MODELS OF EPIGENETICS

Bi- and Multi-stability in Biological Systems

Anne Katrine Alsing

Supervisor: Prof. Kim Sneppen

PhD thesis submitted at
University of Copenhagen
Faculty of Science
Niels Bohr Institute
Center for Models of Life
Denmark

November 2012
Preface

This thesis has been submitted for the assessment of the reward of the PhD degree at the Niels Bohr Institute, Faculty of Science, University of Copenhagen. The PhD study was conducted under the supervision of professor Kim Sneppen, in the period from October 2008 to November 2012, including 11 months of maternity leave. The thesis include two per-reviewed papers and a manuscript currently undergoing the review process in Nucleic Acid Research. The PhD study was supported partly by the Faculty of Science, University of Copenhagen and partly by the Danish National Research Foundation through the Center for Models of Life.

I should like to thank professor Kim Sneppen for his kind supervision and encouraging comments, in particular during the last intensive months of my PhD study. The positive atmosphere and the fine colleagues at the Center for Models of Life, is something I will surely come to miss. Throughout my efforts at describing and comprehending the genetic switch of phage TP901-1, I have been guided by Professor Karin Hammer and Postdoc Margit Pedersen. Their profound knowledge on the switch and molecular biology in general has eased my way into the world of biochemical procedures and proper experimental set ups. For their patience and guidance I am truly grateful.

In addition I would like to thank Professor Steen Pedersen, who open heartedly let me work in his laboratory and together with his lab technician, Tony Hansen, lent valuable experience for my first attempts on protein synthesis and protein-DNA interaction experiments. Also I am grateful for the opportunity to make a scientific visit in the lab of research director Christian Cambillau at Architecture et Fonction des Macromolécules Biologiques at the University of Marseille. Even though none of my experimental work made it into this thesis, my stay in Luminy under kind supervision from Mariella Tegoni meant a great deal to me.

Finally, but foremost, I wish to thank my family for never ending support and for continuously adjusting my perspectives. You stabilize my life.
Abstract

It is becoming increasingly clear that differences in DNA sequence can not be the sole carrier of the phenotypic differentiation that arises when a cell expresses different genes to different degrees. A simple example, such as the cells of the human body, clearly shows that cells carrying the same genomic material can show quiet diverse phenotypes characterized by organ specific expression patterns. The mechanisms responsible for this phenotypic plasticity are characterized as epigenetic, as they inflict their effect "epi-" (Greek for “above” or “on top”) of the genetic code. For a gene regulatory mechanism to be classified as epigenetic, it is required that it is self-sustainable in the sense that the governed gene expression or repression should prevail for the lifetime of the cell and must be inherited by possible daughter cells.

An example of epigenetic differentiation is the bistable switch of phage TP901-1, in which the devotion to a distinct state is governed by a small network of trans-acting factors. TP901-1 infects lactococcal bacteria and either exploit the host for the production of new phage leading to the death of the host, or become integrated on the host genome as a non-lethal tag-along during subsequent cell divisions. The switch of the phage, responsible for the dedicated choice, is a small DNA fragment comprising two promoters set back-to-back and their adjacent genes. The gene products work as regulators of the promoters, joining the regulatory loop of the switch. In its lethal path, it is proposed that the switch should implement an interaction between the two antagonistic proteins, and the switch DNA. Such an interaction is novel within the biological context of phage switches. However, the network partly resembles a theoretic mixed feedback module able to show bistability. From comparisons between experimental data and model predictions based on mathematical descriptions, it is verified that TP901-1 could be a first biological example of a phage switch achieving bistability through the trans-acting effect of a repressing heteromer.

While networks such as that of TP901-1, relies on diffusible factors for the maintenance of memory over cell divisions, another way of staying devoted to a distinct state is marking the activity state of a gene on its DNA. In eukaryotes such marks are known, either as methylation of cytosine bases or modifications of the nucleosomes that wind the DNA into chromatin structures. Once established the patterns may be conserved over many cell generations. The self-sustainable nature of the patterns is attributed to the cis-acting mechanism of read-write enzymes that facilitates the same histone modifications as they recognize.

Developing olfactory neurons become devoted to expression of one version of hundreds of olfactory receptor genes present in the mammalian genome. Ol-
factory receptor genes are very similar in sequence and upstream promoter regions. This disqualify simple gene regulation by a realistic pool of diffusible trans-acting factors as the activation of one receptor gene must result in the continuous repression of the remaining genes. Consequently differentiation of olfactory neurons presents a first example of a system that seems to require a local cis-acting gene regulatory mechanism.

A novel model is presented that describe the choice of olfactory receptor by a system of coupled bistable regions. The bistability of each receptor gene is attributed to patterns of similar nucleosome marks, governed by the cis-acting mechanism of read-write enzymes. Genes are coupled by a negative feedback on this mechanism by a global factor given by any active gene. The model allows for a detailed study of a mechanism generating multistability, and shows how single gene expression can be obtained in a family of multiple similar genes all bound by the same regulatory mechanisms.
Danish abstract / Dansk resume

I kølvandet på kortlægningen af det fulde genom fra flere arter er det blevet klart, at forskelle i DNA-sekvens ikke alene kan forklare de forskelle i fænotype der opstår når forskelige gener i forskellig grad udtrykkes til proteiner. Et nærliggende eksempel er majoriteten af celler i menneskekroppen. De bærer det samme genetiske materiale, men viser meget forskellige fænotyper kendteget ved proteinproduktion fra organspecifikke gener.

De mekanismer der er ansvarlige for denne fænotypiske plasticitet, betegnes epigenetiske fordi de virker “epi-“ (græsk for “over” eller “i tillæg til”) den genetiske kode, der dikterer rækkefølgen af aminosyrer i de enkelte proteiner. En epigenetisk genregulering skal således styre udtryksgraden af de forskellige proteiner i cellens genom uden, at der er foretaget ændringer i DNAet. I øvrigt skal en epigenetisk tilstand bevares over hele cellens levetid og i særlighed nedarves til mulige datterceller, nøjagtigt ligesom en ændring i DNA-sekvensen nedarves. Denne afhandling studerer to systemer, der hver især viser den bi- eller multistabilitet, der ofte kendteger epigenetisk regulering, gennem deres evne til udelukkende at udtrykke et gen ud af to eller flere mulige.

Det første system er den bistabile fag TP901-1, der inficerer mælkesyrebakterier. Ved en infektion kan fagen vælge at udnytte den inficerede vært til produktion af nye fager, hvilket i sidste ende dommer værten til døden. Alternativt implementeres fagens genom i værterne genom og ducereres fredsommeligt sammen med dette ved efterfølgende celledelinger. Fra genetisk manipulerede systemer hvor den dødelige udgang frakobles, er det konstateret at valget mellem de to udviklinger er yderst dedikeret og fastholdes gennem værternes celledelinger.


Selv for et simpelt bistabilit system som switchen, kommer vores forestillingsevne til kort på grund af systemets dynamiske og ikke-linearke natur. I anal-
yser af tilsidste netværk fra andre systemer der vare tager genregulerende
beslutninger, udnyttede i stedet kvantitative metoder til at forstå systemernes
dynamik. I denne afhandling foretages en kvantitativ analyse af forskellige
konkrete udlægninger af de molekyler/biologiske modeller der er opstillet for
de to tilstande a TP901-1 switchen. Målet er at teste modellerne pålidelighed
i forhold til opretholdelsen af bistabilitet. Herudover testes hver model for
hvorvidt den kan reproducerere tidligere observerede forhold mellem promotor
aktiviteter i de to stabile tilstande. Desuden testes modellerne mod et datasæt
af promotoraktivitet for forskellige koncentrationer af det ene protein CI. Analy-
syerne tillader, at en model der er baseret på direkte binding mellem DNA og
de to proteiner kan afskrives. Desuden viser analysen, at de biologiske modeller
må revurderes, for at kunne beskrive alle eksperimentelle observationer.

Den anden epigenetiske mekanisme der betragtes i af handleren er den genreg-
ulerende effekt der tilskrives modifikationer af histoner. I eukaryoty er DNAet
viklet omkring nukleosomer og tæt pakket i de kromatin strukturer, der danner
vores kromosomer. Hver nukleosom består af par af fire centrale histoner. His-
toner er proteiner og deres sidste 10-20 aminosyrer i hver ende, modificeres efter
histonerne samles i nukleosomerne. Forskellige konkrete modifikationer er sat i
forbindelse med aktive og inaktive genregioner. Endelig er histonmodifikationer
erne arvelige, idet de vedligeholdes gennem celledeling, og de fremstår derfor
som indlysende epigenetiske træk.

I af handleren undersøges det teoretiske grundlag for histonmodifikationers epig-
enetiske potentielle med en stokastisk model. Modellen giver en ny beskrivelsen
af differentieringen af pattedyrs olfaktoriske nerveceller. De olfaktoriske nerve-
celler sidder i epitelet i næsehulen og giver os vores lugtesans. Cellebestanden i
epitelet fornyes konstant gennem et pattedyrs levetid, så nye nerveceller udvikles
jævnligt fra stamceller. Hver olfaktorisk neuron er kendtegnet ved ekslusiv
ekspressionen af én udgave af de olfaktoriske receptorer (ORer) der vare tager
opfangelser af lugte. Receptorgenet vælges tilfældigt ud af en familie af 900-
1400 gener. Det betyder, at hvert olfaktorisk receptorgen skal være i en af to
tilstande; enten udtrykt og dermed valgt, eller ikke udtrykt og altså fraværdigt.
Desuden skal valget af et gen udelukke, at nogen af de andre gener udtrykkes.
Valget skal være stabilt og een olfaktorisk neuron kendtegnes således ved dens
hengivenhed til et enkelt af OR-generne.

Det olfaktoriske valg, af ét blandt tusinde mulige gener, repræsenterer et system
med flere adskilte stabile tilstande. Der er opstillet adskilte forskel på forsk på
t at forklare, hvordan denne tilsyneladende tilfældige multistabilitet skabes. Her
undersøges differentieringen af de olfaktoriske neuroner i en ny model, inspireret
af forsk i forslag om at histonmodifikationer kan være involveret i valget. Modellen im-
plementerer et tidligere teoretisk studie af lokal bistabilitet mellem dominerende
histonmodifikationer i en afgrenset genregion. I den modeller, præsenterer her,
beskrives hvert OR gen som sådan en bistabil genregion der kobles til de
øvrige OR gener via en globalt virknings proteinfaktor. Modellen tillader dermed
den detaljeret undersøgelse af mekanismerne bag multistabiliteten, og viser hvor-
dan singular genekspression kan opstå i en familie af mange ens gener der alle
er bundet af de samme reguleringsemekanismer.
List of appended papers

Two papers on the genetic switch of TP901-1 found in appendixes A and B.


One manuscript on the differentiation of olfactory neurons found in appendix C.

