB \tag{3.1}$$

The formation rate of AB depends on the concentrations of the two molecular species, $[A]$ and $[B]$, as well as the reaction rate k_1 (3.2), while the dissociation rate of the complex depends on the concentration of the complex $[AB]$ and the dissociation rate k_{-1} (3.3).

$$A + B \xrightarrow{k_1} AB \tag{3.2}$$

$$AB \xrightarrow{k_{-1}} A + B \tag{3.3}$$

When the reactions in both directions are equally fast, the system is said to be in *equilibrium*. This does not mean that no reactions take place, just that they cancel out so there is no detectable change in concentrations. A more general

term sometimes used is *steady state*. Steady state however also includes the static state where no reactions take place, as well as a when there is a constant flux at non-equilibrium.

Even though it is to some extent possible to measure reaction rates, many times it is sufficient to look at the system in equilibrium, and hence only the rate quotient is needed:

$$\begin{aligned} k_1[A][B] &= k_{-1}[AB] \\ K_d &= \frac{k_{-1}}{k_1} = \frac{[A][B]}{[AB]} \\ K_a &= \frac{1}{K_d} \end{aligned}$$

The constant K_d is the equilibrium dissociation constant, and K_a the equilibrium association constant. The equilibrium constant can also be expressed in terms of fractional occupation, Y; the amount of bound B, $[AB]$, out of the total available molecules $[B_{total}]$= $[B] + [AB]$:

$$Y = \frac{[AB]}{[B_{total}]} = \frac{[AB]}{[AB] + [B]} = \frac{[A]}{K_d + [A]}$$

K_d is then the concentration at which the complex is half saturated.

Cooperative processes

One important biochemical mechanism in kinetics is the cooperative process, where the first binding of a protein or ligand affects further bindings of the same kind. Often a cooperative process is mentioned when a protein has several ligand binding sites, where one ligand bound increases the affinity for the subsequent ligands. But cooperativity is also present when proteins bind together forming multimers, or bind DNA together. A quantitative measure of cooperativity is the Hill coefficient, where a coefficient of 1 indicates independent binding while a higher positive integer indicates positive cooperation. The higher coefficient, the higher degree of cooperativity. Suppose a complex with A_xB is formed, *i.e.* x number of A molecules and one B molecule bind together. If each A molecule bind individually to B with dissociation constant K, the fractional occupation reads:

$$\frac{[A_xB]}{[B_{total}]} = \frac{1}{1 + (\frac{K}{[A]})^x} \tag{3.4}$$

From this one can compute the Hill coefficient by plotting $log\frac{Y}{1-Y}$ versus log of free concentration,$[A]$:

$$\begin{aligned} log(\frac{Y}{1-Y}) &= log(\frac{[A_xB]}{[B]}) \\ &= log((\frac{[A]}{K})^x) \\ &= xlog([A]) - log(K^x) \end{aligned}$$

Plotting $log(\frac{Y}{1-Y})$ versus $log[A]$, will then generate a curve with slope x, defined as the Hill coefficient. Per definition, one commonly refers to the effective Hill coefficient, taken at half saturation, $\frac{Y}{1-Y} = 1$.

3.2. Thermodynamics and statistical mechanics

Thermodynamics is an area mainly concerned with transformations between different kinds of energies; heat, mechanical work or chemical work, and the macroscopic properties of a studied system, such as pressure and temperature. In thermodynamics a system is defined as any region completely enclosed within a well defined boundary. This means that one can define a system as completely isolated from the surroundings, with no possible energy transformation, or allow energy flow and particle flow in and out of the system. Statistical mechanics enables macroscopic predictions based on microscopic properties of the atoms or molecules in the system. The macro-state of the system is computed from probability distributions on ensembles of the microscopic configurations possible.

Thermodynamics and statistical mechanics are broad areas in physics, applicable in many different research areas. This thesis' focus is the use of statistical mechanics in description of genetic regulation. Hence, fundamental concepts of importance for these purposes will be introduced and derived, while other areas of these fields are left out.

3.2.1. Energy and entropy

Two central quantities in thermodynamics are the internal energy and the entropy. The internal energy of a system consist of the sum of the kinetic and potential energies of all particles in the system. According to the first law of thermodynamics, the internal energy of a closed system cannot change. This is one of the most fundamental principles in physics. Most often, the absolute energy in a studied system is not of interest, only the changes brought upon it due to changes in the environment.

The quantity entropy is sometimes referred to as a measure of the molecular disorder. In thermodynamics, entropy can be defined as the extent to which energy is dispersed in a disorderly manner, due to heat transfer to the system during a reversible process (Atkins, 1998);

$$dS = \frac{dq_{rev}}{T} \tag{3.5}$$

where q_{rev} is the transferred heat and T is the temperature of the system.

In statistical mechanics, the entropy is instead defined as a function of the number of available configurations in the system;

$$S = k_B ln(g) \tag{3.6}$$

where g is the number of accessible configurations and k_B is the Boltzmann constant. In a closed system the entropy will always increase, until maximum entropy is reached, where the system is in equilibrium.

3.2.2. Thermodynamic potentials

The mathematical representation of the internal energy uses extensive parameters, such as the entropy (S) and the volume (V) and the number of particles (N), as independent variables. An extensive parameter is proportional to the number of particles in the system. However, these parameters are not always easily controlled for all processes, and alternative functions to describe the energy changes in a system are needed. Transforming the original mathematical representation into functions using the intensive parameters, temperature (T), pressure (P) and chemical potential (μ), as the independent variables, results in more convenient representations. In thermodynamics these representations are referred to as thermodynamic potentials.

There are three thermodynamic potentials used under different conditions: the Helmholtz energy (F), the enthalpy (H) and the Gibbs free energy (G). In the Helmholtz energy the entropy parameter has been replaced by the temperature. For the enthalpy, the volume parameter is replaced by the pressure parameter. In the Gibbs free energy, both the entropy and the volume parameters have been replaced with the temperature and the pressure. In most chemical and biological processes studied, both the temperature and the pressure are held constant, leading the Gibbs potential to be the preferred potential to work with.

$$\begin{aligned} U(S,V,N) \qquad & dU = TdS - PdV + \mu dN \\ F(T,V,N) \qquad & dF = -SdT - PdV + \mu dN \\ H(S,P,N) \qquad & dH = TdS + VdP + \mu dN \\ G(T,P,N) \qquad & dG = -SdT + VdP + \mu dN \end{aligned}$$

3.2.3. Statistical ensembles

An ensemble is an *imaginary* collection of many systems, all replicas of the *actual* system of interest, that are all considered to give statistical information about the macroscopic state of the system. Each *copy* should represent the system in one configuration, so in theory the ensemble consists of at least the same number possible configurations the system can exist in. Studying the statistics of an ensemble is the theoretical equivalent of repeating a physical experiment with the same macroscopic conditions, several times and looking at the statistics.

There are three types of ensembles used for different types of systems; the micro-canonical, the canonical and the grand canonical ensemble. In the micro-canonical ensemble all systems are thermally isolated, thereby having the same total energy. The systems in a canonical ensemble are assumed to be in contact with a large heat reservoir, enabling heat exchanges at a fixed temperature. In the grand canonical ensemble, the systems are in thermal contact with a reservoir and can in addition to heat also exchange particles. Which ensemble to use for a study depends on how one defines the system of interest.

Canonical ensemble

A closed system with total energy E_i can exist in $g(E_i)$ number of configurations. Each configuration is characterized by the position and momentum of the particles in the system, their separate energies adding up to the total energy E_i. The fundamental postulate in statistical mechanics is that given the closed system in equilibrium, it is found equally likely in any one of its configurations (Kittel and Kroemer, 2000). If the system is in thermal contact with a reservoir, the total energy of the reservoir and the system, E_{tot}, will be held constant. However, the systems energy E_i can vary. Since all configurations are equally probable, the likelihood of the system to have energy E_i is related to the number of configurations giving the energy E_i of the system and $E_{tot} - E_i$ for the reservoir (Kittel and Kroemer, 2000);

$$\begin{aligned}\frac{P(E_1)}{P(E_2)} &= \frac{g(E_{tot}-E_1)g(E_1)}{g(E_{tot}-E_2)g(E_2)} \\ &= \frac{e^{S(E_{tot}-E_1)/k_B+S(E_1)/k_B}}{e^{S(E_{tot}-E_2)/k_B+S(E_2)/k_B}}\end{aligned}$$

Expanding the entropy around E_{tot}, gives;

$$\frac{P(E_1)}{P(E_2)} = \frac{e^{S(E_{tot})/k_B - E_1(\frac{\delta S}{\delta E_{tot}})/k_B+S(E_1)/k_B}}{e^{S(E_{tot})/k_B - E_2(\frac{\delta S}{\delta E_{tot}})/k_B+S(E_2)/k_B}} \tag{3.7}$$

Written more conveniently, with the reservoir entropy canceling out:

$$\frac{P(E_1)}{P(E_2)} = \frac{g(E_1)e^{-E_1/K_BT}}{g(E_2)e^{-E_2/k_BT}}$$

The probability to find the state with a certain energy is hence proportional to the number of configurations of this state and the exponential of the relative energy in that state, E_i. Most often one is interested not only in the relative probabilities of two states, but the general probability of being in one particular state amongst all other, $P(E_1)$. This is achieved by normalizing the weight of state E_1 with the sum of all statistical weights, $Z = \sum_i g(E_i)e^{-E_i/k_BT}$:

$$P(E_1) = \frac{1}{Z}g(E_1)e^{-E_1/k_BT} \tag{3.8}$$

The quantity Z is known as the canonical partition function.

Grand canonical ensemble

For a system which has both energy flow and particle flow, the probability of being in a certain state $s(E_i, N_i)$ can be found in a similar manner as in the canonical ensemble. The difference is that the particle number, N, needs to be taken into account when looking at the individual states of the system. In the same way presented for the canonical ensemble, $P(E_1, N_1)$ can be derived from the number of configurations $g(E_1, N_1)$, normalized with the sum of all configurations.

$$P(E_1, N_1) = \frac{g(E_1, N_1)e^{N_1\mu_1/k_BT - E_1/k_BT}}{\sum_i g(E_i, N_i)e^{N_i\mu_i/k_BT - E_i/k_BT}} \tag{3.9}$$

As particle flow is allowed in this ensemble, the probability now includes a term taking into account the number of molecules and their chemical potential. The chemical potential is defined as the Gibbs free energy per particle and in spontaneous reactions, particles flow from higher potentials to lower potentials. In a biological system the chemical potential is essentially the logarithm of the concentration (Atkins, 1998).

As described briefly in section 3.2.2, most biological systems are studied at constant pressure and temperature. Therefor the experimentally measured energies available are most often given as Gibbs free energies. Mathematically, the expansion of the entropy, at constant pressure and temperature, allows us to use the Gibbs free energy (see Appendix).

3.3. Statistical mechanics in gene regulation models

The idea behind thermodynamic models of gene regulation is that the expression level of a gene can be computed from the equilibrium probabilities that transcription factors and polymerase are bound to the promoter region. This is of course to simplify the complex regulation of gene transcription that occurs *in vivo* into the question of probability of transcription. There may be several processes intervening after the initiation of transcription occurs, hindering a full transcription. However, these models give a straightforward relation between the gene expression and the concentration of transcription factors and polymerase given for the system, and they have been shown to be valuable in predicting and understanding promoter activities (Shea and Ackers, 1985; Arkin *et al.*, 1998; Dodd *et al.*, 2004; Aurell and Sneppen, 2002).

3.3.1. Model assumptions and framework

The most fundamental assumption in the usage of the equilibrium statistical mechanics for computing molecules binding to DNA, is that the system is in equilibrium. This can only be justified by looking at the time-scales for the different events in the transcription process. Equilibrium binding can only be assumed if any significant changes in molecular concentrations are slower than the on and off rates for transcription factor binding to DNA (Bintu *et al.*, 2005).

For a single molecule, the association rate can be estimated from diffusion-limited association, yielding a rate in order of 10ths of seconds for most operator sites (Aurell *et al.*, 2002). Therefore, it has generally been accepted that the occupational state of the promoter sites can be given by the equilibrium constants. However, one needs to be aware that for formation of larger complexes, the association and dissociation time-scales might be on the limit of this equilibrium approximation.

When constructing a model it is necessary to know the architecture of the promoter and which transcription factors are important for inducing or preventing transcription. To give a correct description, experimental data concerning equilibrium binding constants for the involved transcription factors are then necessary. From the promoter structure one can thereafter set up the model where the binding sites, with or without bound factors, can be considered the small system, exchanging particles and energy with the larger reservoir, the cytoplasm. This then allows for a statistical mechanical description with the grand canonical ensemble.