# Contents

<table>
<thead>
<tr>
<th>Preface</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td>Danish abstract / Dansk resume</td>
<td>vii</td>
</tr>
<tr>
<td>List of appended papers</td>
<td>ix</td>
</tr>
<tr>
<td>Contents</td>
<td>x</td>
</tr>
<tr>
<td>1 Aim and outline</td>
<td>1</td>
</tr>
</tbody>
</table>

## Part I. Epigenetics

<table>
<thead>
<tr>
<th>2 Decisions on a molecular scale</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Epigenetics through gene regulatory networks</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Histone modifications as epigenetic traits</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Epigenetic systems considered in this thesis</td>
<td>11</td>
</tr>
</tbody>
</table>

## Part II. Switch of Phage TP901-1

<table>
<thead>
<tr>
<th>3 Bistability of a temperate phage</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Phage TP901-1</td>
<td>19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4 Modelling the switch</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Model basics</td>
<td>24</td>
</tr>
<tr>
<td>4.2 Extending model to full system</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5 Results from modelling a modified model system</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Model A - Direct binding at O_M</td>
<td>33</td>
</tr>
<tr>
<td>5.2 Model B - Formation of a heteromer in the cytoplasm</td>
<td>36</td>
</tr>
<tr>
<td>5.3 Model A vs Model B</td>
<td>38</td>
</tr>
<tr>
<td>5.4 Relating parameters to reality</td>
<td>40</td>
</tr>
</tbody>
</table>


Chapter 1

Aim and outline

The aim of this PhD study was to use a theoretical approach to investigate two different mechanisms at play in cells that achieve stable expression of a single gene chosen among two or more possible genes. In both of the systems the choice is made without any changes to the genetic code and may in principle be undone by external forces. Thus each cell within a population of cells with identical DNA can differentiate into two or multiple distinct states characterized by specific gene expression profiles. Such differentiation is called epigenetic, which also covers the heritability of the states to subsequent generations.

In a classical view gene expression patterns arise as results of binding between molecules in the cytoplasm. E.g binding of RNA polymerase at a genes promoter region is a prerequisite for transcription initiation. Additionally the association may be affiliated or inhibited by binding of other proteins at the promoter region. Networks of such trans-acting transcription factors can show oscillations in the expression of involved genes, inflict continuous expression of single genes, or give the system a build-in bistability, that let it choose a distinct state and stay devoted to it for many cell generations.

Networks of transcription factors relies on pools of diffusible factors for the maintenance of memory over cell divisions, another way of staying devoted to a distinct state is marking the activity state of a gene on its DNA. In eukaryotes such marks are known, either as methylation of cytosine bases or modifications of the nucleosomes that wind the DNA into chromatin structures. Patterns of certain histone modifications and DNA methylations may be dictated by diffusible trans-acting factors. However, once established the patterns may be conserved over many cell generations and two copies of a chromosome region within the same cell may carry differently dictating patterns, supposedly without the need for the initial regulating factors. The self-sustainable nature of the patterns is attributed to the cis-acting mechanism of read-write enzymes that facilitates the same modifications as they recognize. The system of olfactory neuron differentiation studied in the second half of this thesis, presents a first example of a biological system where such local gene regulation is required.

The first system is the genetic regulatory switch of phage TP901-1. This bistable switch implements a small network of trans-acting transcription factors. The
network closely resemble that of the mixed feedback motif that have only been
proposed theoretically. The system is studied in a quantitative analyses with
the objective to capture the underlying mechanisms of bistability. To match the
bistability in promoter activity in the phage, the network contains a repressive
regulatory link in additionel to the links of the mixed feedback motif. The
results of the study were communicated in two papers and are reported in Part
II of the thesis. In the first paper a simplified version of the regulatory system
is considered, while the second paper consider the full bistable switch of phage
TP901-1 in a combination of experimental and theoretical work.

Part III reports the advance of a novel model of the multistable system of
olfactory receptor gene expression made in the third appended manuscript. The
differentiation of olfactory neurons is strongly connected to the devotion to
expression of a single olfactory receptor gene from a large family of possible
genes. The singularity of olfactory receptor gene expression has been the subject
of intense experimental studies since the Nobel prize winning discovery of the
receptors and their coupling to odour sensation. This theoretical study of the
maintenance of exclusive gene expression couples bistable systems through a
globally acting negative feedback. The bistability of the subsystems is modelled
by local positive feedback between DNA bound nucleosomes. The olfactory
differentiation model thus combines both cis- and trans-acting mechanisms of
regulation.

A short introduction to epigenetics in general and the epigenetic nature of the
two considered systems in particular are given in Part I. Conclusion and per-
spective remarks are given individually for each of the two parts, II and III.
References are listed alphabetically by first author and are found before the
appended papers.
Part I

Epigenetics
Part I frontage illustration collected from www.
Chapter 2

Decisions on a molecular scale

Life is characterized by its immense diversity. Bacteria, plants, insects, vertebrates, take any biosphere on the planet and life will present itself in some form. Majestic emperor penguin males sticking it out in breeding colonies during the blistering cold of the Antarctic winter; water vapour harvesting by the Namib desert beetles living in one of the driest deserts of the world; the transparent appearance of the jellyfish; or the humongous trees found in the forests of California stretching more than 100 meters into the air. The diversity is fascinating. Even more so, when considering differences within individual life forms. Take the straightforward example of humans. Noses may be big or small, straight or crooked. Variances in height are evident from a walk in any metropolitan and eyes come in any colour from grey to almost bright yellow. In our post-genomic era differences are often attributed to differences in the genome, but even between monozygotic twins, carrying identical DNA, phenotypic differences can be observed, see Fig. 2.1.

Eusocial insects such as honeybees, ants and wasps are astonishing examples of this phenotypic plasticity. The delightfully productive nests of the honeybee consists of a reproductive fertile queen, a multitude of sterile female workers and seasonally a few thousand drones. The honeybee queen is bigger than its minions and have fully developed ovaries but in contrast lacks the glands producing wax for building the combs of the hive. Initial mating with several drones enables the queen to produce progeny of both sexes for the remainder of its life. Female honeybee larvae all posses the potential to become queens as no genetic differences dictate the caste. Rather it is the nutritional and social environment of each larva that sets its fate. Larvae destined to be queens are solely nourished by royal jelly, while to-be-worker larvae must make do with lower quality food [138]. In the undisturbed nest the caste fate is conserved, however if a honeybee colony looses its queen a worker bee may develop ovaries underpinning that the caste differentiation is not genetic.

Increasing the magnifying glass, further lends to the astonishment. Simply consider the diversity among the cells constituting the human body all developed
Decisions on a molecular scale

Figure 2.1: Differentiation between individuals with identical DNA. A) Monozygotic twins. Despite having the same DNA some differences can be observed between the sisters. Figure adapted from [30]. B) Honeybee queen and workers. The queen is generally larger, longer lived and have developed ovaries in contrast to the sterile female workers. However, any female larva posses the potential to develop into a queen without changes in the heritable genetic code. C) Army ant Eciton burchelli soldier and minor worker. In some ant species the workers are found in two or more morphological distinct types. Figures B and C adapted from [138]. D) Bacillus subtilis in the process of sporulating. The bright circles are spores being developed within the mother cell. Figure adapted from [67].

from the same fertilized egg. In the path from embryo to fully developed species, cells are continuously being devoted to certain tasks. Nerves, muscle cells, liver cells, or the epidermal cells of our skin; all are easily separated under a microscope. Even a simple unicellular species like the bacteria Bacillus subtilis develop population heterogeneity. In the face of a limiting supply of resources about half of the B. subtilis in a population will sporulate while others may switch to utilise other metabolites [132]. Spores are tolerant to extreme environmental conditions and the organism can stay sporulated for years waiting for better times. When these arrives, the spore germinates and daughter cells of the renewed cell division are genetic and phenotypic indistinguishable from their distant relatives before the sporulation. Even in mid-exponential growth under conditions of plenty metabolites, B. subtilis are found as two strikingly different cell types. Most of the bacteria swim by means of a flagellum but a subpopulation grow in long immobile chains [48]. Both bet-hedging strategies are found within isogenic populations.

Some of the examples of phenotype diversity given above describes transient changes that may be reverted such as the sporulation of B. subtilis. However, in many cases the choice of phenotype should be upheld for the life of the cell or even more extreme; remain unaffected by subsequent cell divisions. You do not want the cells of your eye to suddenly start excreting gastric acid as the cells lining the stomach. Moreover, in all the examples the diversity of phenotypes
Figure 2.2: Central Dogma of genetics and the packaging of DNA into chromosomes. The genetic information is generally transferred in a single direction from DNA to protein through transcription and translation by RNA polymerase and ribosomes, respectively. A cell is characterized by which genes are expressed into proteins through this information pathway and by the expression level of the individual genes. This defines the gene expression pattern of the cell. Expression of genes are regulated from transcription initiation to final protein folding. In eukaryotes gene expression is additionally regulated by the winding of the DNA around nucleosomes, consisting of four core histones. Specific post translational modifications of the tails of the bound histones have been proposed to alter the accessibility of the transcriptional machinery to the genes. Figures adapted from [44, 139].

are achieved without changes to the heritable genetic code. This dedication to a fixed functional identity, heritable through generations, without changes in the underlying DNA sequence is what defines epigenetic differentiation.

On a molecular level different cells are specified by their expression (or repression) of given genes, resulting in the proteins that govern the characteristic function of the cell. E.g. the same gene expression pattern needed for the establishment of a functional liver cell should not be found in nerve cells constituting the brain and vice versa. The central dogma of molecular biology first stated by Francis Crick in 1958 sets the basics of the differentiation. The dogma couples the heritable code of DNA to the functional form of proteins, see Fig. 2.2 [18, 19]. The DNA of active genes is transcribed by RNA polymerase resulting in messenger RNA (mRNA) which in turn is translated into proteins by ribosomes, completing expression of the gene.

Obviously it is not for every gene to be expressed in every cell. Mechanisms of gene expression regulation are found on every level of the path from transcription initiation to final protein folding. Transcription of a gene initiates at the promoter of the gene. Transcription factors may bind at these promoter regions and prevent or recruit the binding of the transcriptional machinery. microRNA can target mRNA resulting in rapid degradation or simple abortion of translation. On top of this, post-translational modification may render the final protein inactive though ready for activation. In eukaryotes further levels of regulation are achieved by methylation of the DNA and the winding of DNA in chromatin structures. Regions of DNA carrying methylations at the cytosines show reduced gene expression. Hymenoptera, like the honeybees, are differen-
Decisions on a molecular scale

Figure 2.3: Small networks of regulatory feedback systems that may exhibit bistability. Apart from an overall positively reinforcing feedback, systems must include an ultra-sensitive regulation of at least one of its components and appropriately balanced strengths of its parts, to obtain bistability. Figure a-c adapted from [4].

2.1 Epigenetics through gene regulatory networks

In bacteria the classical example of gene regulation is networks of trans-acting transcription factors. E.g. Sporulation of \textit{B. subtilis} is preserved by the auto-stimulating architecture of the protein Spo0A. In its phosphorylated form Spo0A stimulates transcription from its own promoter and expression of the gene \textit{kinA}, which encode a kinase involved in the phosphorylation of Spo0A. Under non-sporulating conditions expression of the protein Spo0E which acts to dephosphorylate Spo0A is increased, preventing the establishment of an endospore [83].

From a physicist perspective epigenetic differentiation of a cell into one of two states, simply describes as a bistable system. However, it is not a simple task to stick to a decision when everything is more or less transient as is the case in cellular processes. No binding between molecules is unbreakable and thermal fluctuations cause probabilistic outcomes of transcription and translation events [75, 50]. Moreover, the differentiation must be preserved even when the stimulus that triggered it, is withdrawn. In \textit{B. subtilis} the sporulated form is maintained as Spo0A is self-activating and -phosphorylating, providing an introductory example of the positive feedback that is one requirement for bistability in networks of trans-acting factors.

In a study of transcriptional networks J.E. Ferrell Jr. listed three requirements of a system to execute bistability. First, the system must include some sort
of feedback that positively enforce the chosen state. Positive feedback between gene products is one example, another is the double negative feedback between two mutually inhibiting gene products, A and B. In the latter case the distinct states are either A expressed and repressing expression of B or conversely, B being expressed and repressing A. Networks containing more than two components and both positive and negative regulation may also show bistability, required the result of the feedback is positive, see Fig. 2.3. Secondly the system must posses some kind of non-linearity in the feedback circuit. Non-linearity may be achieved through some of the involved gene products responding ultra-sensitively to their regulators, e.g. through cooperative binding. Finally the system must be rightfully balanced. If any part of the system is too powerful it will render the system monostable rather than bistable [46, 4].

As mentioned the true transcriptional networks include many other regulatory mechanisms than simple transcription regulation. In fact theoretical studies have identified a core genetic module, called the mixed feedback module, that only include transcription regulation of one of the two gene products, X and Y, see Fig. 2.3d. Instead the two distinct states of high X/low Y and low X/high Y are governed through sequestration. Say X inhibits expression of Y, then the former case is established by high X expression as in the double negative motif. However, in this case the latter state does not include repression of X but the sequestration of X into a complex with Y. In a reasonable balanced system excess expression of Y will sequestrate all the produced X allowing for a pool of free Y [27].

2.2 Histone modifications as epigenetic traits

While trans-acting transcription factors do provide a sound basic understanding of gene regulation in prokaryotes they can not cover the full picture with regard to eukaryotes. In contrast to prokaryotes, the genome of eukaryotes is restricted to the nucleus of the cell and additionally is wound around octamers of histones in the nucleosome complex [53]. Neighbouring nucleosomes resembles beads on a string and form the chromatin fibre that is condensed into chromosomes 10,000fold shorter than the length of the constituting DNA, illustrated in Fig. 2.2.

Nucleosomes are protein structures composed of two duplicates of each of the four core histones; H2A, H2B, H3 and H4, Fig. 2.4. The core histones all have highly dynamic amino terminal tail extruding from the nucleosome [58]. The tails are approximately 20-35 residues in length and are subjects of intense post transcriptional modification (PTM) that impact the transcriptional activity of the genes at the associated DNA. Modifications such as methylation, acetylation, phosphorylation, and others to certain histone tail residues have been identified as active or silencing marks corresponding to the state of the associated regions of the genome [103, 74, 9]. E.g. trimethylation of lysine 27 in histone H3 is found in the nucleosomes comprising the chromatin at inactive genes, while the same modification at lysine 36 marks chromatin associated with active genes [14, 8]. The different histone modifications may loosen or tighten the histone-DNA association easing or restricting the availability of the associated DNA to transcription factors and the transcriptional machinery, or simply act as
Decisions on a molecular scale

Figure 2.4: Illustration of a nucleosome. A nucleosome is composed of duplicates of each of the four core histones; H2A, H2B, H3 and H4. Each core nucleosome wraps $\sim 150 \text{ bp}$ of DNA and is separated by $\sim 15\text{bp}$ to the neighbouring nucleosome [74]. The N-terminal tails of the core histones are subject to extensive post translational modifications that impact the transcriptional activity of the genes of the associated DNA. The illustration include the covalent post transcriptional modifications; methylation (Me), acetylation (Ac), ubiquitination (Ub), and phosphorylation (Ph) on the N- and C-terminal tails of the histones. Examples of marks associated with inactive genes are indicated by red dots while the green dot indicate a mark of transcriptional active genes. In general active genes are enriched in H3, H4 and H2A acetylation [103]. Figure modified from [129]

recruiting factors them self [74].

The addition or removal of certain tail modifications thus present an additional mean for regulating gene expression effectuated by the enzymes responsible for the post translational modifications. Writer enzymes include histone methyl transferases (HMTs) and histone acetyl transferases (HATs), responsible for methylation and acetylation, respectively. The removal of modifications are effectuated by eraser enzymes like histone demethyl transferases (HDMTs) and histone deacetylases (HDACs) [129]. The modification of certain DNA regions have classically been attributed to interactions between DNA, regulating transcription factors and such writer or eraser enzymes. While this provide for a way of establishing regions of specific modifications it does not resemble a mean for self-sustained devotion to a distinct pattern upon removal of the initial transcription factor as we expect of epigenetic differentiation.

A third group of proteins termed readers will recognize and bind histone modifications. The polycomb repressive complex 1 (PRC1) presents a combination of a reader and writer as it recognizes trimethylation on lysine 27 of H3 and subsequent ubiquitimates lysine 199 of H2. More interesting in the context of epigenetics is the recent discovery of such read-write combinations that will
2.3 Epigenetic systems considered in this thesis

recognize the same modification they effectuate. The gene silencing activity of the PRC2 is governed by the histone lysine methyltransferase activity of one subunit, but the complex also include the subunit EED that recognizes this exact modification [66]. A similar constitution is seen in methyltransferase Arabidopsis ASHH2 in which different domains can distinguish different methylated histone tails [42]. The read-write enzymes provide a mean of preserving the state of a cluster of nucleosomes of identical modification and can explain how patterns of histone modifications are propagated and inherited through many cell divisions. In comparison to the trans-acting mechanism of diffusible transcription factors, combined read-write enzymes give rise to the phenomenon of epigenetic inheritance through a cis-acting mechanism of positive feedback loops [145].

An example of epigenetic differentiation proposed to be governed by histone modifications is the silenced mating-type locus of fission yeast Schizosaccharomyces pombe (S. pombe). The wild-type locus is enriched for nucleosomes in which histone 3 carries the silent mark of methylated lysine 9. Mating-type switching in S. pombe involves three loci of which only one locus, mat1, is expressed and determines the mating type of the cell. The remaining loci are transcriptional silent but provide genetic information to mat1 dictating the switch between the two mating types of S. pombe. Deletion of the region between the two silent loci and nearby placement of the reporter ura4+, result in bistability in mating type switching phenotype as well as bistability in expression of ura4+ [36, 126]. It is observed that in the silent state of Ura4+ cells, nucleosomes of the locus are enriched in the same silent methylation marks as the wild type locus while silent marks are absent in the active state cells. The two states are highly stable with switching occurring less than once in 10,000 generations underpinning the heritable nature of the histone marks.

Stable X-chromosome inactivation is another example of epigenetic differentiation preserved by histone modifications. In female mammals one of the two X-chromosomes must be silenced to avoid double dosage of X bound gene products compared to their male counterparts. In the early development of placental animals, each cell of the blastocyst independently and randomly inactivates one X-chromosome copy. Inactivation is associated with upregulation of the local Xist gene, however effectively maintained through compaction of the silenced chromosome into heterochromatin, characterized by repressive histone marks. All descendants of a cell with a particular X chromosome inactivated, will have the same X-chromosome inactivated [6]. In female tortoiseshell cats this independent and random inactivation of one of the X-chromosomes is visible in the mottled coloration of the fur. Alleles directing fur colour reside on the X-chromosome. Thus for any given patch of fur, coloration results from the allele positioned at the chromosome that was not inactivated.

2.3 Epigenetic systems considered in this thesis

The present thesis studies two different means of achieving epigenetic differentiation, either as the bistable choice between two states, or the multistable case where a choice is made of one out of multiple possible states.
First up is the bistable switch of temperate phage TP901-1 that infects lactic acid bacteria. In this context bistability establishes a choice of gene expression activity from only one of two early promoters determining the outcome of an infection. The choice is governed by mutual transcriptional regulation by two phage encoded proteins, CI and MOR. Thus the phage switch is an example of the networks of trans-acting transcription factors that have been identified as effectors of gene regulatory decisions. The phage repressor, CI, represses expression of MOR by binding at the $P_L$ promoter region. Oppositely, repression of CI expression requires the presence of both MOR and CI itself. It is proposed that CI and MOR repress CI expression by forming a complex on the DNA in vicinity of the promoter $P_R$ that governs CI expression. However, such interactions are still to be confirmed experimentally. With the proposed interaction the network of the phage partly resembles that of the theoretically described mixed feedback module, but including an additional repressive link to account for the mutual exclusive activity from the two promoters, see Fig. 2.5A.

Small switch regulatory networks have been the subject of intense study for many decades providing a solid basis for a modelling approach to study the bistability of TP901-1. TP901-1 is the first system described that could utilize a combination of sequestration and double negative trans-acting feedback. Our study identify the requirement for sequestration and points at an inconsistency in the description of the switch when comparing models to experiments.
2.3 Epigenetic systems considered in this thesis

Epigenetic differentiation through the effects of trans-acting factors, generally relies on the continues upholding of a pool of the regulating factors for maintaining the devotion over time and cell divisions. In a trans-acting system, multiple copies of a switch element like that of phage TP901-1, will all settle into the same state, governed by the winning pool of regulators assessable from all parts of the system. In contrast systems where two copies of the same genome is found in different states of activity may not readily be described simply by networks of transcription factors, without the incorporation of clustering effects or very strong cooperativity between regulators [106, 52].

In the differentiation of developing olfactory neurons a choice is made between multiple stable states. Each mature olfactory neuron is characterized by the expression of a single olfactory receptor (OR) gene, chosen randomly from a large family of highly similar genes. Each OR gene represents a bistable system, coupled to the activity of every other OR gene to ensure that only one OR gene is actively expressed. The promoter region of the OR genes is mostly conserved, and the choice of an OR gene is thus unlikely to be governed by networks of transcription factors.

Instead we present a model of this epigenetic differentiation based on the regulatory effect of histone modifications and associated read/write enzymes, and a global negative feedback from any active gene on the activation of others, illustrated in Fig. 2.5B. Our model achieves to encompass or reproduce all the main characteristics of olfactory neuron differentiation, and sets a framework for coupling the intracellular state of a neuron to its axon connections.
Decisions on a molecular scale
Part II

Switch of Phage TP901-1
Part II frontage illustration: Transmission Electron Micrograph of two phage TP901-1. Adapted from [133]
Chapter 3

Bistability of a temperate phage

In the simplest imaginable epigenetic differentiation a choice is made between one of two states. Temperate phage present a textbook example of such bistable devotion. Phage are virus-like structures that infects bacteria, explore the transcriptional and translational machinery for production of new phage, and eventually lyses the host releasing the progeny for new rounds of infection. A temperate phage may, however, abort this lethal path and instead become silenced as a prophage, often by the implementation of its genetic luggage in the hosts genome, see Fig. 3.1. Once silenced the prophage genome is stably duplicated along with the hosts in subsequent bacteria generations. This stable state is only aborted by changes in cellular conditions such as activation of the host SOS response system. This decision between a lytic path of a virulent phage or the lysogenic path leading to phage silencing represents epigenetic differentiation. The lysogenic choice is past along to new bacterial offspring without alternations in the genome of the phage.