The various involved transcription factors $[TF_1], [TF_2], \ldots, [TF_m]$ can bind into different states, s, at the promoter region. For each state, $N_i(s)$ notes the number of bound molecules of each kind, and ΔG_s, the Gibbs free energy. From eq. 3.9, the probability of state s for the promoter region then reads:

$$P_s = \frac{1}{Z}[TF_1]^{N_1(s)}[TF_2]^{N_2(s)} \ldots [TF_m]^{N_m(s)} exp(-\frac{\Delta G_s}{k_B T}) \tag{3.10}$$

where Z is the grand canonical partition function, T is the system temperature and k_B is the Boltzmann constant. If there is degeneracy, where the set of transcription factors can bind in several configurations, with the same ΔG energy, one also includes the number of configurations $\xi(N_1, N_2, \ldots, N_i)$. The net transcription rate, $\Re$, from the promoter is given by summing over the rates for all states:

$$\Re = \sum_s P_s \Re_s \tag{3.11}$$

Some states, s, will naturally not contribute to the transcription rate, depending on the biological requirements for the promoter activation. For transcription to be possible, the binding of RNA polymerase is of course required at the start site. Whether to include the binding of the polymerase in the state description depends on the focus of the model. If one not particularly focuses on the polymerase binding differences between different promoters, one may assume it is a constant factor in the system, and not include it in the state description. Both approaches have been used in gene regulation models (Dodd *et al.*, 2004; Aurell *et al.*, 2002; Aurell and Sneppen, 2002; Shea and Ackers, 1985).

3.3.2. Example of the simplest scenario

The simplest example of applying the modeling approach described here is when a promoter has one single binding site for a protein A. Suppose the promoter is

active when A is bound and otherwise silent. Then there are two possible states of the promoter, with no bound A and gene transcription being off, and with one bound A molecule and gene transcription active.

To find the likelihood of the promoter transcription being active, $P(on)$, one then computes the probability of A being in the bound state out of the two possible states:

$$\begin{aligned} P(on) &= \frac{e^{\mu_A/k_BT - G_1/k_BT}}{\sum_i e^{A_i\mu_i/k_BT - G_i/k_BT}} \\ &= \frac{[A]e^{-G_1/k_BT}}{1 + [A]e^{-G_1/k_BT}} \end{aligned}$$

The energy for the unbound state, G_0, can be set to zero, since one only uses relative energies. In this example, the expression is relatively simple since there are only two states, and only one possible configuration for both states. Most often however, as discussed in chapter 2, the promoter architecture is more complex and can include both inhibitors and activators binding to several sites, thus generating more possible configurations $\xi(n_i, G_i)$. Note however that this simple scenario described here is not an efficient genetic switch, since the transcription is a linear function of the transcription factor concentration. To create a functional switch, that can turn on gene transcription with only a small change in transcription factor concentration, one needs cooperative actions like dimers or even bigger complexes, forming at the operator region.

For some promoters, there can also be DNA looping , where bound transcription factors at a promoter region further upstream interact with the basal machinery. This interaction then comes in both in the configuration term and with as additional energy term for the looped states. The additional energy term then includes both the entropy cost in forming a loop, as well as an eventual gain in interaction energy. For processes where there is cooperative binding between DNA bound transcription factors, this extra energy gain is also included in the Gibbs free energy for each state that allows the interaction.

3.4. Deterministic or stochastic models?

When modeling any system at the macroscopic level, studying the amount of molecules in the system and their interactions, one can chose to take two different model approaches. One way is to study the average level of molecules at each time-point, and use a deterministic description of all the processes in the system. One then assumes a continuous number of molecules that are homogeneously distributed in the system volume, and that reactions occur instantly and in a deterministic manner. A deterministic description of the kinetics presented in eq 3.1 would be:

$$\begin{aligned} \frac{d[A]}{dt} &= k_{-1}[AB] - k_1[A][B] \\ \frac{d[B]}{dt} &= k_{-1}[AB] - k_1[A][B] \end{aligned}$$

with $[A]$, $[B]$ and $[AB]$ as continuous variables.

A model can also be constructed to capture the exact number of molecules, which fluctuate around the expected average number. This is achieved by assuming that each reaction take place with a certain probability at a certain time, hence the processes are discrete and not continuous. Typically the stochastic description starts with the master equation, which describes the probability for the system to be in all possible states, as a function of time (van Kampen, 1992). For simpler systems, the master equation, or approximations of the master equation, might be possible to solve analytically or numerically (Elf *et al.*, 2003; Werner, 2004).

However, the most common approach to model biological systems stochastically is to use the Gillespie algorithm, which gives an exact simulated solution to the master equation (Gillespie, 1977). In the Gillespie algorithm, one computes the reaction *propensities*, a_i, for each possible reaction i. The reaction propensity is the probability, per time unit, that the reaction will occur, and is directly correlated with the deterministic rate constant. To determine which reaction in the model to occur, μ_i, and when, τ, these two parameters have to be chosen from the distribution $P(\mu, \tau)$:

$$P(\mu,\tau) = \begin{cases} a_\mu e^{-a_0\tau} & if\ 0 \leqslant \tau < \infty \\ 0 & \text{otherwise} \end{cases}$$

A stochastic model is especially important for systems where molecules are present at a very low level, where noise in production and decay plays a considerable role. Recent research points strongly towards gene expression as a stochastic process, where the promoter transcribes genes in bursts, rather than in a constant flow ones it is active (Kaufmann and van Oudenaarden, 2007; Raser, 2005; Blake *et al.*, 2003). Due to this, many models of molecular interactions and gene expression might be relevant to model stochastically. But the model approach taken depend on the experimental data available and the purpose of the study. For systems where one expects eventual stochastic effects to be small, the less computationally demanding deterministic study is just as informative as the stochastic (Andrews and Arkin, 2006).

Chapter 4

EPSTEIN-BARR VIRUS

The Epstein-Barr virus (EBV) was discovered in 1964 by Anthony Epstein's group studying cell lines cultured from Burkitts lymphoma tumours in Africa. The virus belongs to the gamma herpes family, and although it was first found in cells associated with Burkitts lymphoma, it is now known to be very widespread. EBV infects over 90 % of the human population and besides from the association with Burkitt's lymphoma other cancer types are also correlated with EBV infections.

The most important feature of EBV is that it has a unique ability to transform resting B-cells into permanent latently infected lymphoblastoid cell lines, *i.e.* immortalized cells (Thorley-Laswon and Gross, 2004). The virus manages to hijack the cellular pathways of its host cell and gain control over the proliferation process. The mechanisms behind this are not completely understood, although EBV has been shown to up-regulate various kinds of cellular genes, like receptors, growth factors and adhesion molecules, that are involved in important cellular pathways (Bornkamm and Hammerschmidt, 2001).

Like all herpes viruses, EBV can persist latently in infected host cells. The remarkable property of EBV is however that it can enter four different latency programs, characterized by different gene expressions. This chapter aims to give an overview of the virus life-cycle, including description of the latency programs and viral genes, as well as associated malignancies. Finally the viral genetic switch which is the topic of our research will be described.

4.1. The virus life-cycle

EBV is spread through saliva and primary infection is believed to occur in the the epithelial cells in the oropharynx, the throat part just behind the mouth. From there it can spread to circulating naive B-cells in close vicinity to these infected cells. Virus shedding into the saliva mostly occurs during the first infection, but can be triggered even later on during the latent stable infection

period (Amon and Farrell, 2005). The primary infection often occurs in infancy in less developed countries, and will then most likely be asymptomatic. In more developed areas of the world infection is often delayed until adolescence and can then result in infectious mononucleosis, also known as the "kissing disease".

When B-cells are infected, the viral proteins induce cell proliferation, causing the immune system to react. Around 10 % of the B-cells carry the virus during the acute infection, a number which is rapidly reduced due to the immune system response (Thompson and Kurzrock, 2004). But in order to be preserved in the host, the virus has to lay latent in a reservoir, escaping the immune system. For EBV this is achieved by infected B-cells passing through the germinal center, forming memory B-cells (Amon and Farrell, 2005). This is a natural process in creating memory B-cells, and it is not clear exactly how the EBV infected cells manages to use this mechanism for its own benefit (Bornkamm and Hammerschmidt, 2001).

In the pool of memory B-cells, the viral gene expression is very restricted, to ensure that the cells remain unnoticed by the immune system. The virus can however shift between its four different latent programs and thereby somewhat regulate the cellular phenotype, including transforming a resting cell to a proliferating cell, or *vice versa*. The lytic pathway can get triggered when memory cells differentiate further due to an antigen stimuli (Amon and Farrell, 2005). Figure 4.1 gives a schematic illustration of the viral life cycle.

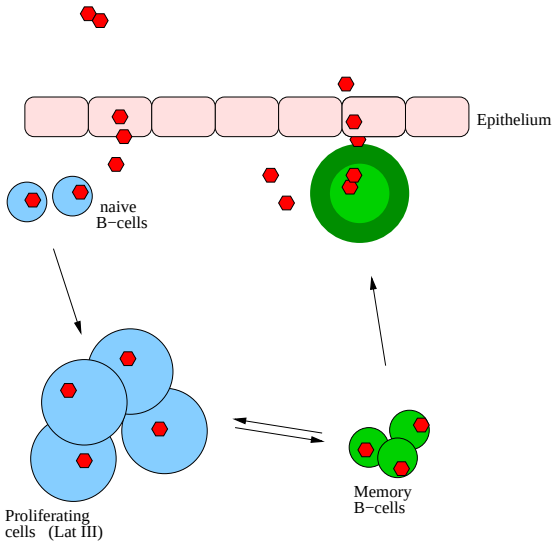


Figure 4.1. *Schematic illustration of the EBV life cycle. The virus is spread through saliva and enters the body through epithelial cells in the mouth. From there it can spread to circulating naive B-cells and stimulate them to enter the growth program. Eventually the virus transforms the B-cells into memory B-cells, which circulate in the blood system. As response to antigen stimuli the virus can switch the cells to produce new viruses that are shed in the saliva.*

4.2. Malignancies

Even though a majority of the worlds population harbors EBV, most people have no symptoms except possibly during the first acute infection. But EBV was discovered in cell lines of Burkitts lymphoma, and has since then been strongly linked to other lymphomas as well as carcinomas.

Burkitt's Lymphoma (BL) is a very aggressive tumour appearing in the jaw and exists in two forms; endemic and sporadic. Endemic BL is somewhat geographically restricted and occurs primarily in Papua New Guinea and equatorial Africa. It is believed that the malaria infection together with EBV stimulate the B cell proliferation (Young and Murray, 2003; Baumforth *et al.*, 1999). The sporadic BL is a worldwide lymphoma type but occurs much less frequently than the endemic (Young and Murray, 2003). Burkitts lymphoma is also observed in HIV positive patients (Young and Murray, 2003).

Another lymphoma linked to EBV is the Hodgkin's disease (HD), characterized by the presence of Reed-Sternberg cells, giant multi-nuclei tumour cells derived from B cells. The role of EBV in the development of HD is still not fully understood, but there is a bimodal infection distribution over age, where older patients and boys under 10 seem to have a higher association with EBV (Young and Murray, 2003).

Another malignancy strongly associated with EBV infection is Nasopharyngeal carcinoma (NPC), a tumour affecting the epithelial cells in the nasopharynx, the upper part of the throat behind the nose. NPC is common in China and south-east Asia, with 20-30 cases a year per 100,000 people (Young and Murray, 2003). Well known is also that central nervous system lymphomas in AIDS patients almost always contain EBV, and that nearly all post-transplant lymphomas are associated with EBV (Thompson and Kurzrock, 2004). There are also studies linking EBV with certain breast cancer types, although this is still under dispute (Thompson and Kurzrock, 2004; Baumforth *et al.*, 1999). In summary, EBV is involved in several serious malignancies and therefore serves as a highly interesting model system in tumour biology.

4.3. EBV genes and latencies

The Epstein-Barr virus was the first human virus to have its genome completely sequenced in 1984 (Baer *et al.*, 1984; Young and Rickinson, 2004). The genome is around 170 kbp long, double stranded and is maintained in the cells as circular episomes. It contains 84 open reading frames and until now 12 latent genes are known to be expressed during the different types of latencies, out of which nine are translated into proteins. The proteins produced include six nuclear antigens, EBNA 1-6, and three membrane proteins LMP1, LMP2A and LMP2B. The remaining genes expressed are RNAs named EBER1, EBER2 and BART/BARF0. The expression pattern vary between the four different latency programs, named latency 0, I, II and III (for review of latencies and viral proteins see Klein *et al.* (2007); Bornkamm and Hammerschmidt (2001)).

Latency 0 has the most restricted expression pattern, with LMP2A and possibly the EBERs being expressed. This restricted latency form is found in memory B-cells in healthy individuals (Bornkamm and Hammerschmidt, 2001). The latency I program expresses the nuclear antigen EBNA-1 together with the EBERs, an expression pattern found in BL cell lines as well as memory B cells in healthy individuals (Rowe *et al.*, 1987). In latency II, found in both HD and NCP cells, EBNA-1 is produced together with all membrane proteins and the RNAs. The most extensive gene expression is within latency III cells, where all 12 latent genes are transcribed. This latency program is also referred to as the growth program, and is present in AIDS-related lymphomas as well as lymphoblastoid cell lines (LCL), cells with unlimited proliferation.

The focus of the research presented in this thesis, is on EBNA-1 transcriptional regulation, and EBNA-1 is presented in more detail in the next section. but it is important to mention the function of at least a few other viral proteins here. First of all, EBNA-1 is not the only transcription factor of the EBV gene products. Also EBNA-2 and EBNA-3 has DNA-binding properties and are involved in regulating viral genes (Bornkamm and Hammerschmidt, 2001). Secondly, the LMP proteins are the most direct candidates for the oncogenic features of the virus. LMP1 can up-regulate anti-apoptotic genes, and provide growth and differentiation signals to the cell. Also LMP2A induces expression of genes that activates cell-cycle and hinders apoptosis (Young and Murray, 2003).

4.3.1. EBNA-1

The nuclear protein EBNA-1 is a DNA binding protein with vital functions for the virus. It initiates viral replication and ensures viral episome partitioning during cell division, as well as functions as a transcriptional regulator. EBNA-1 consists of 641 aa and is the most studied EBV protein, with its DNA-binding domain and dimerization region crystallized (Barwell *et al.*, 1995). The protein forms very stable dimers in solution, its degradation hampered by a Gly-Ala repeat domain that inhibits proteosomal degradation in the cell (Levitskaya *et al.*, 1997). The half-time is estimated to be at least 36-48 hours (Davenport and Pagano, 1999).

EBNA-1 binds DNA as a dimer to a 16 base-pair long palindromic sequence 'G(A/G)TATCAT-ATGCTA(C/T)C' (Frappier and O'Donnell, 1991b). Three loci on the viral genome are known to bind EBNA-1; two in the origin of replication, upstream of the C promoter (see 4.4.1), and one downstream of the Q promoter (see 4.4.2). In the origin of replication, the first EBNA-1 binding locus is called the family of repeats (FR) and consists of 20 consecutive binding sites. The second locus is the Dyad Symmetry (DS), which bind four EBNA-1 dimers (Summers *et al.*, 1996). In the Q promoter region there are two identified binding sites. The binding affinity of EBNA-1 to these three loci differ, due to minor sequence variations (Jones *et al.*, 1989). EBNA-1 binds strongest to sites in FR, with intermediate strength to the DS and the weakest to the Q promoter sites. EBNA-1 binding to DS results in assembly of the replication machinery, while binding to the other loci is known to regulate promoter activities (for review on EBNA-1 see Leight and Sugden (2000)).

Of interest is also that EBNA-1 might be involved in controlling cellular genes directly, since it has the ability to destabilize histones and has been shown to activate transcription from promoters integrated in the cellular genome (Avolio-Hunter *et al.*, 2001; Kennedy and Sugden, 2003). It is also known that EBNA-1 mediates looping between the FR region and DS (Frappier and O'Donnell, 1991a)

4.3.2. Oct-2

Although this chapter concerns the Epstein-Barr virus, there is one human protein that is relevant in the viral gene regulation and therefore should be discussed in this context. The human transcription factor Oct-2 is specific for B cells and some neuronal cells, and belongs to the POU domain family of proteins (Kemler and Schaffner, 1990). The POU domain enables DNA specific binding, and contains two sub-domains, the homoedomain and the specific domain, held together with a flexible linker (Sturm and Herr, 1988). The Oct proteins are named after the octamer motif to which they bind, 'ATGCAAAT'. This motif is found to regulate the crucial expression of immunoglobulins in B-cells as well as histone genes (Kemler and Schaffner, 1990). Oct-2 binds the single octamer site as monomer, but can dimerize and bind to longer palindromic sequences (Botquin *et al.*, 1998; Tomilin *et al.*, 2004).

Oct-2 is known as an activator of immunoglobulin genes in B-cells, but can also function as a repressor, depending on context and cofactors involved (Lillycrop *et al.*, 1991; Malin *et al.*, 2005). Highly relevant for the Epstein-Barr virus is that Oct-2 can regulate transcription from the C promoter. Oct-2 on its own has been shown to activate Cp transcription, while in complex with the cofactor Groucho/TLE it represses the promoter (Almqvist *et al.*, in press, 2005). *In vivo* essays further established that Oct-2 with Groucho/TLE bind to the family of repeats, competing with EBNA-1 (Zou *et al.*, 2006).

4.4. EBV promoters

The Epstein-Barr viral genome contains three latent promoters that governs EBNA production; the W, C and Q promoters (Wp,Cp and Qp). Initially after infection, the viral genes are transcribed under control of the W promoter, driving production of EBNA-2 and EBNA-5 (Nilsson *et al.*, 2001). Within 36 hours, gene expression control is switched to the C promoter, a switch induced by EBNA-2 (Woisetschlaeger *et al.*, 1991). Cp activity results in bicistronic transcripts encoding all six EBNA proteins, and Cp is the promoter active during the latency III program (Bodescot *et al.*, 1987). In the more restricted latencies, Cp and Wp are down-regulated by transcription factors and become hyper-methylated and thereby silenced (Salamon *et al.*, 2001). Production of EBNA-1, which is necessary for the EBV plasmid partitioning and replication, is then governed by the Q promoter (Qp) (Zetterberg *et al.*, 1999).

Transcripts for the latent membrane proteins and the lytic proteins are controlled through separate promoters. Our research focuses on the switch between the C and Q promoters, and therefore they will be described in more detail in

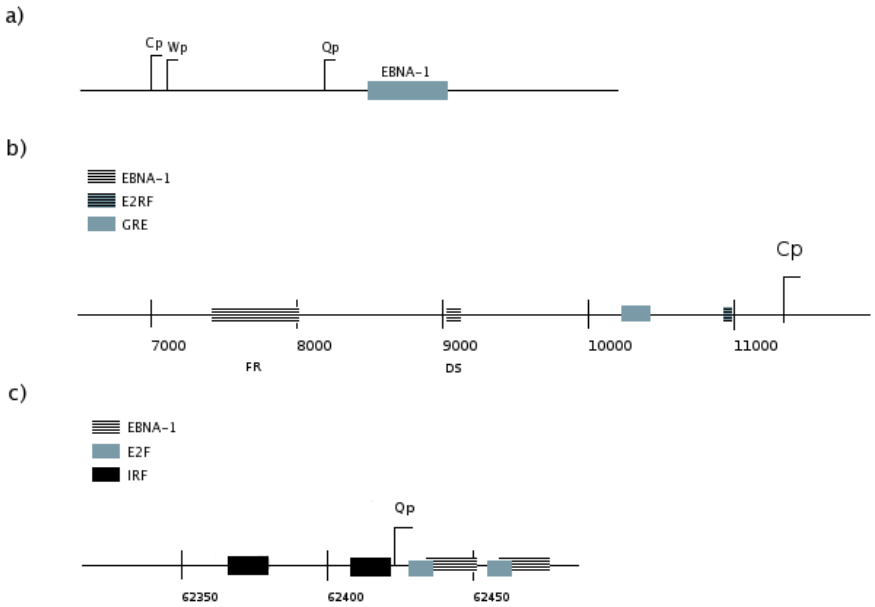


Figure 4.2. *Illustration over the latent promoter architecture in EBV. a) The relative position of Wp,Cp and Qp on the viral genome, with the EBNA-1 exon positioned downstream of the Qp initiation site. b) More detailed illustration of the Cp region, with the two EBNA-1 binding loci positioned; family of repeat (FR) and dyad symmetry (DS), together with the glucocorticoid-responsive element (GRE) and the EBNA-2 responsive enhancer element (E2RF). c) More detailed illustration of the Qp region, with the three different regulatory sites; EBNA-1, E2F and IRF.*

the following sections. The positions of the three latent promoters is illustrated in fig 4.2 a.

4.4.1. The C promoter

The C promoter region has been thoroughly studied experimentally, and many transcriptional elements have therefore successfully been identified. The most prominent regulatory sequence of Cp is the family of repeats (FR), located approximately 3000 bp upstream of the transcription start site. FR has long been known to consist of 20 repeated binding sites for EBNA-1, and EBNA-1 binding is essential for Cp activity (Reisman and Sugden, 1986). For full promoter activity, at least 7-8 bound EBNA-1 are required (Zetterberg *et al.*, 2004; Wysokenski and Yates, 1989). Recently, also binding sites for the human transcription factor Oct-2 was found in FR, interspersed between the EBNA-1 sites (Almqvist *et al.*, in press). This is of great interest since it indicates a more complex regulation of these viral proteins than earlier believed, a regulation also involving specialized

human transcription factors. The sequences to which EBNA-1 and Oct-2 have been found to bind are listed in table 4.1. There are five different EBNA-1 sequences and six different octamer sites, of which the best Oct-binding sites are preferably located next to the worst EBNA-1 binding sites (Zou *et al.*, 2006).

Other regulating elements found in the C promoter includes a glucocorticiod-responsive element (GRE), an EBNA-2 responsive element and binding sites for Egr, Sp1 and NF-Y transcription factors (Sung *et al.*, 1991; Ling *et al.*, 1993; Kupfer and Summers, 1990; Nilsson *et al.*, 2001). Deletion analysis has however revealed that the GRE is not necessary for promoter activity, and likewise that the EBNA-2 enhancer is not sufficient for activation (Puglielli *et al.*, 1996). Moreover, NF-Y binding sites can be found in 30 % of all eukaryotic promoters, and together with the fact that NF-Y and Sp1 are ubiquitously expressed proteins, these elements seem to have a more general function in transcriptional control. This leaves EBNA-1 as the main activating factor of the C promoter, together with the possibility of Oct-2 acting as an inhibitor. For illustration of the Q promoter see fig. 4.2 b.

Table 4.1. *Table of the FR sequences found to bind EBNA-1 and Oct-2. (Sequence data taken from (Zou* et al.*, 2006).)*

Nr		EBNA-1 site		octamer site
1)	GAG	GATAGCATATGCTACC	CGG	ATACAGAT
2)	TAG	GATAGCATATACTACC	CAG	ATATAGAT
3)	TAG	GATAGCATATGCTACC	CAG	ATATAGAT
4)	TAG	GATAGCCTATGCTACC	CAG	ATATAAAT
5)	TAG	GATAGCATATACTACC	CAG	ATATAGAT
6)	TAG	GATAGCATATGCTACC	CAG	ATATAGAT
7)	TAG	GATAGCCTATGCTACC	CAG	ATATAGAT
8)	TAG	GATAGCATATGCTACC	CAG	ATATAGAT
9)	TAG	GATAGCATATGCTATC	CAG	ATAT- - -T
10)	TGG	G-TAGTATATGCTACC	CAG	ATATAAAT
11)	TAG	GATAGCATATACTACC	CTA	ATCTCTAT
12)	TAG	GATAGCATATGCTACC	CGG	ATACAGAT
13)	TAG	GATAGCATATACTACC	CGG	ATATAGAT
14)	TAG	GATAGCATATGCTACC	CGG	ATATAGAT
15)	TAG	GATAGCCTATGCTACC	CAG	ATATAAAT
16)	TAG	GATAGCATATACTACC	CAG	ATATAGAT
17)	TAG	GATAGCATATGCTACC	CGG	ATATAGAT
18)	TAG	GATAGCCTATGCTACC	CAG	ATATAGAT
19)	TAG	GATAGCATATGCTATC	CAG	ATAT- - -T
20)	TGG	G-TAGTATATGCTACC	CAT	GGCAACAT

4.4.2. The Q promoter

The Q promoter is believed to be a housekeeping, hypo-methylated TATA-free promoter (Schaefer *et al.*, 1995; Salamon *et al.*, 2001). Three types of binding sites have been located in the Qp region; sites for E2F, EBNA-1 and IRF factors. The two sites binding E2F are interspersed with two EBNA-1 sites, and E2F has been shown to compete with EBNA-1 and activate Qp transcription (Sung *et al.*, 1994). These results have however been questioned, instead suggesting a role for E2F as both an inhibitor and activator, depending on cell type (Ruf and Sample, 1999). IRF2 binding upstream of the transcription initiation site, lead to activation of Qp (Nonkwelo *et al.*, 1997b). But since IRF2 is a constitutively express protein, its function is believed to be of a general character, guiding the transcriptional machinery to TATA-free promoters (Nonkwelo *et al.*, 1997a; Schaefer *et al.*, 1997a).

There is however no doubt that EBNA-1 is involved in regulating Qp activity and thereby its own production. The binding affinity of EBNA-1 to its Qp sites is 14 times lower than for the sites in FR (Ambinder *et al.*, 1990), and binding inhibits Qp activity (Sample *et al.*, 1992; Schaefer *et al.*, 1995). EBNA-1 is thereby a negative auto-regulator in resting cell types. To block transcription it is sufficient with one bound EBNA-1 protein (Schaefer *et al.*, 1997b). For illustration of the C promoter see fig. 4.2 c.

4.5. Switch between promoter activity: Cp/Qp

That the Epstein-Barr virus is associated with various lymphoma types is by now well established. The gene expression patterns for the different latency programs are also more or less identified, with restricted gene expression in the resting cells and a more extensive protein production in the growing cells. But the virus' oncogenic capabilities, and how it manages to control the cells growth program, thereby switching between programs, is still an important question to answer.