In the intensively studied *E. coli* phage λ the choice is founded in a small DNA fragment identifying the genetic switch of the phage. The genes, promoters and operator sites of the switch comprises an example of the before mentioned networks of trans-acting transcription factors, see Fig. 3.2. At the base of the two stable states of the switch is the mutual exclusive expression of the regulator proteins CI and Cro [86]. CI is the gene product of cI positioned immediate downstream of the PRM promoter, while Cro results from expression of gene Cro downstream of PR [85, 25]. The lytic path is maintained when the switch is dominated by expression from the PR promoter, resulting in high levels of Cro protein. Binding of Cro at operator site OR3 prevents RNApolymerase binding at the PRM promoter and thus prevents expression of CI. Similarly the lysogenic path is maintained by expression of CI from PRM. CI will bind cooperatively at OR1 and OR2 to repress transcription from PR [89]. The lysogenic path is re-enforced by CI as binding at OR1 and OR2 enhance RNApolymerase association at PRM. At high enough CI concentration CI will bind additionally to OR3 and repress its self-activation [88].
Bistability of a temperate phage

Figure 3.1: Illustration of the two paths resulting from the bistable switch of a temperate phage. Left side presents the lytic path leading to host lysis and release of phage progeny for new rounds of infections. Right side presents the lysogenic path of genomic insertion and prophage like silencing. The full circle of the lysogenic path indicate that the prophage may be induced by changes in cellular conditions such as activation of the host SOS response system.

It is important to note that the interactions between regulators and operator sites need not be extremely stable in order for the states of the switch to be stable. In fact the bistability is coupled to the double negative feedback loop, the cooperativity of CI binding at OR1 and OR2, and the self-activation of CI expression that reinforces the lysogenic state [5, 7].

Regulatory networks of transcription factors enforcing mutual repression through loops of positive feedback are well studied mechanisms for epigenetic memory. However, as already mentioned, simple gene regulation with individual transcription factors mutually excluding the expression of each other as in phage

Figure 3.2: Illustration of the gene regulation in the switch of *E. coli* phage λ. A) In a lysogen *cl* is expressed from PRM and binds cooperatively at the two operator sites OR1 and OR2 to inhibit RNApolymerase (RNAp) binding at PR and expression of Cro. In lytic growth OR1 and OR2 are free for RNAp to bind and express Cro. Cro in turn will bind at OR3 and repress CI expression by inhibiting RNAp binding at PRM. Figure adapted from [87]. B) The core motif of the switch. CI repress Cro production and Cro repress CI production. The dashed arrow from CI upon itself indicate the self activation by RNAp recruitment by CI bound at OR2 and OR3.
λ are not sufficient for understanding epigenetic memory of the cells of higher prokaryotes. Even simple species like phage may implement more evolved networks including transcription factor interactions and yet to be discovered mechanism.

3.1 Phage TP901-1

*Lactococcus Lactis* temperate phage TP901-1 have a genetic switch with a twist. Though very similar to λ in its genomic setup, the early lysogenic promoter of this phage is only fully repressed in the presence of both regulators [63].

Temperate phage TP901-1 belongs to the type P335 phage a subspecies of the *Siphoviridae* family, identified by having double-stranded DNA, isometric head and non-contractile tail [15, 11], see Fig. 3.3B. TP901-1 is induced by UV light from *L. Lactis* sups. *cremoris* 901-1. The lytically propagated phage will resume the lysogenic path in 3% of infections in laboratory strains of lactococci illustrating the temperate nature of the phage. Resulting lysogenic bacteria are immune to re-infection by TP901-1 but will exhibit the spontaneous release of phage that is a criterion differentiating the lysogenic state from that of a prophage. During infection of exponential growing *L. lactis* the latent period of TP901-1 is 65 minutes with a mean burst size of 40 ± 10 new phage.

Initially the molecular study of phage TP901-1 was taken on with the aim of developing genetic tools and discovering global gene regulatory systems in lactococci [15]. In fact, the characterization of the site-specific integration of the circular phage DNA of TP901-1 into its hosts genome during lysogenic devotion, allowed for the construction of two promoter-reporter integration vectors. Such vectors are important tools for single-copy cloning and identification of promoters, terminators and signal sequences [13, 12]. The latter aim found its subject in the study of the regulatory mechanisms responsible for the choice between lysogeny and lysis.

Characterization of the transcription of different regions in the 38.4 kb genome of phage TP901-1 identified three temporal phases of transcription. The region transcribed in the early phase includes two opposing promoters termed P_R and P_L, see Fig. 3.3A. *cl*, the gene transcribed immediate downstream of P_R, shows local homology to repressors of other lactococcal phage and is succeeded by genes necessary for the chromosomal integration of the phage DNA. In contrast the first transcribe from P_L, the *mor* gene, was initially proposed to have a Cro-like function, and is followed by genes related to phage replication [62]. Early transcription from either of the promoters thus dictate the fate of the phage.

Functional analysis of a ~1kb genetic fragment containing only P_R and P_L, and the genes, *cl* and *mor* revealed the switch responsible for the bistability observed in the activity of the early promoters of TP901-1. Introduced on a plasmid in *L. lactis* this fragment settles into one of two distinct states each characterized by activity from only one of the promoters [63]. The repressive nature of C1, the gene product of *cl*, was concluded from a series of measurements of the activity of P_R and P_L by the reporter β-galactosidase. The measurements showed 31- and 62-fold decreases in promoter activity of P_R and P_L, respectively, in the
Bistability of a temperate phage

Figure 3.3: The genetic switch of bacteriophage TP901-1. A) The full genome of the phage covers 38.4 kb and is integrated at the attB site in the host genome during the lysogenic development. The sketched switch region covers 979 bp and encompasses two promoters, P_R and P_L, set back-to-back and directing, respectively, the transcription of genes cI and mor. Three operator sites are found in the region, O_R, O_L and O_D. Figure adapted from [81]. B) Transmission electron micrograph of phage TP901-1. Figure adapted from [133]. C) Atomic model of putative oligomeric structure from rigid body modelling against scattering data. The structure presents one of more possible models. Figure adapted from [81]. D) Example of clonal variation of the switch region. Colonies of L. lactis subsp. cremoris containing plasmids with the switch region in a P_R-lacLM fusion. Grown on plates containing X-Gal expression of lacLM is identified as blue colonies. Upper part show a restreak of a single colony of white transformants with P_R activity repressed. Lower part show restreak of single colony of blue transformants with P_R partly de-repressed.

presence of cI downstream of P_R, compared to measurements after a frameshift mutation was introduced in cI. In the absence of both cI and mor, P_L showed about 4 times the activity of P_R and the activity of P_L was the same regardless of the presence or absence of P_R, showing that P_L activity is unaffected by activity of P_R. In contrast to the repressive effect of CI, the presence of mor alone does not affect the promoter activities. The tight repression of P_R inflicted during lytic growth requires the presence of both cI and mor [63].

During phage infections, P_L promotes expression of the genes involved in the early state of the lytic development in 97 % of infections, while P_R initiates the lysogenic silencing of the phage in the remaining 3 %. Colonies of L. lactis transfected with plasmids carrying the switch in a P_L-lacLM fusion show the same frequency between blue (P_L active leading to lacLM expression) and white (P_L closed) phenotypes when grown on plates containing X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside). Thus the small DNA fragment contains all the genetic material required for repression of both promoters and consequently constitute the switch for the establishment of bistability [63].

Bacteria containing the plasmid with the switch in the state of repressed P_L are immune to infections by phage TP901-1. Working with the switch of TP901-1
in plasmid systems one thus recognize the state of repressed P_L as the immune state of the switch, while the state of an active P_L promoter is classified as anti-immune. The immunity obtained by the switch-carrying plasmid is governed by the presence of cI. Intriguingly, large expression of mor will also convey immunity to TP901-1 infections. Both CI and MOR (the gene product of mor) are trans-acting as their supplement from a second plasmid complement introduced mutations in the switch fragment [63].

The phage repressor, CI, expressed from P_R show large similarities in protein sequence to the repressors of other P335 type phage with its helix-turn-helix motif associated to DNA binding [63]. Accordingly purified CI bind with different affinities to different regions of the switch fragment. Combined with sequence analysis of the switch this identified three CI DNA binding operator sites, O_R, positioned between the promoters, O_L just downstream of P_L, and a distant site, O_D, flanking the end of mor [45], see Fig. 3.3A. The three operator sites are all composed of inverted repeats suitable for binding of a dimer of CI and consistently CI forms dimers in solution as many known phage repressors. The distance between O_R and O_L amounts to 63bp or six helical turns positioning the operator sites on the same site of the DNA and allowing for multiple CI dimers to bind cooperatively at the operators [45, 82]. Gelfiltration reveal the formation of higher CI multimers compatible in size to a CI hexamer. From analysis of one dimensional SAXS data the most likely structure of such hexamers was found to be a flat disc with the six CI subunits presenting their DNA binding regions in pairs of two at the periphery of the disc [82], see Fig, 3.3C. In such a structure the distance between pairs of DNA binding regions is approximately 160˚A not far from the 204˚A separating six helical turns. With some bending of the DNA, the disc structure thus could associate at both O_R and O_L.

Proposed models describing the states of the genetic switch

Two models have been proposed for describing the two different states of the bistable switch from phage TP901-1. In the immune state the repressor, CI, represses transcription from P_L through cooperative binding at the operator sites O_L and O_R [45]. Further cooperative binding of CI to a distant operator site O_D tightens this repression [81]. Binding of CI at O_R also represses the early lysogenic promoter P_R. O_R is a relatively weak CI binding site thus with falling concentrations of CI, repression of P_R should be released before repression of P_L, allowing for replenishing of the CI deposit and renewed binding at O_R, Fig. 3.4A.

The phage encoded protein MOR functions as an anti-repressor of the phage [63]. However, MOR alone does not repress any of the two promoters in vivo. Tight repression of the P_R promoter in the anti-immune state is only obtained in presence of both CI and MOR, and is apparently independent of the known CI operator sites [81]. Hence, the model of the tight repression of P_R in the anti-immune state involves interaction between CI, MOR and supposedly the DNA in vicinity of the P_R promoter, see Fig. 3.4A, even though the interaction has not been verified experimentally [82]. The descriptive models of the steady state interactions of proteins and DNA in the two stable states of the switch are
Figure 3.4: Biological model of protein interactions in the two stable states of the genetic switch of TP901-1. **A)** In the immune state transcription initiation from promoter $P_L$ is repressed by binding of CI dimers to the operator sites $O_L$ or $O_R$. In the model three CI dimers are envisioned to form a hexamer increasing the chance of simultaneous binding to all three operator sites. The anti-immune state requires presence of both CI and MOR. Expression of CI is tightly repressed in the anti-immune state and the biological model envision a binding between CI, MOR and the DNA in the vicinity of promoter $P_R$ [81]. Promoters are shown as black arrows indicating the direction of transcribing RNA polymerase. RNA polymerases are omitted for clarity. **B)** The core motif of the switch comparable to Fig. 3.2B. CI repress MOR production but tight repression of CI requires both CI and MOR.

the subject of the theoretical approach presented in the following chapters.
Chapter 4

Modelling the switch

In phage TP901-1 only one of the two proteins required for bistability of the genetic switch have been shown to interact with DNA related to the repression of the lytic promoter $P_L$. Full repression of the lysogenic promoter $P_R$, requires presence of both of the proteins. The described biological models of the interactions between CI, MOR and DNA in the two stable states of the switch provide a plausible and compact description of the system. However, some questions remain unattended and phage TP901-1 continues to be a puzzling twist on the understanding of mechanisms that may ensure bistability. Generally speaking, it is not easy to predict the behaviour of a bistable switch without quantitative analysis because a bistable switch is a dynamical and highly nonlinear system. In the present case of TP901-1 the switch could involve a number of mechanisms, such as cooperativity binding via homo-/hetero- dimerization, sequestration via heterodimerization, and intertwined loops of negative and positive feedback via protein interactions.