Although many of the viral promoters have been investigated for over a decade there are missing links, and no complete description of the promoter control mechanisms exist. One especially important point is to understand how the growth program is controlled, *i.e.* how the C promoter is turned on and off. If this switch can be understood, it might be controllable, meaning an opportunity to disturb the virus life-cycle with therapeutic aim.

One hypothesis on how the Cp/Qp switch in EBV-infected cells is governed, is presented in fig 4.3. In a cell with a high level of Oct-2-Grg/TLE present, these will be bound to FR, inhibiting transcription from Cp and the EBNA-1 present in the cell are produced from Qp. This scenario corresponds then to a resting cell, latency I. However, if Oct-2 levels drop EBNA-1 can instead bind FR and induce activity of Cp, thereby switching the cell into a proliferating state, latency III. The Qp binding sites for EBNA-1 will afterwards also be bound by EBNA-1, silencing Qp. For this state to reverse, there is a need for increased Oct-2 levels once more, since EBNA-1 is a very stable protein. This regulating hypothesis is

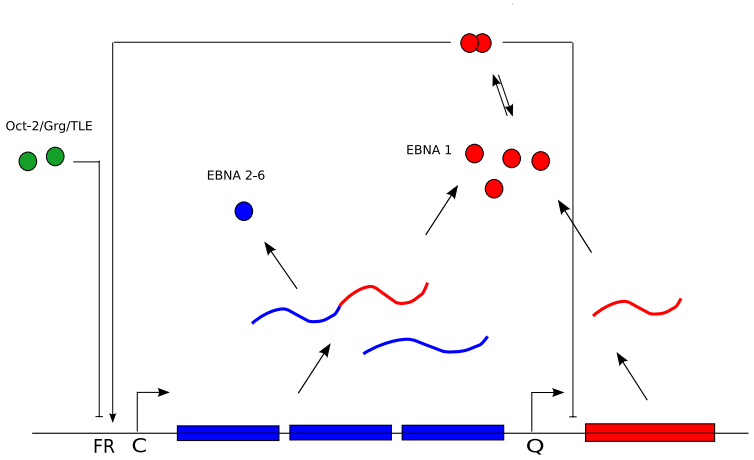


Figure 4.3. *Illustration over the Cp/Qp switch hypothesis studied. EBNA-1 proteins are produced from transcript originating either from Qp or Cp, depending on the cellular state. Oct-2 proteins are the human transcription factors, regulated by external signals. Oct-2 and EBNA-1 competes for binding to the 20 sites in FR, where Oct-2 bound, in complex with Groucho/TLE, inhibits transcription from Cp while EBNA-1 activates the promoter. EBNA-1 present in high levels while also negatively regulate Qp activity.*

what has been the focus of our research concerning EBV.

Chapter 5

BACTERIOPHAGE λ

One of the most well studied genetic systems, both experimentally and theoretically, is the bacteriophage λ. The virus was discovered in the early 50ies, when *Escherichia Coli* strains were subjected to UV-radiation (Lederberg and Lederberg, 1953; Lwoff, 1953). From this accidental discovery, the λ phage infection became a model system for many researcher in the studies of the phage infected bacteria (Gotesman and Weisberg, 2004). When the first theoretical gene regulation models appeared, they were constructed to describe the promoter activities of the λ phage (Ackers *et al.*, 1982; Shea and Ackers, 1985). Still today the phage is an important model system for genetic logic control, with it's relative small genome and yet interesting and complex regulatory functions.

5.1. The viral life-cycle

Infection of the bacteria by λ leads to either a lysogenic or a lytic response of the cell. In the lysogenic response all but one viral gene, the CI repressor, are turned off and the viral genome is integrated into the bacterial genome. The viral genome is then replicated with the cells genome and the virus is passively spread when the cell grows and divides. The lytic response program results in the onset of program to produce new viral particles, eventually causing the cell to burst and release new viruses into the surrounding. The central regulatory protein in the lytic pathway is the Cro repressor. Which path that is chosen depends on the conditions in the bacteria. In the case of starvation or high multiplicity of infection (many phages taken in by the cell), the lysogenic response is favored (Herskowitz and Hagen, 1980; Ptashne, 1992).

In case of UV irradiation or other DNA damage reactions, lysogens are activated into production of new phage particles, enabling the phage to abandon the host cell that is being destroyed. Under UV irradiation, the RecA protein cleaves the CI repressor, thereby switching the cell from lysogenic to lytic state (Roberts

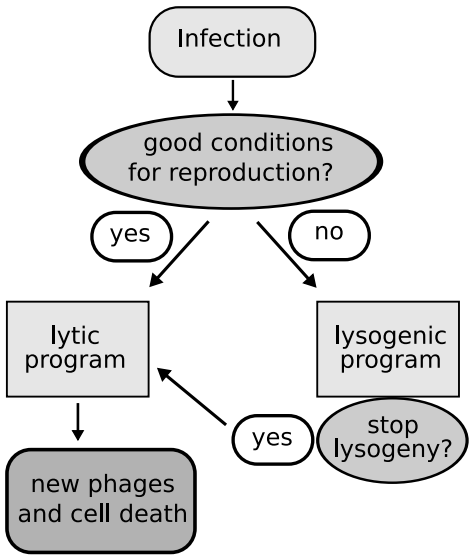


Figure 5.1. *Schematic drawing of the life/cycle of bacteriophage lambda. After infection, the virus enters one of the two possible programs; the lysogenic or the lytic. For cells that form lysogens, these can at later stages be activated to enter the lytic pathway and start producing new phage particles. (Courtesy of Erik Aurell.)*

and Roberts, 1975). Figure 5.1 schematically summarizes the life-cycle of the bacteriophage.

5.2. The viral promoters and the genetic switch

The λ phage genome has seven promoters, which are active in different stages of infection and depending on which pathway that is chosen (Herskowitz and Hagen, 1980). After infection, the right promoter (PR) and the left promoter (PL) are active, producing the lytic repressor Cro and the anti-termination protein N. N binds to termination sites downstream of PR and PL and allows transcription to continue past them, transcribing a number of different genes necessary for both the lysogenic (CII and CIII) and the lytic pathway (O and P) (Herskowitz and Hagen, 1980). Then, depending on the environmental conditions affecting the CII levels, either CI production is activated and the phage enters the lysogenic pathway, or Cro and its downstream genes accumulates in the cell, and the lytic path is chosen.

There are two essential operator regions; OL controlling PL, and OR controlling PR and PRM, the promoter for Cro and CI production. The essential switch between the two pathways lies at the right operator, where there are three binding sites for both Cro and CI, see figure 5.2. These sites have different affinities for the two repressors, where CI binds strongest to OR1, then OR2 and finally

OR3. CI dimers bound to two adjacent sites can bind cooperatively. CI bound to OR1 and/or OR2 represses transcription from PR, hence inhibits production of the lytic proteins. At the same time cI can stimulate transcription from PRM by facilitating the RNA polymerase to bind in (Meyer and Ptashne, 1980). At high concentrations of CI it also binds to OR3, and thereby down-regulates its own production. The relative affinities for Cro to OR are exactly the opposite, where Cro binds strongest to OR3. Cro thereby blocks PRM transcription while allowing PR to continue transcribing. At high Cro concentrations also OR2 and OR1 will be bound and PR silenced.

The left operator also has three binding sites for CI and Cro, with a similar affinity grading, where both cro and cI can inhibit transcription. What has been discovered in the recent decade is that OL and OR can loop together, and CI bound at the two operator regions form octamer and tetramer structures (Revet *et al.*, 1999). The looped structure still allows for transcription from PRM, as long as OR3 is free. This looped structure increases repression of the lytic PR and stabilizes the lysogenic state (Dodd *et al.*, 2001; Revet *et al.*, 1999). Figure 5.2 WT displays the control circuit for the wild-type phage λ.

5.3. Lambda virus modified with LacR

Since the lambda phage is well studied and reasonably simple to work with, there has been various studies done with genetically modified λ genomes. One of the latest studies were done by Little and Atsumi in 2006, where they modified the lambda genome by swapping the lytic control protein Cro, with the Lac repressor, LacR (Atsumi and Little, 2006). Their goal was to test the modularity of the lambda circuit, *i.e.* to determine if specific vital properties of the λ phage would be affected by an exchange in lytic repressor protein.

The modified genetic λ-lac constructs concerned the three promoters that are directly involved in the switch between the lysogenic and the lytic state; PR, PRM and PL. First the gene coding for the Cro protein was replaced by the Lac repressor gene, lacI. In order for the PR product to maintain its repressor function, lac operator sites, lacO, were introduced at PL and PR (Circuit A, fig 5.2). The lacO sites are positioned after the transcription start site, leaving the original operator sites intact. Some mutants were also constructed where the OR3 operator was replaced by the lacO, allowing lacI to inhibit PRM, but cutting the negative auto-regulation of cI on PRM (Circuit B, fig 5.2).

Due to uncertainties on how tight the new feed-back regulations should be in order to create a functional phage, a large library of mutants was created in the experimental study. The setup was that each lacO could be one out of five variants (A-E), and the Shine-Dalgarno sequence could be one out of six variants (A-F), in total summing up to 900 mutants. The five different operator sites had varying affinities for LacR, with A being the wild-type operator with the best affinity, and B to E having gradually worse affinities. The altered Shine-Dalgarno sequences differed in their translational efficiencies, with A being the most effective.

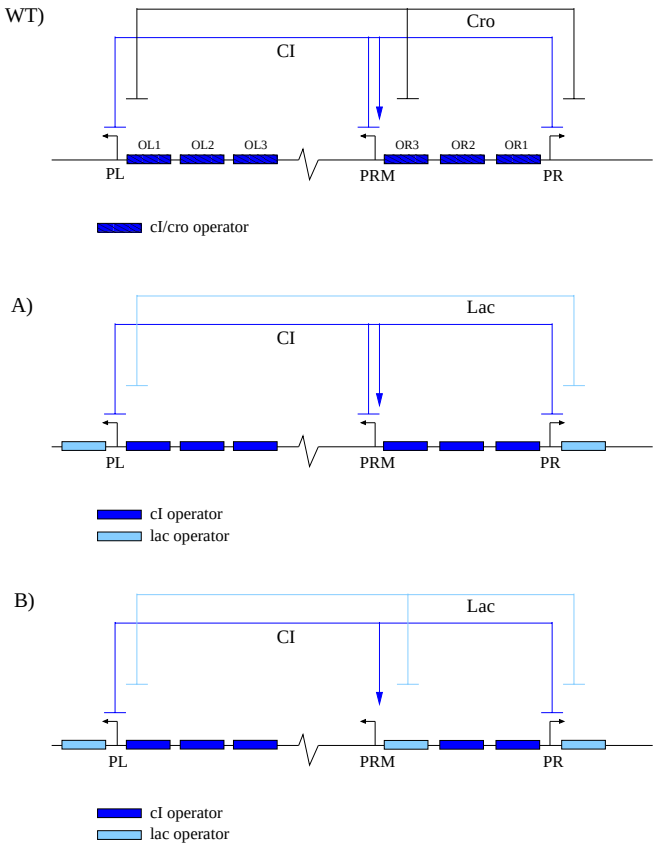


Figure 5.2. *The wild-type circuit (WT) and the two different genetic circuits that were created by Little & Atsumi (A and B). A) The genetic construct when all three OR sites are intact and only the lac repressor sites have been added (circuit A). B) The genetic construct were the OR3 sites has been switch to a lac operator site (circuit B).*

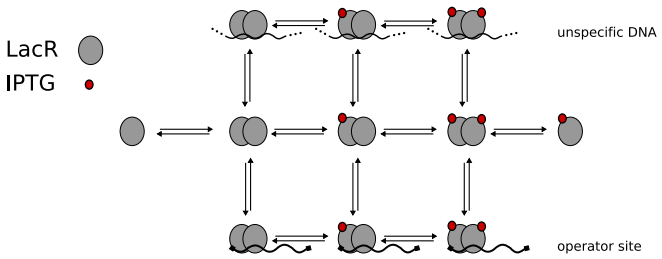


Figure 5.3. *Illustration over all LacR and IPTG interactions possible for the dimeric mutant used in the study by Little & Atsumi. The Lac repressor forms dimers, where each monomer can bind the inducer molecule. The dimer can then bind both to operator sites and nonspecifically to the genome.*

5.3.1. Lac repressor

The Lac repressor is a repressor involved in the metabolic regulation for lactose uptake in bacteria. The first theory about metabolic regulation in bacteria came from Jacob and Monod in 1961, and the Lac repressor was isolated five years later (Jacob and Monod, 1961; Gilbert and Muller-Hill, 1966). Since the discovery it has been widely used in experimental studies of regulatory control, also in eukaryotic systems (Hu and Davidson, 1987). The Lac repressor is an inducible repressor, meaning that it binds ligands that affect its binding to DNA. *In vivo* the inducer is allolactose, but more commonly used in experimental setups is IPTG (Isopropyl β-D-1-thiogalactopyranoside), a very similar sugar molecule. The inducers reduce the binding affinity of the repressor for its specific operator sites by changing the protein conformation. The non-specific binding is however not affected.

The repressor naturally forms tetramers, and each monomer within the structure can bind one ligand. It has a very high affinity for its natural operator site, which however can be reduced over 100-fold for saturated amounts of bound inducer (O'Gorman *et al.*, 1980). The binding of inducer molecules to the repressor are also affected by the presence of operator. For operator bound repressors, the IPTG affinity is reduced 20-fold but the binding becomes cooperative (O'Gorman *et al.*, 1980; Dunaway *et al.*, 1980).

In the Little & Atsumi study, the Lac repressor used was a dimeric mutant. A tetrameric mutant would possibly form looped structures between OL and OR, affecting the regulation. The tetrameric Lac repressor is very stable while the dimeric protein has a half-life of about 20 minutes in the cell (Platt *et al.*, 1970). The DNA binding properties of the dimeric mutant is however the same as the the wild-type repressor's (Chen and Matthews, 1992). Figure 5.3 displays all the kinetic reactions present in the λ-Lac system.

Chapter 6

RESULTS AND DISCUSSION

6.1. Paper I - Switch between Cp and Qp

We have investigated the hypothesis presented in chapter 4, regarding Oct-2 and EBNA-1 as the major regulatory factors involved in the switch between the C promoter and the Q promoter activity in EBV-infected cells. A deterministic thermodynamic model was built, based on available experimental data including the binding affinities for EBNA-1 and Oct-2 to the involved binding sites in Cp and Qp, as well as dimerization constants and steady state number of EBNA-1 molecules in latency III. EBNA-1 and Oct-2 were assumed to compete for the 20 binding sites present in FR, meaning each bound protein blocks binding of the other species to the closest neighboring site.

The goal was to test whether this hypothesis could describe observed promoter activities, and to test the conditions for stability and transitions between latency states.

Stability of latency states

Looking at various levels of Oct-2 in the system, the model displays regions of mono-stability, with either latency I or latency III as the stable state, or bi-stability. For low Oct-2 levels, only latency III is stable since the EBNA-1 produced from the Q promoter will bind to FR and stimulate Cp transcription. The positive feedback of EBNA-1 therefore results in only latency III being stable. For intermediate levels of Oct-2, both latency states can exist, while elevated Oct-2 levels pushes the cell to latency I mono-stability.

In fig. 6.1 the potential landscape of the system is shown, for three different levels of Oct-2. The production potential is computed as the derivative of the transcription and dilution rates in the system, *i.e.* the net changes of EBNA-1. The promoter production rate is calibrated to fit measured levels of EBNA-1 in latency III cells, which is around 30,000 for low Oct-2 levels. As can be seen

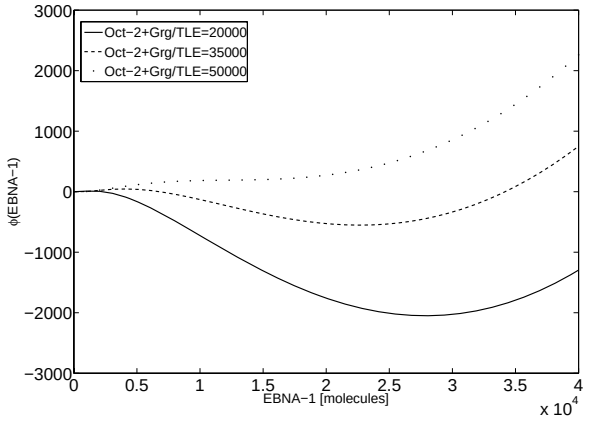


Figure 6.1. *The potential landscape of the EBV system as function of EBNA-1, for three different levels of Oct-2. For low levels of Oct-2, the landscape has two minima, one for a high EBNA-1 level (latency III) and one for a very low level (latency I). With increasing Oct-2 levels, the latency I minima diminishes and eventually disappears completely.*

in fig 6.1 (solid line), for an Oct-2 level of 20,000 or lower, there is essentially one potential minimum, at a high EBNA-1 level (latency III). For increasing Oct-2 levels (dashed and dotted lines), the latency III minimum is shifted towards lower EBNA-1 levels, and a minimum at around a few hundred EBNA-1 appears (latency I).

The stability of both states were tested for three different cell volumes and EBNA-1 dimerization constants, since these are not completely verified parameters. For most levels of Oct-2 in the cell, latency III is markedly more stable, meaning a larger change in EBNA-1 levels is needed to transit from latency III to I than in the opposite direction. This property remains for the tested parameter alterations, although the boundaries are shifted. For a smaller cellular volume, the affect of changed dimerization of EBNA-1 naturally affects the system more. Less dimerization requires higher EBNA-1 levels to shut down Qp and switch on Cp, raises the latency I level and thereby shrinks the bi-stable region.

Promoter activities

Of central importance is whether the model can reproduce the experimental available data concerning promoter activities in the two latency types. Shaefer *et al.* present a Qp/Cp activity ratio of 5-100 in latency I cells, and 0.05;0.1 in latency III cells (Schaefer *et al.*, 1997b). Another study by Zetterberg *et al.* show that in latency I cells, 76-83 % of all EBNA-1 transcripts are produced from Qp, while less than 1 % in latency III cells (Zetterberg *et al.*, 1999). Our model predictions of Qp/Cp activities correspond well to this data were we see

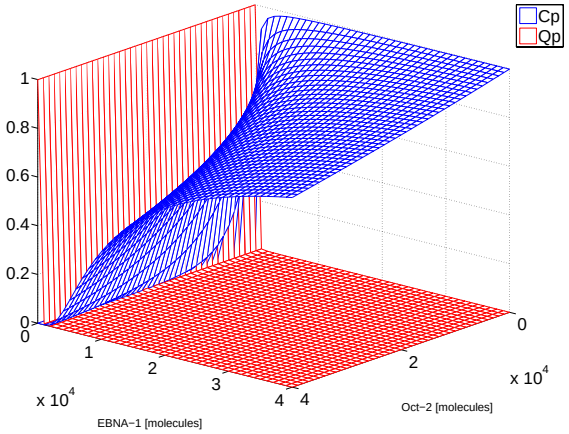


Figure 6.2. *Plot of Cp (blue) and Qp (red) probability of activation as function of EBNA-1 and Oct-2 levels in the system. Qp is only on for extremely low levels of EBNA-1 and is not directly affected by changes in Oct-2 levels. The Cp activity however is strongly dependent on Oct-2 where a higher level competes out EBNA-1 bindings at FR. The Cp activity is however always remarkably higher than Qp, and reaches 100 % activity for very low Oct-2 levels.*

a dramatic difference in the two promoter activities. The Q promoter activity is always low, with no detectable activity at all in latency III and around 1 % in the latency III states. The C promoter on the other hand has a very high activity, being 40-90 % active in the bistable latency III states, and more than 90 % active in the mono-stable latency III. In figure 6.2 the probability of activation for Qp and Cp is shown as function of EBNA-1 and Oct-2 levels.

Transitions

A final interesting point is to see what is required for the system to switch between the two different latency states. In figure 6.3 the levels of EBNA-1 and Oct-2 over time is plotted. The system starts in the resting state, latency I, with a low level of EBNA-1 and relatively low Oct-2 level. Lowering the Oct-2 just below the threshold for inducing Cp activity, switches the cell into latency III where EBNA-1 production is elevated. Within a few days, stable latency III levels of EBNA-1 is reached and in order to switch the cell back to latency I, a dramatic increase in Oct-2 levels is needed. There is yet no experimental setup that allows testing this exact scenario of Oct-2 inhibition of Cp. However, the experimental results available are by EBNA-1 RNA interference and CD40 ligand exposure, where it has been shown that proliferating cells can be switched back to resting states within 5 days (Hong *et al.*, 2006; Pokrovskaja *et al.*, 2002). Since EBNA-1 degradation is very slow, our model requires rather extreme levels of Oct-2, at least temporarily, to switch back to latency I in five days.

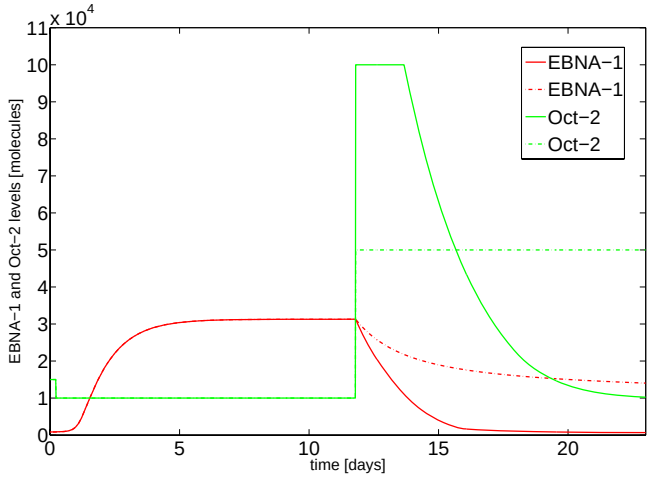


Figure 6.3. *Plot displaying the possible transitions between latency I and latency III, showing changes in molecular levels of EBNA-1 (red) and Oct-2 (green) as function of time. The system starts out with low a EBNA-1 level when a drop in Oct-2 induces Cp to become active and the EBNA-1 production increases. Within approx. five days, the EBNA-1 level has reached its latency III steady state level. To induce the system to switch back to latency I, a dramatic increase in Oct-2 is needed. The time it takes to return to the resting state depends on the fold-change. For a 10-fold increase (solid green line), the system can switch within a few days, where the limit is set by the degradation of EBNA-1. For smaller elevations in Oct-2 (dashed green line), the EBNA-1 level remains relatively high for a long period of time.*

Conclusions

The main conclusion to draw from this study is that EBNA-1 and Oct-2 as only regulating transcription factors, may well account for the observed promoter activities of Cp and Qp. We see a dramatic difference between the two latency states, with latency III being much more stable. A small decrease in Oct-2 levels, as could be produced from an externally imposed signal, rapidly induces a drift to latency III due to the high affinity of EBNA-1 to FR. Experimental observations is that resting cells do drift into proliferating cells, and not in the opposite direction.

There are still certain open question in this system, such as eventual viral feedback on the Oct-2 regulation, as well as parameters that need to be more exactly determined, such as protein levels measurements and cellular volume alterations. However, it is clear that this hypothesis is of interest to explore further in order to completely understand the proliferation regulation in EBV infected cells.

6.2. Paper II - Cooperativity mechanisms at FR

As a follow up to paper I, we investigated the cooperativity properties of transcription factor bindings to FR in the Cp region. In paper I, as simplification, effectively the number of binding sites were reduced from the total 40 to 20 sites, were each could be bound by either protein. This corresponds to each protein blocking only its closest neighboring site for the opposite transcription factor. Since the binding sites are relatively closely spaced, it might however be possible that one bound factor blocks neighboring binding on both sides. Also, DNA bound EBNA-1 can interact cooperatively on other sites on the viral genome.

The goal of paper II was to investigate how a double sided blocking event , as well as EBNA-1 cooperative bindings, would influence the effective cooperativity of the C promoter. As the measure of effective cooperativity we looked at the Hill coefficient at half saturation, for the various cases.

Blocking competition influence

To visualize the cooperative effects of blocking and/or cooperative interactions, we plot the ratio $\frac{P}{1-P}$ for different EBNA-1 concentrations, versus the local Hill coefficient. For very low concentrations, low P, the Hill coefficient approaches the limit, 8, since $P \approx [E_{free}]^8$ (see paper for discussion). For high concentration of EBNA-1, $1 - P \approx [E_{free}]^{13}$, and the local Hill coefficient therefore tends to 13 in the high concentration limit.