The proposed interaction between CI and MOR in the repression of $P_R$, can be split into two general cases. Either the interaction occurs in the cytoplasm, prior to $P_R$ repression, or it could be restricted to combined binding at the DNA. Both cases apparently compatible with the experimental observations. In the former case the switch of TP901-1 would implement a mechanism very similar to the mixed feedback module described in part I [27]. In the immune state CI represses expression of MOR, giving rise to a state of high CI and low MOR concentration, while in the anti-immune state CI and MOR form heteromers which would provide for a mean to achieve a state of high MOR concentration and low free CI levels. However, simply achieving reversed relations between concentrations of free CI and MOR is not the bi-stable situation needed in the switch. In the switch we need the bistability to apply to the promoter activities. Not only has this been observed in experiments on plasmid inserted switch regions, it also provide a mean of excluding expression of downstream genes involved in lysogony or lysis, respectively [63].

In order to unveil how the two states of the phage switch are stabilized we perform quantitative analysis of specific models based on the biological molecular model presented by Pedersen and Hammer in 2008. By confronting numerical results with experimental data, we can restrict possible mechanisms with plau-
Modelling the switch

Figure 4.1: Genetic switch of TP901-1. a) Wild-type genetic switch with the three known CI operator sites $O_R$, $O_L$, and $O_D$. Also shown are the direction the promoters $P_R$ and $P_L$ and the relative position of the $cl$ and $mor$ genes. The cyan box represents the putative $O_M$ site. b) Sketch of the modified switch investigated in the first appended paper [76]. c) Immune state of the modified switch. CI represses transcription from $P_L$ approximately 1000-fold by binding to the $O_L$ operator site with transcription from $P_R$ being allowed. d) Anti-immune state of the modified switch. Repression of $P_R$ depends on both CI and MOR, which repress transcription from $P_R$ approximately 100-fold with transcription from $P_L$ being allowed [81]. Figure adapted from [76]

4.1 Model basics

In the late 1970’s Shea and Ackers presented their statistical mechanical model of promoter activities in the switch of *E. coli* phage $\lambda$. The model assigns statistical weights to every possible binding configuration between proteins and switch DNA as function of protein concentrations, with the dependence characterized by the Hill coefficient and the affinities of the proteins to the operator sites [1]. The probability of finding a given promoter in a transcription permissive state are then given by the equilibrium statistical weights with which the corresponding operators are not bound by the regulators. We will assume that the probability of finding the promoter in a transcription permissive state is proportional to the chance of active transcription, providing a description of the activity probability of the promoters. This approximation holds when the time that RNA polymerase needs to start elongation after binding to DNA is much shorter than the time scales of binding/unbinding of RNA polymerase and repression factors to the promoter/operator sites [75, 115].

To establish the theoretical approach let’s consider the modified switch system that is the subject of the first appended paper [76]. This regulatory circuit consists of the two promoters $P_L$ and $P_R$, which direct the production of MOR and CI, respectively, the known CI binding site $O_L$ and presumably the putative
site for MOR dependent repression of P_R, termed O_M (Fig. 4.2b). The promoter P_L is repressed by CI binding at O_L leading to the P_L activity given as a function of the CI concentration:

\[ p_L([CI]) = p_{L0} \cdot f_{OL}([CI]) \]  

(4.1)

where \( p_{L0} \) is the bare activity of the promoter P_L. The function \( f_{OL}([CI]) \) represents the repression factor. In the absence of CI, there is no repression: \( f_{OL}([CI] = 0) = 1 \). The P_R activity, on the other hand, depends on the concentrations of both CI and MOR. Accordingly, the P_R activity can be written as

\[ p_R([MOR],[CI]) = p_{R0} \cdot f_{OM}([MOR],[CI]) \]  

(4.2)

with \( p_{R0} \) being the bare activity of the P_R promoter. The function \( f_{OM}([MOR],[CI]) \) is the repression factor due to the binding of MOR and CI at O_M and satisfies \( f_{OM}([MOR] = 0,[CI]) = f_{OM}([MOR],[CI] = 0) = 1 \).

As P_L produces MOR and P_R produces CI, the total concentrations of each protein is governed by the dynamic equations,

\[
\frac{d}{dt}[MOR]_{\text{total}} = \frac{1}{\tau_M} (p_L([CI]) - [MOR]_{\text{total}})
\]  

(4.3)

\[
\frac{d}{dt}[CI]_{\text{total}} = \frac{1}{\tau_C} (p_R([MOR],[CI]) - [CI]_{\text{total}})
\]  

(4.4)

where \( \tau_M \) and \( \tau_C \) are the degradation times for MOR and CI, respectively.

We which to consider the steady states of the system where the protein production and degradation should balance. With the promoter activities rescaled by the degradation times and thus given in terms of the steady state protein concentrations, we identify the steady state conditions:

\[ p_L([CI]) = [MOR]_{\text{total}} \]  

(4.5)

\[ p_R([CI],[MOR]) = [CI]_{\text{total}} \]  

(4.6)

Not all steady states are stable against small perturbations. If a small perturbation drives the system out of the steady state the state is unstable. The stability of a state is governed by the dynamic equations for the total concentration of CI and MOR. If there are two stable steady states, the system shows bistability. With these conditions we may graphically inspect the promoter activities of the steady states once explicit forms for both sites are established.

Within the initially mentioned approximations we assume that the activity of a promoter is repressed from its bare activity by the statistical weight of the binding configurations where the associated operator sites are not bound by repressing proteins. In table 4.1 the four binding configurations considered in the model of the modified switch are marked by stars. This identify the form of the repression factors \( f_{OL} \) and \( f_{OM} \), equivalent to the first statistical model on phage \( \lambda \) [1, 110]:

\[
p_L([CI]) = p_{L0} \cdot \frac{\sum \text{(weights of permitting states)}}{\sum \text{(weights of all states)}}
\]  

(4.7)

\[
\downarrow
\]

\[
p_L([CI]) = p_{L0} \cdot \frac{1}{1 + [CI_2]/K_{OL}}
\]  

(4.8)
In the repression of $P_L$, CI forms dimers which in turn bind to the $O_L$ operator site, blocking activity from $P_L$. For the $P_R$ promoter repression two scenarios are considered. In Model A CI and MOR form a complex on the DNA in vicinity of the $P_R$ promoter thus blocking transcription. In Model B CI and MOR form a heterodimer prior to binding at the DNA and the resulting repression of $P_R$ activity. Figure adapted from [76].

Where it is assumed that CI dimers effectuate the repression by binding at $O_L$, see Fig. 4.2, as $K_{OL}$ represents the $CI_2$ concentration at which $O_L$ is occupied half of the time. For the MOR independent repression of $P_L$ we suppose CI dimers effectuate the repression by binding at $O_L$, see Fig. 4.2, The concentration of CI dimers is related to the CI monomer concentration through:

$$[CI_2] = \frac{[CI]^2}{K_{CI_2}}$$  \hspace{1cm} (4.9)
with dissociation constant $K_{CI_2}$, allowing us to give the activity of $P_L$ as

$$p_L([CI]) = pL_0 \cdot \frac{1}{1 + [CI]^2/(\tilde{K}_{OL})^2} \quad (4.10)$$

where $\tilde{K}_{OL} \equiv \sqrt{K_{CI_2} \cdot K_{OL}}$ is the effective affinity of $O_L$ binding, setting the CI monomer concentration at which $O_L$ is occupied for 50% of the time.

For the repression of $P_R$ we test two scenarios, resulting in different explicit expressions for the two sides of Eq. 4.6. In Model A we assume CI and MOR form a complex on the DNA in vicinity of $P_R$ to repress activity from the promoter, see figure 4.2. The activity of the $P_R$ promoter is repressed from the bare activity $pR_0$ by the statistical weight that the putative operator site $O_M$ is unoccupied by CI and MOR:

$$pR([MOR],[CI]) = pR_0 \cdot \frac{\sum \text{(weights of permitting states)}}{\sum \text{(weights of all states)}} \quad (4.11)$$

$$\Downarrow$$

$$pR([MOR],[CI]) = pR_0 \cdot \frac{1}{1 + [MOR][CI]/(K_{OM})^2} \quad (4.12)$$

$$\Downarrow$$

$$pR(pL([CI]),[CI]) = pR_0 \cdot \left[1 + \frac{pL_0\tilde{K}_{OL}/(K_{OM})^2}{([CI](K_{OL}) + (K_{OM}/[CI]))} \right]^{-1} \quad (4.13)$$

Where we used the fact that total concentration of MOR may by approximated by the concentration of MOR monomers; $[MOR]_{total} = [MOR]$, and the steady state condition 4.5 to eliminate the MOR concentration from the explicit form of the $P_R$ activity. The affinity $K_{OM}$ is the concentration $\sqrt{MOR \cdot CI}$ where $O_M$ is occupied by CI and MOR half of the time. Finally the total CI concentration in Model A is given as:

$$[CI]_{total} = [CI] + 2 \frac{[CI]^2}{K_{CI_2}} \quad (4.14)$$

establishing the form of the right side of Eq. 4.6 and allowing for a graphical solution to the steady state condition by plotting both sides as a function of [CI].

In Model B we assume that CI and MOR form a heteromer in solution prior to binding at the DNA and repression of $P_R$, see figure 4.2. The concentration of the MOR:CI heteromer is given as:

$$[MOR:CI] = \frac{[MOR][CI]}{K_{MOR:CI}} \quad (4.15)$$

with the dissociation constant $K_{MOR:CI}$. Now the activity of the $P_R$ promoter is repressed from the bare activity $pR_0$ by the statistical weight that the putative operator site $O_M$ is occupied by the MOR:CI heteromer, and the activity given by statement 4.11 becomes:

$$pR([MOR],[CI]) = pR_0 \cdot \frac{1}{1 + [MOR:CI]/K_{OM}} \quad (4.16)$$
With $K_{OM}$ here being the concentration of MOR:CI heteromer where O_M is occupied for 50% of the time. The formation of the heteromer couples the total concentrations of CI and MOR, and consequently the expression for $[\text{MOR}]$ is a function of $[\text{CI}]$ and $[\text{MOR}]_{\text{total}}$:

$$[\text{MOR}]_{\text{total}} = [\text{MOR}] + [\text{MOR} : \text{CI}] \Rightarrow [\text{MOR}] = \frac{[\text{MOR}]_{\text{total}}}{1 + [\text{CI}] / K_{\text{MOR:CI}}}. \quad (4.17)$$

This relation and Eq. 4.5 allow us to once again give the explicit form of $P_R$ activity as function of $[\text{CI}]$:

$$p_R \left( \frac{p_L([\text{CI}])}{1 + [\text{CI}] / K_{\text{MOR:CI}}} \cdot [\text{CI}] \right) = p_{R_0} \left[ 1 + \frac{p_{L_0}\tilde{K}_{OL}/(\tilde{K}_{OM})^2}{([\text{CI}]/K_{OL}) + (K_{OL}/[\text{CI}])} \cdot \frac{1}{1 + [\text{CI}] / K_{\text{MOR:CI}}} \right]^{-1} \quad (4.18)$$

with the effective affinities: $\tilde{K}_{OM} \equiv \sqrt{K_{\text{MOR:CI}} \cdot K_{OM}}$ and $\tilde{K}_{OL} \equiv \sqrt{K_{CL} \cdot K_{OL}}$.