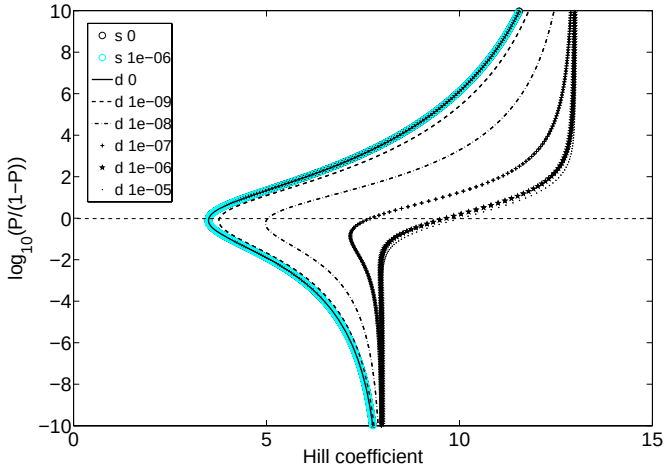


Figure 6.4. *Hill coefficient curves for the single (circles) and double blocking (black curves) model. There is no cooperative binding included here, only the impact on increased Oct-2 levels in the system. For the single blocking cell, there is no impact at all on the effective Hill coefficient, while the double blocking model goes towards a maximum of 10.5 for concentrations around* 10^{-5} *M and above.*

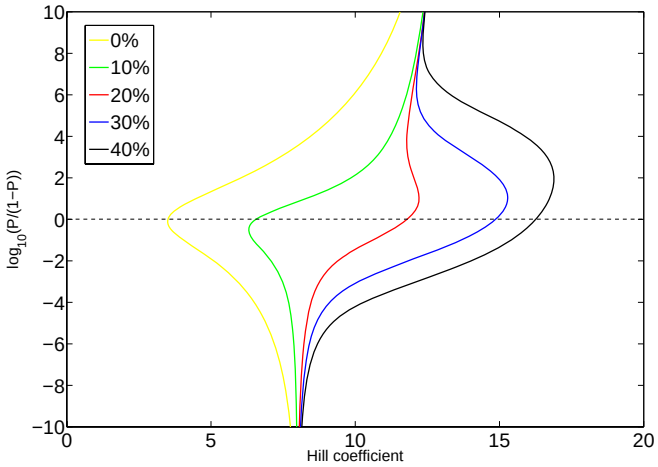


Figure 6.5. *Hill coefficient curves for changing cooperative binding strengths between EBNA-1 at FR, with no Oct-2 present. The binding energies are expressed as percentage of the DNA binding. Here, the curves for five different strengths are shown, from 0 (0 %) to -6.18 kcal/mol (40 %).*

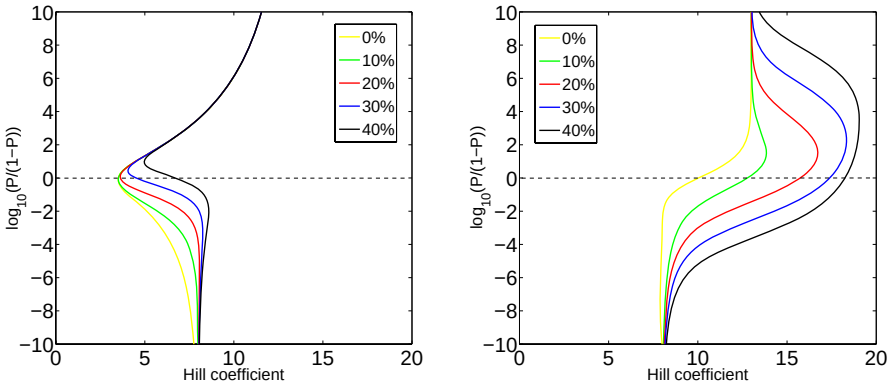
In figure 6.4, the Hill coefficient curves for the single and double blocking models, at various concentrations of Oct-2 are shown. In the single blocking, there is no change in the cooperativity, since the fractional activity scales with the Oct concentration in the same way as the Hill coefficient function. For the double blocking model on the other hand, there is no simple scaling, and the effective cooperativity is markedly increased with the Oct level. For saturating levels of Oct, all sites not occupied by EBNA-1 are filled with Oct, making the switch very sharp from off to on as soon as eight EBNA-1 get to bind.

Cooperative interactions of EBNA-1

Looking at the single blocking model without Oct-2 in the system, but with added cooperative binding, one can clearly see the effective Hill coefficient increases. In figure 6.5, the Hill coefficient curves for 0-40 % added cooperative interactions are plotted. Essentially, adding a cooperative binding simply increases the on response, which is what is seen. The stronger binding, the sharper response, until the FR is either free or bound by EBNA-1 at all 20 sites, yielding an effective Hill coefficient of 20.

Blocking and cooperative interactions

In the cell however, the Oct protein is likely to be present, which is why it is of interest to look at the Hill coefficient curves for high levels of Oct-2. For



(a) Single blocking model (b) Double blocking model.

Figure 6.6. *Hill coefficient curves for the two models of blocking. Figure a) shows the response of increased cooperative EBNA-1 bindings for the single blocking model, and b) for the double blocking. The cooperative strengths are varied from 0 (0 %) to -6.18 kcal/mol (40 %).*

the same cooperative binding strengths as in the previous figures, figure 6.6 displays the Hill coefficient curves for the single blocking model (a) and double blocking model (b). For the single blocking model, the addition of cooperative binding to not have the same impact when there are high levels of Oct-2 in the system. Instead of a 4-fold change, from 3.5 to 16, the effective Hill coefficient is merely doubled, up to 7. The double blocking model has a high effective Hill coefficient even without cooperative bindings, as already seen in 6.4. Hence the added cooperative strength further establishes the all-or-nothing scenario, much like in the single blocking model without Oct-2, with the effective Hill coefficient approaching 20.

Conclusions

The main conclusion from this study is the importance of steric hindrance in the effective cooperativity of the promoter. Essentially, in order to make the switch sharp, there is no need for cooperative interactions between EBNA-1 molecules, if the competition with Oct-2 is helped through enough steric hindrance. This is of general importance when considering models of promoters where the architecture includes many interspersed binding sites.

One hypothesis for this viral example might be based on EBNA-1 as a necessary regulator of Cp, but at the same time essential for a complex formation with DS during replication initiation. More closely spaced sites at FR will most likely result in strong cooperative interactions of EBNA-1, and possibly not allow the replication complex to form correctly. However, with the competition with Oct-2, enough steric hindrance still allows a sharp switch, and an effective promoter

activity regulation.

6.3. Paper III - Computational study of the λ-lac mutants

In the experimental study by Atsumi & Little, where a library of λ-Lac mutants were created, they identified only a few mutants that behaved like the wild-type λ phage, regarding lysogenic frequencies and induction properties. To investigate what made it possible for only a few out of the 900 created mutants to behave like the wild-type, we constructed models of the new genetic circuits. The aim was to identify the characteristic properties of the altered circuits and hopefully explain the experimental results.

Equilibria in circuit A

For circuit A, with the intact OR3 site, all variants of lacO at PR and Shine-Dalgarno essentially behave in the same way. There is no negative control of PRM other than from cI itself, while PR is inhibited by both cI and lacI. Therefore, the only possible stable fixed point is the 'lysogenic' state, where PRM is higher than PR. Figure 6.7 shows the fixed points for all mutants within this circuit. The lysogenic fixed point has a slightly higher CI level than in the wild-type circuit, were there are two negative control mechanisms acting on PRM. The LacR level is consistently low, but varies with the translational efficiency of the lac transcript. The better translation, the more Lac repressor. For the best translational efficiency, the PR lacO also affects the LacR level. However, with increased IPTG level in the cell, this small impact disappears. The PRM promoter is about 40 % active and PR is completely silent, both with and without IPTG in the system.

Equilibria in circuit B

The mutants in circuit B behave quite differently than the ones with circuit A. Figure 6.8 shows the fixed points for all these phages, and their corresponding promoter activities. One group of phages have a stable fixed point corresponding to the lytic state, with a low CI level and a very high LacR repressor level. A small group of mutants have two stable fixed points, one lytic and one lysogenic, while most of them have only the lysogenic fixed point.

Since there is always a negative feedback on PR, the phages with the 'lytic' fixed point, still have a relatively low PR activity, with maximum 25 % activity. Generally the mutants with very low Lac repressor affinity for the PR lacO have higher PR activity. The mutants with two stable fixed points, generally have one with low activity for both promoters and one with full PRM activity and a silent PR. Regarding the mutants with only the lysogenic fix points, the PRM activity for these differ widely. The most important factor for the PRM activity is the Shine-Dalgarno sequence, where low efficiency results in high PRM activities. All of these mutants have very low LacR levels, resulting in the loss of negative regulation of PRM which therefore can be fully active. This is in large contrast to the wild-type phage where cI inhibition holds down the CI level.

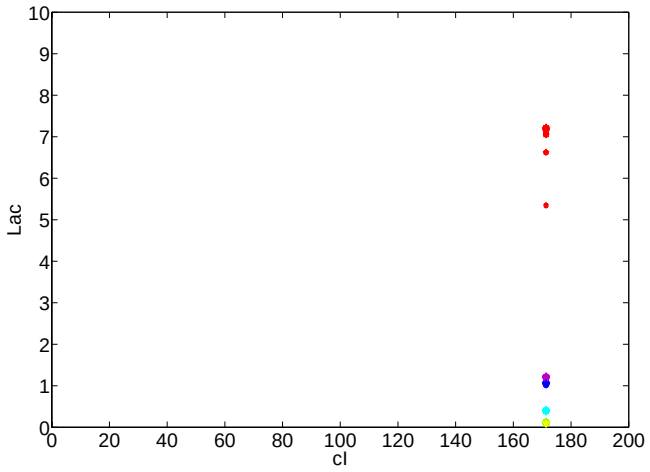
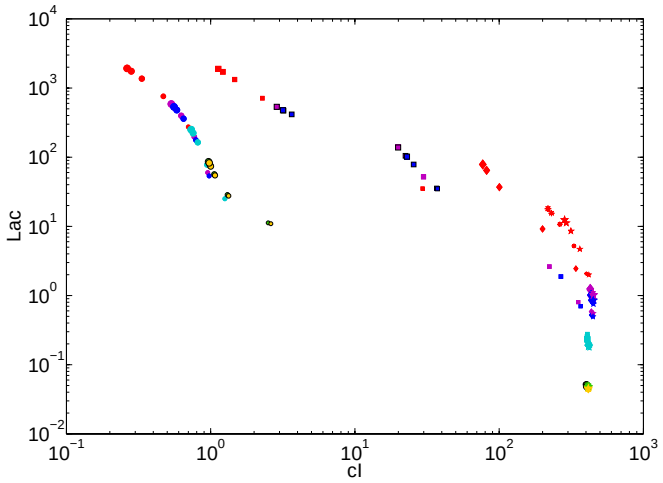
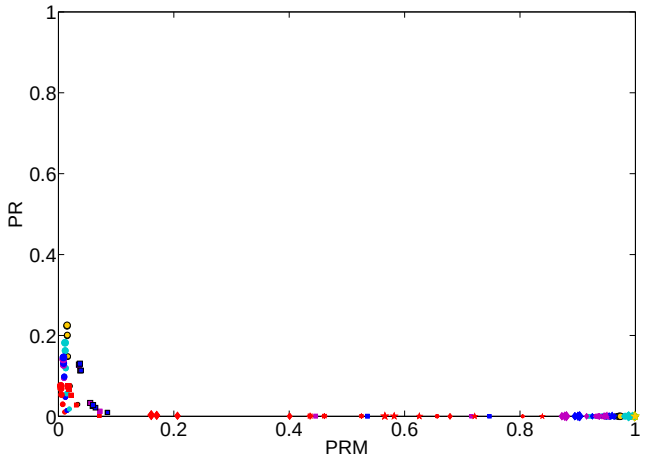


Figure 6.7. *The fixed points for the mutants with an intact OR3 operator site, circuit A.The mutants are coded with three parameters; colour, marker type and marker size. The lacO site at PL is indicated with marker type, going from the highest affinity ∘ (A) through □ (B),⋄ (C),∗ (D) and finally ⋆ (E). The PR lacO is indicated by the marker size, with the highest affinity (A) having the smallest marker, going larger with weaker affinities. The Shine-Dalgarno sequence is indicated by color, with the best efficiency (A) coded as red, going down with purple (B), blue (C), cyan (D), green (E) and finally yellow (F). All mutants of circuit A have only one stable fixed point, with a high PRM activity and zero PR activity.*



(a) Molecular levels



(b) Promoter activities

Figure 6.8. *Plots showing the stable fixed points (a) with corresponding promoter activities (b) for the 180 phages with the best PL lacO at the OR3 operator site. The mutants are coded with three parameters; colour, marker type and marker size. The lacO site at PRM is indicated with marker type, going from the highest affinity ∘ (A) through □ (B),⋄ (C),∗ (D) and finally ⋆ (E). The PR lacO is indicated by the marker size, with smallest markers for the best operator (A), then increasing marker size with decreasing affinity. The Shine-Dalgarno sequence is indicated by color, with the best efficiency (A) coded as red, then decreasing with purple(B), blue (C), cyan (D), green (E) and finally yellow (F). The mutants can be separated into three groups, either with one lytic fixed point or one lysogenic, or with two stable fixed points (marked with black edges).*

Conclusions

What is apparent from our computational study, is that the λ-lac mutants, indeed behaves very differently from the wild-type phage. They have two completely different regulatory circuits, where feed-back regulation mechanisms have been lost.

In the experimental study, a few mutants from circuit A were found to behave similarly to the wild-type phage. However, the repression of PRM in the lytic state is lost, and hence the fixed point that corresponds to the lytic state no longer exist in this circuit. Any induction into the lytic cycle therefore should originate from transitionally large PR activities, or another unknown mechanism. The lysogenic state for mutants in groups A is however quite similar to the wild-type lysogeny regarding PRM activity and CI level. But if this is all that is required for the functional phage, there is no direct reason for why not any of the mutants within circuit B, with similar lysogenic properties, should be able to form stable lysogenies. And especially why not more mutants with circuit A were experimentally isolated, since they all share the same characteristics.

In summary, the comparison between these new mutants and the wild-type is not as straightforward as presented in the experimental study, due to the distinct changes in control regulation of the central operator. Since such a few mutants were experimentally examined, our systems approach study intended to clarify the differences between the various mutants. Due to the inconsistency between the computational and experimental results, a central question that comes to mind is that if indeed only a small subset of all mutants are functional *in vivo*, should one reevaluate the central role of the right operator to control the lysogenic/lytic switch?

Chapter 7

APPENDIX

Gibbs free energy in grand canonical ensemble

As mentioned in chapter 3, the Gibbs free energy is most often used in statistical models describing biological systems. The entropy in the grand canonical ensemble, at constant pressure and temperature, is a function of the energy E, the volume, V, and the number of particles, N; $S(E, V, N)$. The probability of a the smaller system being in state s_1, with energy E_1 is then written as:

$$\begin{aligned} P(E_1, N_1) &= \frac{g(E_{tot} - E_1, V_{tot} - V_1, N_{tot} - N_1)}{\sum_i g(E_{tot} - E_i, V_{tot} - V_i, N_{tot} - N_i)} \\ &= \frac{e^{S(E_{tot}-E_1, V_{tot}-V_1, N_{tot}-N_1)}}{\sum_i e^{S(E_{tot}-E_i, V_{tot}-V_i, N_{tot}-N_i)}} \end{aligned}$$

Assuming the reservoir is much bigger than the small system, $N_{tot} >> N_1$, and that the total energy is much larger than the systems energy, $E_{tot} >> E_1$, we can expand the entropy at constant pressure and temperature:

$$\begin{aligned} P(E_1, V_1, N_1) &= \frac{e^{S(E_{tot})/k_B - E_1(\frac{\Delta S}{\Delta E_{tot}})/k_B - V_1(\frac{\Delta S}{\Delta V_{tot}})/k_B - N_1(\frac{\Delta S}{\Delta N_{tot}})/k_B + S(E_1,V_1,N_1)/k_B}}{\sum_i e^{S(E_{tot})/k_B - E_i(\frac{\Delta S}{\Delta E_{tot}})/k_B - V_i(\frac{\Delta S}{\Delta V_{tot}})/k_B - N_i(\frac{\Delta S}{\Delta N_{tot}})/k_B + S(E_i,V_i,N_i)/k_B}} \\ &= \frac{e^{S(E_{tot})/k_B - E_1/Tk_B - V_1 P/Tk_B - N_1\mu_1/Tk_B + S(E_1,V_1,N_1)/k_B}}{\sum_i e^{S(E_{tot})/k_B - E_i/Tk_B - V_i P/Tk_B - N_i\mu_i/Tk_B + S(E_i,V_i,N_i)/k_B}} \\ &= \frac{e^{(-E_1 - V_1 P - N_1\mu_1 + S(E_1,V_1,N_1)T)/Tk_B}}{\sum_i e^{(-E_i - V_i P - N_i\mu_i + S(E_i,V_i,N_i)T)/Tk_B}} \\ &= \frac{e^{-G_1/Tk_B - N_1\mu_1/Tk_B}}{\sum_i e^{-G_i/Tk_B - N_i\mu_i/Tk_B}} \end{aligned}$$

The Gibbs free energy includes the entropy term of having the setup giving state s_1. If there are different ways of achieving this state, for example if there

at a promoter are two sites that can be occupied by one molecule, both with the same binding energy, one has to add the combinatorial factor ξ:

$$P(E_1, V_1, N_1) = \frac{\xi(N_1)e^{-G_1/Tk_B}}{\sum \xi(N_i)e^{-G_i/Tk_B}} \tag{7.1}$$

Bibliography

ACKERS, G. K., D.JOHNSON, A. AND SHEA, M. A., Quantitative model for gene regulation by λ phage repressor, *Proc. Natl. Acad. Sci*, 79:1129–1133, **1982**.

ALMQVIST, J., ZOU, J., BORESTROM, C., PETTERSON, S., RYMO, L. AND ERNBERG, I., Functional interaction of Oct transcription factors with the family of repeats in Epstein-Barr virus *oriP*, *Journal of General Virology*, 86:1261–1267, **2005**.

ALMQVIST, J., ZOU, J., LINDERSON, Y., BORESTROM, C., ALTIOK, E., ZETTERBERG, H., RYMO, L., PETTERSON, S. AND ERNBERG, I., Repression of Epstein-Barr virus enhancer family of repeats mediated transcription by Oct and Grg/TLE transcriptional regulators, suggests an involvement in switching between latency programs, *Journal of General Virology*, pp. –, **in press**.

AMBINDER, R. F., SHAH, W. A., RAWLINS, D. R., HAYWARD, G. S. AND HAYWARD, S. D., Definition of the sequence requirements for binding of the EBNA-1 protein to its palindromic target sites in Epstein-Barr virus DNA, *Journal of Virology*, 64:2369–2379, **1990**.

AMON, W. AND FARRELL, P. J., Reactivation of Epstein-Barr virus from latency, *Reviews in Medical Virology*, 15:149–156, **2005**.

ANDREWS, S. S. AND ARKIN, A. P., Simulating cell biology, *Current biology*, 16:R523–R527, **2006**.

ARKIN, A., ROSS, J. AND MCADAMS, H. H., Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected Escerichia coli cells, *genetics*, 149:1633–1648, **1998**.

ARNOSTI, D. N. AND KULKARNI, M. M., Transcriptional enhancers: intelligent enhanceosomes or flexible billboards?, *Journal of Cellular Biochemistry*, 94:890–898, **2005**.

ATKINS, P., *Physical chemistry*, Oxford University Press, **1998**.

ATSUMI, S. AND LITTLE, J. W., Role of the lytic repressor in prophage induction of phage lambda as analyzed by a module-replacement approach, *PNAS*, 103(12):4558–4563, **2006**.

AURELL, E., BROWN, S., JOHANSON, J. AND SNEPPEN, K., Stability puzzles in phage lambda, *Physical review E*, 65, **2002**.

AURELL, E. AND SNEPPEN, K., Epigenetics as a first exit problem, *Physical Review Letters*, 88(4):048101, **2002**.

AVOLIO-HUNTER, T. M., LEWIS, P. N. AND FRAPPIER, L., Epstein-Barr nuclear antigen 1 binds and destabilizes nucleosomes at the viral origin of latent DNA replication, *Nucleic Acid Research*, 29:3520–3528, **2001**.

BAER, R., BANKIER, A. T., BIGGIN, M. D., DEININGER, P. L., FARRELL, P. J., GIBSON, T. J., HATFULL, G., HUDSON, G. S., SATCHWELL, S. C., SEGUIN, C., TUFFNELL, P. S. AND BARRELL, B. G., DNA sequence and expression of the B95-8 Epstein-barr virus genome, *Nature*, 310:207–211, **1984**.

BARWELL, A. B. J. A., PFUETZNER, R. A., WILLIAM FUREY, J., EDWARDS, A. M. AND FRAPPIER, L., Crystal structure of the DNA-binding domain of the Epstein-barr virus origin-binding proteins EBNA1, *Cell*, 83:39–46,, **1995**.

BAUMFORTH, K. R. N., YOUNG, L. S., FLAVELL, K. J., CONSTANDINOU, C. AND MURRAY, P. G., The Epstein-Barr virus and its association with human cancers, *Journal of Clinical Pathology*, 52:307–322, **1999**.

BILU, Y. AND BARKAI, N., The design of transcription-factor binding sites is affected by combinatorial regulation, *Genome Biology*, 6:R103, **2005**.

BINTU, L., BUCHLER, N. E., GARCIA, H. G., GERLAND, U., HWA, T., KONDEV, J., KUHLMAN, T. AND PHILLIPS, R., Transcriptional regulation by the numbers: models, *Curr Opin Genet Dev*, 15:116–124,, **2005**.

BLAKE, W. J., KAERN, M., R.CANTOR, C. AND COLLINS, J., Noise in eukaryotic gene expression, *nature*, 422:633–637, **2003**.

BODESCOT, M., PERRICAUDET, M. AND FARRELL, P. J., A promoter for the highly spliced EBNA family of RNAs of Epstein-Barr virus, *Journal of Virology*, 61:3424–3430,, **1987**.

BORNKAMM, G. W. AND HAMMERSCHMIDT, W., Molecular virology of Epstein-barr virus, *The Royal Society*, 356:437–459,, **2001**.

BOTQUIN, V., HESS, H., FUHRMANN, G., ANASTASSIADIS, C., GROSS, M. K., VRIEND, G. AND SCHOLER, H. R., New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 ans Sox-2, *Genes & Development*, 12:2073–2090, **1998**.

BRASSET, E. AND VAURY, C., Insulators are fundamental components of the eukaryotic genome, *Heredity*, 94:571–576, **2005**.

CHEN, J. AND MATTHEWS, K. S., Deletion of lactose repressor carboxyle-terminal domain affects tetramer formation, *The Journal of Biological Chemistry*, 267(20):13843–13850, **1992**.

COOKSON, S., OSTROFF, N., LEE PANG, W., VOLFSON, D. AND HASTY, J., Monitoring dynamics of single-cell gene expression over multiple cell cycles, *Molecular Systems Biology*, p. 2005.0024, **2005**.

DAVENPORT, M. G. AND PAGANO, J. S., Expression of EBNA-1 mRNA is regulated by cell cycle during Epstein-Barr virus type I latency, *Journal of Virology*, 73:3154–3161, **1999**.

DODD, I. B., PERKINS, A. J., TSEMITIDIS, D. AND EGAN, J. B., Octamerization of λ CI repressor is needed for effective repression of PRM and efficient switching from lysogeny, *Genes & Development*, 15:3013–3022, **2001**.

DODD, I. B., SHEARWIN, K. E., PERKINS, A. J., BURR, T., HOCHSCHILD, A. AND EGAN, J. B., Cooperativity in long-range gene regulation by the lambda CI repressor, *Genes & Development*, 18:344–354, **2004**.

DUNAWAY, M., OLSON, J. S., ROSENBERG, J. M., OLGA B. KALLAI, DICKERSON, R. E. AND MATTHEWS, K. S., Kinetic studies of Inducer Binding to *lac* Repressor-Operator complex, *The Journal of Biological Chemistry*, 255(21):10115–10119, **1980**.

ELF, J., LI, G.-W. AND XIE, X. S., Probing transcription factor dynamics at the single-molecule level in a living cell, *Science*, 316:1191–1194, **2007**.

ELF, J., PAULSSON, J., BERG, O. G. AND EHRENBERG, M., Near-critical phenomena in intracellular metabolite pools, *Biophysical journal*, 84:154–170, **2003**.

FRAPPIER, L. AND O'DONNELL, M., Epstein-Barr nuclear antigen 1 mediates a DNA loop within the latent replication of origin of Epstein-Barr virus, *Proc.Natl. Acad. Sci*, 88:10875–10879, **1991**a.

FRAPPIER, L. AND O'DONNELL, M., Overproduction, purification and characterization of EBNA1, the origin binding protein of Epstein-Barr virus., *The Journal of Biological Chemistry*, 266:7819–1826, **1991**b.

GILBERT, W. AND MULLER-HILL, B., Isolation of the Lac repressor, *Proc. Natl. Acad. Sci.*, 56:1891–1898, **1966**.

GILLESPIE, D. T., Exact stochastic simulation of coupled chemical reactions, *The Journal of Physical Chemistry*, 81:2340–2361, **1977**.

GOLDBERG, A. D., ALLIS, C. D. AND BERNSTEIN, E., Epigenetics: a landscape takes shape, *Cell*, 128:635–638, **2007**.

GOTESMAN, M. E. AND WEISBERG, R. A., Little lambda, who made thee?, *Microbiology and Molecular Biology Reviews*, 68:796–813, **2004**.

HERSKOWITZ, I. AND HAGEN, D., The lysis-lysogeny decision of phage λ: explicit programming and responsiveness, *Annual Reviews Genetics*, 14:399–445, **1980**.

HONG, M., MURAI, Y., KUTSUNA, T., TAKAHASHI, H., NOMOTO, K., ANS SHIN ISHIZAWA, C.-M. C., ZHAO, Q.-L., OGAWA, R., HARMON, B. V., TSUNEYAMA, K. AND TAKANO, Y., Suppression of Epstein-barr nuclear antigen 1 (EBNA1) by RNA interference inhibits proliferation of EBV-positive Burkitt's lymphoma cells, *Journal of Cancer Research and Clinical Oncology*, 132:1–8, **2006**.

HU, M. C.-T. AND DAVIDSON, N., The inducible lacoperator-repressor system is functional in mammalian cells, *Cell*, 48:555–566, **1987**.

JACOB, F. AND MONOD, J., Genetic regulatory mechanisms in the synthesis of proteins, *Journal of Molecular Biology*, 3:318–329, **1961**.

JONES, C. H., HAYWARD, S. D. AND RAWLINS, D. R., Interaction of the lymphocyte-derived Epstein-Barr virus nuclear antigen EBNA-1 with its DNA-binding sites, *Journal of Virology*, 63:101–110, **1989**.

JONES, P. A. AND TAKAI, D., The role of DNA methylation in mammalian epigenetics, *Science*, 293:1068–1070, **2001**.

VAN KAMPEN, N. G., *Stochastic processes in physics and chemistry*, Elsevier Science, **1992**.