The coupling between CI and MOR also changes $[\text{CI}]_{\text{total}}$ as function of $[\text{CI}]$:

$$[\text{CI}]_{\text{total}} = [\text{CI}] + 2 \frac{[\text{CI}]^2}{K_{CI}} + p_L([\text{CI}]) \cdot \frac{[\text{CI}] / K_{\text{MOR:CI}}}{1 + [\text{CI}] / K_{\text{MOR:CI}}} \quad (4.19)$$

### 4.2 Extending model to full system

In the second theoretical approach on the bi-stable switch of phage TP901-1 we turn the attention to the full switch region including all of the three operator sites for CI binding, $O_R$, $O_L$, and $O_D$, and the putative $O_M$ site for MOR dependent $P_R$ repression. While the theoretical backbone is basically as described for the simplified system, we now need to make some additional assumptions.

In the biological model of the immune state of the switch the presence of all operator sites for CI binding and the fact that CI have been shown to form hexamers in solution, call for considerations on the binding species of CI.

Studies on the structure of CI propose that CI may associate into a hexamer in a disc like structure that would be able to place a DNA binding region of the CI proteins in the proximity of all of the three operator sites, see Fig 3.3 in chapter 3 [82]. Whether this disc forms like a trimer of dimers or a dimer of trimers can not be concluded and it is uncertain if the hexamer forms upon binding to the DNA or if free hexamers are present under the physiological condition of the cell. Mutants of CI that do not form the hexamer in in vitro studies, reproduce the bistability between the immune and anti-immune states of the switch fragment in vivo [82], indicating that hexamer formation is not required for operator binding. Additionally most other CI like phage repressors are only found as dimers in solutions. Consequently we assume that CI binds to the three known operator sites as dimers with the possibility of cooperative interactions between pairs or triplets of bound dimers, the latter describing the formation of a bound hexamer. See Fig 4.3.
4.2 Extending model to full system

\[ \text{PL repression} \]

\[ \text{Generally:} \]
\[ \text{Weight} = [\text{MOR}]^{j(s)}[\text{CI}]^{i(s)} \exp(-\Delta G_s/kB) \]

Figure 4.3: Sketch of additional assumptions on the CI mediated repression of PL. CI form dimers with dimerization energy \( \Delta G_{CI2} \). CI dimers bind at the operator sites with individual binding energies, \( \Delta G_{OL}, \Delta G_{OR} \) or \( \Delta G_{OD} \). Here exemplified by binding at OL. Two bound CI dimers may form cooperative interactions that strengthen the DNA association by energy \( \Delta G_2 \). Exemplified by cooperative binding of two CI dimers at OR and OL, however the dimer-dimer cooperative binding is allowed for all pairs of bound dimers in the model. Cooperative binding between three bound dimers is assigned the additional energy \( \Delta G_3 \). In calculating the weight of each binding configuration the total energy, \( \Delta G_s \), include the energy of all the protein-DNA bindings, possible cooperative energy and a dimerization energy for each of the dimers bound in the configuration. \( j(s) \) and \( i(s) \) is the total number of bound MOR and CI in configuration \( s \), respectively. MOR is not part of the depicted PL repression, however, the form of the assigned weights is general for both PL and PR repression.

The binding configurations included in the model are listed in table 4.1. It is assumed that activity from PL is repressed by CI binding at OR and OL, while the PR promoter is repressed by CI binding at OR and the MOR:CI:DNA association at OM. Including all of the known binding sites the activity of the
Table 4.1: Binding configurations included in model of full switch of phage TP901-1. “—” illustrates a free binding site while CI₂ and MOR:CI indicate binding of the according species. Parentheses indicate CI dimers associated with cooperative binding. “Yes” or “No” in PR and PL columns identify whether the configuration allow or inhibits transcription initiation from the promoter. First seven columns are the fifteen states included in the model of CI mediated repression. Last four columns represent the additional configurations resulting from binding of the MOR:CI complex. Binding at OM is assumed to be independent of binding at the other sites, thus configurations 15-29 also include the bindings of the in-row configuration 0-14. E.g. *18 is the state of CI₂ binding at OD and MOR:CI binding at OM allowing only transcription from PL. Stars indicate configurations included in the model of the modified switch, where OR and OŁ binding is excluded.

PL promoter becomes;

\[ p_L([CI]) = p_L(0) \cdot f_{OR,OL}([CI]) \]  

(4.20)

Where \( f_{OR,OL} \) is the repression factor due to binding of CI at either OR or OL, satisfying \( f_{OR,OL}(0) = 1 \). In a similar manner the PR activity is written as:

\[ p_R([MOR],[CI]) = p_R(0) \cdot f_{OM,OR}([MOR],[CI]) \]  

(4.21)

capturing the repression of PR by CI and MOR occlusion of OM, and CI binding at OR. The repression factor \( f_{OM,OR} \) satisfies \( f_{OM,OR}(0) = 1 \).

Again we assume that the repression factors are given by the equilibrium statistical weights that the corresponding operators are occupied. However, the additional binding species associated with the full system prevents the notion of explicit compact forms of the promoter activities as presented for the modified system. Instead we express the statistical weight of each microscopic binding configuration as a function of [MOR] and [CI] in the general form:

\[ w_s = [MOR]=[CI] \exp(-\Delta G_s/k_BT) \]  

(4.22)

30
with \( j(s) \) and \( i(s) \) being the total number of bound MOR monomers and CI monomers in configuration \( s \), respectively. \( \Delta G_s \) is the sum of all energies associated with binding configuration \( s \) normalised to zero for the configuration of no bound proteins [1]. We here express the weight of each configuration by the Gibbs free energy associated with the binding. This is conceptually the same as we did in the approach on the modified system, as a statistical mechanic description of ligand binding to an operator site allow us to make the connection:

\[
K = [1M] \cdot \exp(\Delta G/k_BT) \quad (4.23)
\]

Adopting the convention for \( \Delta G \) by measuring the prefactor in molar [117]. In the remainder of the thesis concentrations is thus always implicitly given in molar. With the weights we express the steady state promoter activity of the two promoters using Eq. 4.7 and 4.11.
Modelling the switch
Chapter 5

Results from modelling a modified model system

This chapter highlights the results of the publication included in appendix A. We consider promoter activity data reported for $P_R$ and $P_L$ on switch fragments implemented on multi-copy plasmids and introduced into $L. lactis$. Activities of the two promoters were reported as highly mutual exclusive. In the immune state the $P_L$ promoter activity is tightly repressed compared to the anti-immune state. For the $P_R$ promoter the relation is inverted, though the repression fold is less dominant as $P_R$ is auto-regulated by CI binding at $O_R$ in the immune state [81]. In a variant of the switch, where operator sites $O_R$ and $O_D$ have been mutated to restrict binding of CI to the $O_L$ site (see Fig. 4.1b), the bistability in promoter activities is even more pronounced. In the immune state, the $P_L$ promoter is repressed approximately 1000-fold compared to the unrepressed state, but high expression from $P_R$ is allowed due to the lack of auto-regulation by CI binding at $O_R$ (see Fig. 4.1c). In the anti-immune state, $P_R$ is repressed approximately 100-fold compared to the immune state while high expression from $P_L$ is allowed (see Fig. 4.1d). It is this data of promoter fold repressions for the modified version of the switch that is the subject of the first theoretical model.

We examine the possibility for bistability in this variant of the switch when described by each of the two models introduced in chapter 3. We numerically consider the behaviour of the models in the steady states, comparing the obtained repression folds to the experimental observations. This allows us to make restrictions on the possible involved mechanisms based on considerations of plausible parameters.

5.1 Model A - Direct binding at $O_M$

In Model A $P_R$ is repressed by binding of one CI monomer and one MOR monomer to $O_M$ without prior interaction between CI and MOR in the cytoplasm. Eqs. 4.10, 4.13 and 4.14 let us examine the behaviour of the activity
Figure 5.1: Model A with one CI monomer and one MOR monomer binding at O_M. Promoter activities of P_L (green curve), P_R (red curve) and [CI]_{total} (red dashed line) as function of free [CI]. Intersection between the curves of [CI]_{total} and P_R, marked by a filled red circle, represents the steady state. The green circle marks the P_L activity at the steady state.

of the two promoters and the total CI concentration as a function of free CI monomers. Fig. 5.1 shows a typical example of the behaviour. Steady states are graphically identified where the lines of P_R activity (red line) and [CI]_{total} (broken red line) intersect, as according to the steady state condition Eq. 4.6. The P_R activity in the steady state may be read from the ordinate axes, as may the P_L activity (green line) for the corresponding CI concentration.

Inspection of the graph for P_R activity (continuous red line) reveal that P_R is open both at very low and at high CI concentrations but repressed at the intermediate concentration. This characteristic behaviour can be understood by noting that the graph represents equation 4.13 giving the P_R activity in a system where MOR is expressed from P_L under CI control. At low [CI], there is plenty of MOR due to high P_L activity, while at high [CI], no MOR is present. The P_R activity is repressed only at intermediate [CI] because both MOR and CI are necessary for its repression by means of the MOR:CI:DNA complex formation.

Only one intersection point is observed between the curves of P_R activity and total CI concentration, showing that there is only one steady-state solution. This uniqueness of steady state holds true for any given value of the parameters because the activity of P_R given by Eq. 4.13 never increases faster than proportional to CI, whereas [CI]_{total} by Eq. 4.14 always increases faster or proportional to [CI]. Therefore, bistability is never realized in Model A with the assumption of binding of one MOR and one CI for repression of P_R.

A variant of Model A

Bistability may be obtained in Model A if a larger number of proteins are assumed to form the heteromer at O_M. Suppose m MOR monomers and c CI monomers bind at O_M to form the MOR_m:CI_c:DNA complex that represses transcription from P_R, then the expression for the P_R activity, Eq. 4.13, should be replaced by

\[ p_R([MOR], [CI]) = p_{R0} \cdot \frac{1}{1 + [MOR]^m[CI]^c/(K_{OM})^{m+c}} \]  

while all the other equations remain the same. For the variant of Model A to have more than two solutions to the steady state condition the slope of the
5.1 Model A - Direct binding at O

Figure 5.2: Model A with (a) one CI monomer and two MOR monomers or (b) two CI monomers and two MOR monomers binding at O_M. Promoter activities of P_L and P_R represented by the green and red curve, respectively and [CI] total in dashed red line, all as function of free [CI]. Filled circles mark steady state promoter activities while the open circle represent the unstable steady state. Using the steady state activities the repression folds P_L(open)/P_L(closed) and P_R(open)/P_R(closed) in both cases amount to approximately 1000 and 200, respectively.

Increasing part of the curve of P_R activity must exceed the slope of total CI concentration given in equation 4.14. The largest slope of the P_R activity as a function of [CI] is 2m – c in the logarithmic scale. Since the slope of the plot of [CI] total given by Eq. 4.14 is between 1 and 2, the steady-state Eq. 4.6 can have more than two solutions with Eq. 5.1 when 2m – c ≥ 2. In the case of m = 2, we could obtain multiple solutions with c = 1 or 2, that is, two MOR monomers binding together with one or two CI monomers at O_M. Examples for Model A with (m, c) = (2, 1) and (2, 2) are shown in Fig. 5.2. In each example, the intersections between the continuous red line (the P_R activity) and the broken red line (the total CI concentration) represent three steady-state solutions.

Stability analysis of steady states.