KAUFMANN, B. B. AND VAN OUDENAARDEN, A., Stochastic gene expression: from single molecules to the proteome, *Current Opinion in Genetics & Development*, 17:107–112, **2007**.

KEMLER, I. AND SCHAFFNER, W., octamer transcription factors and the cell type-specificity of immunoglobulin gene expression, *The FASEB Journal*, 4:1444–1449, **1990**.

KENNEDY, G. AND SUGDEN, B., EBNA-1,a bifunctional transcriptional activator, *Molecular and Cellular biology*, 23:6901–6908, **2003**.

KITTEL, C. AND KROEMER, H., *Thermal Physics*, W. H. Freeman and Company, **2000**.

KLEIN, E., KIS, L. L. AND KLEIN, G., Epstein-barr virus infection in humans: from harmless to life endagering virus-lymphocyte interactions, *Oncogene*, 26:1297–1305, **2007**.

KUPFER, S. R. AND SUMMERS, W. C., Identification of a Glucocorticoid-responsive element in Epstein-Barr virus, *Journal of Virology*, 64:1984–1990, **1990**.

LEDERBERG, E. M. AND LEDERBERG, J., Genetic studies of lysogenicity in Escherichia Coli, *Genetics*, 38:51–64, **1953**.

LEIGHT, E. R. AND SUGDEN, B., EBNA-1: a protein pivotal to latent infection bt Epsein-barr virus, *Reviews in Medical Virology*, 10:83–100, **2000**.

LEVITSKAYA, J., SHARIPO, A., LEONCHIKS, A., CIECHANOVER, A. AND MASUCCI, M. G., Inhibition of ubiquitin/proteasome-dependent protein degradation by hte Gly-Ala repeat domain of the Epstein-Barr virus nuclear antigen 1, *Proc. Natl. Acad. Sci. USA*, 94:12616–12621, **1997**.

LILLYCROP, K., DENT, C., WHEATLEY, S., BEECH, M., N.N. NINKINA, WOOD, J. AND LATCHMAN, D., The octamer-binding protien Oct-2 represses HSV immediate-early genes in cell lines derived from latently infectable sensory neurons, *Neuron*, 7:381–390, **1991**.

LING, P. D., RAWLINS, D. R. AND HAYWARD, S. D., The Epstein-Barr virus immortalizing protein EBNA-2 is targeted to DNA by a cellular enhancer-binding protein, *Proc. Natl. Acad. Sci*, 90:9237–9241, **1993**.

LWOFF, A., Lysogeny, *Bacteriological Reviews*, 17:269–332, **1953**.

MAERKL, S. J. AND QUAKE, S. R., A systems approach to measuring the binding energy landscapes of transcription factors, *Science*, 315:233–237, **2007**.

MALIN, S., LINDERSON, Y., ALMQVIST, J., ERNBERG, I., TALLONE, T. AND PETTERSON, S., DNA-dependent conversion of Oct-1 and Oct-2 into transcriptional repressors by Groucho/TLE, *Nucleic Acids Research*, 33:4618–4625, **2005**.

MEYER, B. J. AND PTASHNE, M., Gene regulation at the right operator (OR) of bacteriophage λ, III. λ repressor directly activates gene transcription, *Journal of Molecular biology*, 139:195–205, **1980**.

NILSSON, T., ZETTERBERG, H., WANG, Y. C. AND RYMO, L., Promoter-proximal regulatory elements involved in *oriP*-EBNA1-independent and -dependent activation of the Epstein-Barr virus C promoter in B-lymphoid cell lines, *Journal of Virology*, 75:5796–5811, **2001**.

NONKWELO, C., RUF, I. K. AND SAMPLE, J., The Epstein-Barr virus EBNA-1 promoter Qp requires an intitator-like element, *Journal of Virology*, 71:354–361, **1997**a.

NONKWELO, C., RUF, I. K. AND SAMPLE, J., Interferon-independent and -induced regulation of Epstein-barr virus EBNA-1 gene transcription in Burkitts Lymphoma, *Journal of Virology*, 71:6887–6897, **1997**b.

OGBOURNE, S. AND ANTALIS, T. M., Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes, *Biochemical Journal*, 331:1–14, **1998**.

O'GORMAN, R. B., ROSENBERG, J. M., KALLAI, O. B., DICKERSON, R. E., ITAKURA, K., RIGGS, A. D. AND MATTHEWS, K. S., Equilibrium binding of inducer to lac repressor-operator DNA complex, *The Journal of Biological Chemistry*, 255(21):10107–10114, **1980**.

ORPHANIDES, G., LAGRANGE, T. AND REINBERG, D., The general transcription factors of RNA polymerase II, *Genes & Development*, 10:2657–2683, **1996**.

PLATT, T., MILLER, J. H. AND WEBER, K., In vivo degradation of mutant Lac repressor, *Nature*, 228:1154–1156, **1970**.

POKROVSKAJA, K., EHLIN-HENRIKSSON, B., KISS, C., CHALLA, A., GORDON, J., GOGOLAK, P., KLEIN, G. AND SZEKELY, L., CD40 ligation downregulates EBNA-2 and LMP-1 expression in EBV-transformed lymphoblastoid cell lines, *Int J Cancer*, 99:705–712, **2002**.

PTASHNE, M., *A genetic switch, second edition*, Cell Press & Blackwell scientific publications, **1992**.

PUGLIELLI, M. T., WOISETSCHLAEGER, M. AND SPECK, S. H., oriP is essential for EBNA gene promoter activity in Epstein-Barr virus-immortalized lymphoblastoid cell lines, *Journal of Virology*, 70:5758–5768, **1996**.

RASER, J. M., Noise in gene expression: origins, consequences, and control, *Science*, 309:2010–2013, **2005**.

REISMAN, D. AND SUGDEN, B., Trans activation of an Epstein-Barr viral transcription enhancer by the Epstein-Barr viral nuclear antigen 1, *Molecular and cellular Biology*, 6:3838–3846, **1986**.

REVET, B., VON WILCKEN-BERGMANN, B., BESSERT, H., BARKER, A. AND MULLER-HILL, B., Four dimers of λ repressor bound to two suitably spaced pairs of λ operators form octamers and DNA loops over large distances, *Current Biology*, 9:151–154, **1999**.

ROBERTS, J. W. AND ROBERTS, C. W., Proteolytic cleavage of bacteriophage lambda repressor in induction, *Proc. Natl. Acad. Sci.*, 72:147–151, **1975**.

ROWE, M., ROWE, D. T., GREGORY, C. D., YOUNG, L. S., FARRELL, P. J., RUPANI, H. AND RICKINSON, A. B., Differences in B cell growth phenotype reflect novel patterns of Epstein-Barr virus latent gene expressions in Burkitt's lymphoma cells, *The EMBo Journal*, 6:2743–2751, **1987**.

RUF, I. K. AND SAMPLE, J., Repression of Epstein-Barr virus EBNA-1 gene transcription by pRb during restricted latency, *Journal of Virology*, 73:7943–7951, **1999**.

SALAMON, D., TAKACS, M., UJVARI, D., UHLIG, J., WOLF, H., MINAROVITS, J. AND NILLER, H. H., Protein-DNA binding and CpG methylation at nucleotide resolution of latency -associated promoters Qp, Cp, and LMP1p pf Epstein-Barr virus, *J Virol*, 75:2584–2596, **2001**.

SAMPLE, J., HENSON, E. B. AND SAMPLE, C., The Epstein-Barr virus nuclear protein 1 promoter active in type I latency is autoregulated, *J Virol*, 66:4654–4661, **1992**.

SCHAEFER, B. C., PAULSON, E., STROMINER, J. L. AND SPECK, S. H., Constitutive activation of Epstein-Barr virus (EBV) nuclear antigen 1 gene transcription by IRF1 and IRF2 during restricted EBV latency, *Molecular and Cellular Biology*, 17:873–886, **1997**a.

SCHAEFER, B. C., STROMINGER, J. L. AND SPECK, S. H., Redefining the Epstein-Barr virus-encoded nuclear antigen EBNA-1 gene promoter and transcription initiation site in group I Burkitt lymphoma cell lines, *Proc. Natl. Acad. Sci.*, 92:10565–10569, **1995**.

SCHAEFER, B. C., STROMINGER, J. L. AND SPECK, S. H., Host-cell-determined methylation of specific Epstein-Barr virus promoter regulates the choice between distinct viral latency programs, *Molecular and Cellular Biology*, 17:364–377, **1997**b.

SERFLING, E., JASIN, M. AND WALTERSCHAFFNER, Enhancers and eukaryotic gene transcription, *Trends in genetics*, 1:224–230, **1985**.

SHEA, M. AND ACKERS, G., The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation, *J Mol Biol*, 181:211–230, **1985**.

STORMO, G. D., DNA binding sites: representation and discovery, *Bioinformatics*, 16(1):16–23, **2000**.

STRUHL, K., Fundamentally different logic of gene regulation in eukaryotes and prokaryotes, *Cell*, 98:1–4, **1999**.

STURM, R. A. AND HERR, W., The POU domain is a bipartite DNA-binding structure, *Nature*, 336:601–604, **1988**.

SUMMERS, H., BARWELL, J. A., PFUETZNER, R. A., EDWARDS, A. M. AND FRAPPIER, L., Cooperative assembly of EBNA1 on the Epstein-Barr virus latent origin of replication, *Journal of Virology*, 70:1228–1231, **1996**.

SUNG, N. S., KENNEY, S., GUTCH, D. AND PAGANO, J. S., EBNA-2 transcativates a lymphoid-specific enhancer in the *Bam*HI C promoter of Epstein-Barr virus, *Journal of Virology*, 65:2164–2169, **1991**.

SUNG, N. S., WILSON, J., DAVENPORT, M., SISTA, N. D. AND PAGANO, J. S., Reciprocal regulation of the Epstein-Barr virus *Bam*HI-F promoter by EBNA-1 and an E2F transcription factor, *Molecular and Cellular Biology*, 14:7144–7152, **1994**.

THOMPSON, M. P. AND KURZROCK, R., Epstein-barr virus and cancer, *Clinical Cancer Research*, 10:803–821, **2004**.

THORLEY-LASWON, D. A. AND GROSS, A., Persistence of the Epstein-Barr virus and the origins of associated lymphomas, *the New England Journal of Medicine*, 350:1328–1337, **2004**.

TOMILIN, A., REMENYI, A., LINS, K., BAK, H., LEIDEL, S., VRIEND, G., WILMANNS, M. AND R-SCHOLER, H., Synergism with the coactivator OBF-1 (OCA-B,BOB-1) is mediated by a specific POU dimer configuration, *the New England Journal of Medicine*, 350:1328–1337, **2004**.

TURNER, B. M., Histone acetylation and the epigenetic code, *BioEssays*, 22:836–845, **2000**.

WERNER, M., *Numerical solution to the master equation using the linear noise approximation*, Master's thesis, Uppsala University, **2004**.

WOISETSCHLAEGER, M., JIN, X. W., YANDAVA, C. N., FURMANSKI, L. A., STROMINGER, J. L. AND SPECK, S. H., Role for the Epstein-Barr virus nuclear antigen 2 in viral promoter switching during initial stages of infection, *Proc. Natl. Acad. Sci*, 88:3942–3946, **1991**.

WOLFFE, A. P. AND MATZKE, M. A., Epigenetics: Regulation through repression, *Science*, 286:481–486, **1999**.

WRAY, G. A., HAHN, M. W., ABOUHEIF, E., BALHOFF, J. P., PIZER, M., ROCKMAN, M. V. AND ROMANO, L. A., The evolution of transcriptional regulation in Eukaryotes, *Mol Biol Evol*, 20:1377–1419, **2003**.

WYSOKENSKI, D. A. AND YATES, J. L., Multiple EBNA1-binding sites are required to form an EBNA1-dependent enhancer and to activate a minimal replicative origin within *oriP* of EPstein-Barr virus, *Journal of Virology*, 63:2657–2666, **1989**.

YOUNG, L. S. AND MURRAY, P. G., Epstein-Barr virus and oncogenesis: from latent genes to tumours, *Oncogene*, 22:5108–5121, **2003**.

YOUNG, L. S. AND RICKINSON, A. B., Epstein-Barr virus: 40 years on, *Nat Rev Cancer*, 4:757–768, **2004**.

ZETTERBERG, H., BORESTROM, C., NILSSON, T. AND RYMO, L., Multiple EBNA1-binding sites within oriPI are required for EBNA1-dependent transactivation of the Epstein-Barr virus C promoter, *International Journal of Oncolocy*, 25:693–696, **2004**.

ZETTERBERG, H., STENGLEIN, M., JANSSON, A., RICKSTEN, A. AND RYMO, L., Relative levels of EBNA1 gene transcripts from the C/W, F and Q promoters in Epstein-Barr virus-transformed lymphoid cells in latent and lytic stages of infection, *Journal of General Virology*, 80:457–465, **1999**.

ZOU, J., ALMQVIST, J. AND ERNBERG, I., The Epstein-Barr virus enhancer FR - consequences of DNA sequence variation and binding of OCt and its co-regulator Grg/TLE *in vitro* and *in vivo*, **2006**, submitted.