In figures 5.1 and 5.2 the steady states are marked by open or filled circles, identifying the stability of the given steady state as either unstable or stable, respectively. The stability may be recognized in the following way, using Fig. 5.2 as example. For the unstable steady state in the middle: If the CI monomer concentration increases by fluctuation from the steady value of [CI], the CI production from P_R will increase more than the increase in degradation given by [CI] total as the continuous red line of the P_R activity goes above the broken red line of [CI] total upon increasing [CI] from the middle steady state. This means that such a fluctuation causes further increase of [CI]. Similarly a small decrease in [CI] from the steady value, is enhanced as the degradation in that case is larger than the production. Consequently, the state is driven out of the steady state. On the other hand, the steady states at both ends represent stable
states: A fluctuation towards larger [CI] from these states leads to insufficient CI production in comparison with the CI degradation. That is, the continuous red line goes under the broken red line as [CI] increases. Oppositely the degradation curve falls below the production curve for decreases in CI. Both scenarios bring the system back to the original state; therefore, the states at the ends are stable. Thus, the system has two stable steady states, which allow for bistability.

We identify the two steady states as the immune and anti-immune states of the system. The state at the right represents the immune state with open P_R and repressed P_L, while the one at the left represents the anti-immune state with open P_L and repressed P_R. With the graphical representation of Fig. 5.2 we determine the promoter activity in each of the two stable states as previously described. This allows us to estimate the repression folds for P_L and P_R between the immune and anti-immune state in the examples of Fig. 5.2 to approximately 1000 and 200, respectively. The folds should be compared with the promoter activities obtained from the in vivo measurements. To recapitulate, in the modified switch system P_L is repressed approximately 1000-fold in the immune state and P_R is repressed approximately 100-fold in the anti-immune state [81].

5.2 Model B - Formation of a heteromer in the cytoplasm

In Model B we consider the possibility that CI and MOR form a complex in the cytoplasm before binding to the DNA and repression of P_R (illustrated in Fig. 4.2). As for Model A we subject Model B to a graphical inspection in order to estimate the promoter activity repression folds for comparison to experiments. Fig. 5.3 shows a typical example of P_R and P_L activities (red and green line, respectively), and [CI]_{total} (red broken line), all as function of free CI monomer concentration. Comparing the graphs to those of Model A a striking difference is that in Model B the behaviour of [CI]_{total}, now given by equation 4.18, can be non-monotonic. The behaviour comes from the competition for the free CI between MOR:CI and CI_2, see lower frames of Fig. 5.3. When the free CI concentration approaches the dimerization limit, CI_2 are formed and start repressing P_L, leading to the depletion of MOR. A low concentration of MOR means less CI bound in the heteromer even at higher concentrations of CI free. By Eq. 4.19 [CI]_{total} thus decreases until dimerization starts dominating.

Within a certain range of [CI]_{total} there thus exist three possible states with different concentrations of free CI. This suggests that the bistability could be obtained for the system with P_L regulated by CI, even if P_R were not regulated. Namely, even if P_R would produce CI at a fixed rate within the range of multivalued [CI]_{total}. In that case the bistability of the activity of P_L would come about due to the sequestration of CI into the MOR:CI heteromer; At low CI concentrations all CI is bound in the complex allowing full activity from P_L, while high CI values result in CI free to repress P_L. Such bistability is, of course, not the bistability observed in the experiments as activity from P_R would then not be bistable, but one can see that this feature of behaviour in [CI]_{total} (broken red line) makes it easier to have three intersections with the P_R activity curve (red line) than in the case of Model A.
5.2 Model B - Formation of a heteromer in the cytoplasm

Figure 5.3: Model B with MOR:CI binding at O_M. Again promoter activities of P_R and P_L are represented as red and green curves, respectively, while the dashed red curve is [CI]_{total}, all as function of free [CI]. Intersection between [CI]_{total} and P_R are steady states marked by filled circles for stable states and open circles for unstable states. In Model B the relative bare promoter activities affect the system's behaviour due to the interaction between MOR and CI, and plots are thus presented for different ratios between pL_0 and pR_0. In the left plot the bare activities of the promoters are equal and the repression folds pL(open)/pL(closed) and pR(open)/pR(closed) are 10,000 and 50, respectively. In the right plot pL_0 = 10 · pR_0 and the repression folds pL(open)/pL(closed) and pR(open)/pR(closed) are 60,000 and 100, respectively. Lower frames show the ratios of MOR units in the forms of monomer and heteromer, and the ratios of CI units in the form of monomer, dimer, and heteromer. One can see that MOR:CI and CI_2 compete for the free CI units in the intermediate concentration range of [CI].

According to the stability criterion discussed in relation to Model A, the steady states at both ends are stable while the state in the middle is unstable even for the system where both P_R and P_L are regulated. Still, a full dynamical stability analysis shows that there are some cases where the states at both ends can be unstable. In these cases the stability is dependent both on the dynamics of [CI]_{total} and P_R in response to changes in free CI, and the relation between the...
degradation rates of CI and MOR. However, the stability criterion is correct for most cases and in the analysis of the bistability we rely on the stability criterion, ignoring the small possibility that the states at both ends could be unstable.

A variant of Model B

In the preceding case of Model A a variation in the numbers of MOR and CI bound in the heterocomplex was a necessity for the model to show bistability. This is not the case for Model B, but increasing the number of interacting molecules in the heteromer complex to one CI monomer and two MOR monomers improves the model with respect to the parameter ranges that will reproduce the experimental promoter activity repression folds.

Fig. 5.4 shows a typical example of the promoter activities given by this variant of the model. As for Model A, introducing a larger heterocomplex result in a steeper decent of \( P_R \), making it more easy to obtain more than one intersection between the curves of \([\text{CI}]_{\text{total}}\) and \( P_R \) activity as function of free CI concentration.

5.3 Model A vs Model B

An automated graphical inspection for bistability at different combinations of parameters allow us to test more than \(10^6\) parameter sets for each of the models and their variations. For each tested parameter set we looked for bistability and compared the obtained repression folds with the promoter activities obtained from the \textit{in vivo} measurements [81]. In evaluating the results we use the criterion that the \( P_R \) repression fold should be in the range from 50 to 200 (50 < \( P_R(\text{open})/P_R(\text{closed}) \) < 200) and the \( P_L \) repression fold should at least be 500 (\( P_L(\text{open})/P_L(\text{closed}) \) ≥ 500).

To capture these repression folds in Model A with \((m, c) = (2, 1)\) and \((2, 2)\), we test the three parameters, \( K_{\text{C}12} \), \( K_{\text{OL}} \), and \( K_{\text{OM}} \), representing the dimerization constant of CI, the effective binding constant of CI monomer at OL, and the binding constant of the MOR:CI complex at OM, respectively. Fig. 5.5 shows the distributions of accepted values of parameters out of randomly chosen values in the logarithmic scale. \( K_{\text{OL}} \) and \( K_{\text{OM}} \) are narrowly distributed while \( K_{\text{C}12} \) are larger than \(10^{-1}\) for \((m, c) = (2, 1)\) and larger than \(10^{-2}\) for \((m, c) = (2, 2)\) in the unit of CI concentration at full activity of \( P_R \). One can also see that the accepted values for \( K_{\text{C}12} \) are much larger than those for \( K_{\text{OL}} \). This suggests that, in order for Model A to work, CI must exist as a monomer and act by cooperative binding to form CI2 at OL when repressing \( P_L \).

Fig. 5.6 shows the parameters that satisfy the \( P_R \) repression fold criterion versus the corresponding \( P_L \) repression fold for Model A. The vertical green lines are drawn at \( P_L \) repression fold 500; thus, only the sets on the right side of the lines should be accepted by the repression fold criterion. Comparing figure 5.5 and 5.6 we see that the relatively high values for \( K_{\text{C}12} \) in Model A come from the requirement for the large \( P_L \) repression fold. This can be understood as follows: In order to achieve a large repression fold for \( P_L \), the difference in \([\text{CI}]\) for the
5.3 Model A vs Model B

Figure 5.4: Model B with MOR$_2$:CI binding at O$_M$. Again promoter activities of P$_R$ and P$_L$ are represented as red and green curves, respectively, while the dashed red curve is [CI]$_\text{total}$, all as function of free [CI]. Intersection between [CI]$_\text{total}$ and P$_R$ are steady states. States are marked by filled circles for stable and open circles for unstable. In Model B the relative bare promoter activities affect the system behavior due to the interaction between MOR and CI, and plots are thus presented for different ratios between $pL_0$ and $pR_0$. In the left plot the bare activities of the promoters are equal and the repression folds $pL(\text{open})/pL(\text{closed})$ and $pR(\text{open})/pR(\text{closed})$ are 1000 and 50, respectively. In the right plot $pL_0 = 10 \cdot pR_0$ and the repression folds $pL(\text{open})/pL(\text{closed})$ and $pR(\text{open})/pR(\text{closed})$ are 1000 and 60, respectively. Lower frames show the ratios of MOR units in the forms of monomer and heteromer, and the ratios of CI units in the form of monomer, dimer, and heteromer. One can see that MOR:CI and CI$_2$ compete for the free CI units in the intermediate concentration range of [CI].

Two steady states should be large (increasing the horizontal distance between the green dots of figure 5.2). This, in turn, requires smaller slope in the [CI]$_\text{total}$ curve, namely, larger $K_{CI_2}$, because the slope in [CI]$_\text{total}$ changes from 1 to 2 around [CI] $\approx K_{CI_2}$, see Fig. 5.2.

In the columns of figure 5.7, the parameters that give P$_R$ repression fold in the range [50,200] are plotted versus the resulting P$_L$ repression fold for Model B.
Results from modelling a modified model system

Figure 5.5: Distribution of possible parameters for model A with \((m, c) = (1, 1)\) and \((2,1)\), and \(p_{L0} = p_{R0}\). Plotted parameters are those that satisfy \(50 < p_{R(open)}/p_{R(closed)} < 200\) and \(p_{L(open)}/p_{L(closed)} \geq 500\) out of \(10^6\) tested parameter sets. Tested parameters were randomly chosen from the range \(\tilde{K}_{OL}, \tilde{K}_{OM},\) and \(K_{CI2} \in [10^{-7}, 10]\). The tested parameter sets were chosen randomly over the range of \([10^{-7},1]\) for \(K_{CI2}, \tilde{K}_{OL}, \tilde{K}_{OM}, K_{MOR,CI}\). One can see that a broad range of parameter sets satisfy the repression fold criterion for \(P_R\). The resulting repression folds for \(P_L\) are limited to the region larger than 5000 for \((m, c) = (1, 1)\), but 50 for \((m, c) = (2, 1)\).

### 5.4 Relating parameters to reality

In order to asses the validity of the models based on the parameter restrictions

<table>
<thead>
<tr>
<th>((m, c))</th>
<th>((2,1))</th>
<th>((2,2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\tilde{K}_{OL})</td>
<td>(\sim 10^{-2})</td>
<td>(10^{-3} \sim 10^{-2})</td>
</tr>
<tr>
<td>(K_{OM})</td>
<td>(10^{-1})</td>
<td>(\sim 10^{-3})</td>
</tr>
<tr>
<td>(K_{CI2})</td>
<td>(\geq 10^{-2})</td>
<td>(\geq 10^{-2})</td>
</tr>
</tbody>
</table>

Table 5.1: Accepted ranges of parameters for Model A. Parameters are given in units of the concentration that corresponds to \([CI]\) and \([MOR]\) at full activity of the promoter \(P_R\) and \(P_L\) respectively.
5.4 Relating parameters to reality

Figure 5.6: Distribution of possible parameters versus $P_L$ repression fold; $pL(open)/pL(closed)$, for Model A with $(m,c) = (2, 1)$ and $(2, 2)$, and $pL_0 = pR_0$. The plotted parameters are those that satisfy the $P_R$ repression fold criterion $50 < pR(open)/pR(closed) < 200$ out of randomly chosen parameters from the region $\tilde{K}_{OL}$, $\tilde{K}_{OM}$, $K_{CI} \in [10^{-7}, 1]$. Vertical green lines are drawn at $P_L$ repression fold 500, thus only the parameters that are close to the lines are consistent with the experimentally obtained repression folds. $10^6$ parameter sets were tested.

summarized in table 5.1 and 5.2, we consider the unit of the parameters given as the protein concentration at full promoter activity. We initially rescaled the promoter activities by the degradation times, thus measuring the activities in units of protein concentration in steady state. That is, $P_R$ activity is measured in units of total CI concentration in the steady state of full $P_R$ activity, while $P_L$ activity is given in terms of total MOR concentration in the steady state of full $P_L$ promoter activity. As a result we measure $[CI]$ in units of the total CI concentration in the steady state at full activity of $P_R$.

This concentration is estimated as follows: First, we assume that the total concentration of CI in cells carrying the plasmid version of the wild type switch of TP901-1 in a single copy is comparable to the 300 nM $\lambda$-repressor concentration found in the lysogenic state of the wild $\lambda$ phage [92]. Next, we multiply this by the following two factors: 10 for the copy number of the plasmid used in the
Results from modelling a modified model system

Figure 5.7: Distribution of possible parameters versus $P_L$ repression fold; $p_L(open)/p_L(closed)$ for Model B with $(m,c) = (1,1)$ and $(2,1)$, and $pL_0 = pR_0$. The plotted parameters are those that satisfy the $P_R$ repression fold criterion $50 < p_R(open)/p_R(closed) < 200$ out of randomly chosen parameters from the region $\tilde{K}_{OL}$, $\tilde{K}_{OM}$, $K_{CI}$, $K_{MORCI} \in [10^7,1]$. Vertical green lines are drawn at $P_L$ repression fold 500, thus only the parameters that are close to the lines are consistent with the experimentally obtained repression folds. $10^6$ parameter sets were tested. The effective affinity $\tilde{K}_{OM}$ is defined by $(\tilde{K}_{OM})^{m+c} = \tilde{K}_{OM} \cdot (K_{MORCI})^{m+c-1}$.

Experiments and 100 for the relative activity of $P_R$ in our modified system in comparison with the wild-type switch [81]. With these factors, we estimate that $[CI]_{total}$ at the full $P_R$ activity in the present system, could be well over the $10^5$ nM scale.
5.4 Relating parameters to reality

<table>
<thead>
<tr>
<th>(m,c)</th>
<th>( p_{L0} = p_{R0} )</th>
<th>( p_{L0} = 10 \cdot p_{R0} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tilde{K}_{OL} )</td>
<td>( \lesssim 3 \cdot 10^{-3} )</td>
<td>( \lesssim 3 \cdot 10^{-4} )</td>
</tr>
<tr>
<td>( \tilde{K}_{OM} )</td>
<td>( \lesssim 5 \cdot 10^{-3} )</td>
<td>( \lesssim 2 \cdot 10^{-3} )</td>
</tr>
<tr>
<td>( K_{CI2} )</td>
<td>( \lesssim 3 \cdot 10^{-1} )</td>
<td>( &lt; 1 )</td>
</tr>
<tr>
<td>( K_{MORCI} )</td>
<td>( \lesssim 3 \cdot 10^{-1} )</td>
<td>( \lesssim 3 \cdot 10^{-3} )</td>
</tr>
</tbody>
</table>

Table 5.2: Accepted ranges of parameters for Model B. Parameters are given in units of the concentration that corresponds to [CI] and [MOR] at full activity of the promoter \( P_R \) and \( P_L \) respectively.

For Model A we found that the value of \( \tilde{K}_{OL} \) is between \( \sim 10^{-3} \) and \( 5 \times 10^{-2} \), corresponding to a at least 100nM with the estimated unit of CI concentration. This is in striking contrast to the in vitro estimate of 28 nM for the CI concentration at which the \( O_L \) is occupied for 50% of the time [45]. Another point to be taking from the limited parameters and the unit estimate is how the large repression fold of \( P_L \) enforce the restriction that \( K_{CI2} \gg \tilde{K}_{OL} \) for Model A. This means that CI exist as monomers in the cytoplasm and form \( CI_2 \) when they bind to \( O_L \), but this is in contrast with many phage-encoded repressor proteins, such as those encoded by phage lambda, 434, and 186, which tend to exist as dimers or higher oligomers in solution[51, 16, 77, 111]. Actually, most of the 434 and lambda repressors exist in the dimeric conformation at nanomolar concentrations[51, 77]. As described the investigation of the form of the TP901-1 CI repressor has revealed both dimers and larger structures, thus further leading to the dismissal of Model A as descriptive for the system. However, our Model A challenges the presumption that the formation of dimers is a prerequisite for its specific DNA binding.

We found broader distribution of parameter sets that satisfy the repression fold criterion for Model B. In particular, we did not find lower bounds for possible \( K_{CI2} \) in contrast to the case of Model A. In the comparison of the two variants of Model B, our results show that the model with the formation of \( MOR_2:CI \) complex is more favourable than that with \( MOR:CI \). For the latter, the \( P_L \) repression fold is always larger than 5000 for the parameters that give a \( P_R \) repression fold between 50 through 200. Such a high repression fold of \( P_L \) has never actually been observed in vivo. In contrast the model with \( MOR_2:CI \) formation predicts a limit on the resulting \( P_L \) repression fold as low as 50, which covers the observed range of the \( P_L \) repression fold.

In addition to the tested affinity parameters the models employ two parameters for the relative bare promoter activities, \( pR_0 \) and \( pL_0 \). So far the observations presented were made based on the assumption that \( pR_0 = pL_0 \). In the case of Model A the relative strength of the bare promoters does not affect the systems behaviour, such as bistability and repression folds, because there is no interaction between free MOR and CI in this model. In Model B, CI and MOR interact when forming the heterocomplex. In this way the expressions for promoter activities are statistical dependent and the relative promoter strength affects the system behaviours, such as bistability and repression folds, through the equation of total CI concentration, see Fig. 5.3 and 5.4. However, the ranges of parameters that allow for bistability with the experimental repression folds
Results from modelling a modified model system

is mostly unmarked by changes in the relative bare promoter strengths as seen when comparing the acceptable parameter ranges listed in table 5.2. With the above study of the modified switch system we see that the biological description can in fact reproduce bistability and account for the observed repression folds. Next we turn to testing the models in studying the full system including all of the known CI operator sites.
Chapter 6

Results from modelling the full switch system

This section highlights the results of the publication included in appendix B.

The study of the modified switch fragment containing just the O_L site for CI binding allowed us to focus on the mechanisms describing the repression of P_R by CI and MOR. When considering the full system some additional reflections must be made on the self-repression of CI by binding at O_R. We thus initially disregard MOR and only consider how CI inflict this self-repression through cooperative binding at the three known operator sites. Subsequently we re-introduce MOR and compare the model predictions to experimental measurements of P_R activity at different constitutive levels of CI. The combination of model simulations and experimental data, lets us dismiss the model of P_R repression with interactions of CI and MOR restricted to the DNA and indicate an incompleteness of the biological model.

6.1 Experimental data

The promoter activities predicted by the models are compared to activity measurements made on two genetic systems. The systems are based on the minimal switch fragment isolated by Pedersen and Hammer [81] with CI introduced in trans from a synthetic promoter library. Fig. 6.1 show sketches of the systems together with the averaged measurements.

The P_R promoter activity in the absence of MOR, was measured using plasmid pMAP86. pMAP86 contains the genetic switch fragment with frame-shift mutations in cI and mor and a LacLM reporter gene downstream of P_R. CI was introduced in trans using a library of synthetic constitutive promoters in front of cI, succeeded by the reporter gene pepI. The synthetic library-cI-pepI fragment was integrated on the bacterial chromosome in single copy [118, 15]. The promoter activity given by the LacLM activity is plotted against the CI expression determined by the PepI activity in Fig. 6.1 A. A 25-fold decrease in P_R activity is observed over an ~600-fold change in CI expression. The decrease
Results from modelling the full switch system

Figure 6.1: Effect of constitutive CI expression on $P_R$ activity in the absence (A) or presence (B) of a functional $mor$ gene, and $P_L$ activity in the presence of $mor$ (C). Drawings to the left illustrate the components of the reporter system. Genes with a frame-shift mutation are shown in white. Inverted gray arrows and white box indicate operator sites. LacLM and PepI activities are measured in nanomoles of processed Ortho-Nitrophenol-$b$-galactoside or L-proline-para-nitroanilide, respectively, per minute per milliliter of cell extract per OD$_{280}$. Each data point corresponds to a different synthetic promoter and is an average of at least two independent activity measurements. Error bars are standard deviations of the mean of LacLM activity. Insets in A and B show close-ups of PepI activity region 1100 (note the different ordinate values).

in activity apparently contains a plateau stretching over a 10-fold increase in CI concentration where the $P_R$ activity is held at approximately half of its full activity.

A second plasmid, pMAP88, was used to measure the activity of $P_R$ in the presence of MOR at various CI concentrations. pMAP88 contains the switch fragment with a frame-shift mutation in $cI$ and with the $P_R$ activity reported by lacLM [82]. Again CI was introduced from a synthetic library in single copy on
6.2 MOR independent repression of P\textsubscript{R}

...the bacterial chromosome and monitored by the reporter PepI (Fig. 6.1 B, left). In the presence of MOR P\textsubscript{R} is repressed 500-fold as CI concentration increases 600-fold, but the repression is not observed until the last 10-fold increase in CI expression occurs. Thus, the presence of MOR renders the P\textsubscript{R} promoter open, even at CI concentrations at which the promoter was partially repressed in the absence of MOR.

In addition to the studies on P\textsubscript{R} promoter activity, a smaller sampling was made to examine the response of the P\textsubscript{L} promoter to different levels of CI in the presence of MOR. Plotting the P\textsubscript{L} activity (LacLM activity) against the CI expression (PepI Activity) confirms that mor is indeed transcribed in the strains for CI expression beyond levels at which the P\textsubscript{R} promoter is closed (Fig. 6.1C, and insets in A and B).

The linear correlation between the PepI activity and CI concentration in the systems was seen by Western blotting performed on a selection of the assay samples. A linear least-square fit to a plot of PepI activity versus CI concentration in units of the CI concentration in a system containing a wild-type switch in the immune state (i.e., the WIU) was used to rescale the experimental PepI activity units, to units of total CI concentration in the immune state before comparing model and experimental data.

I constructed the genetic systems and performed the activity measurements as part of my master thesis. A full description of materials and methods is given in the appended paper in appendix B [2].

6.2 MOR independent repression of P\textsubscript{R}

Eq. 4.21 gives the P\textsubscript{R} promoter activity in units of the bare promoter activity, p\textsubscript{R}0, as function of the free CI concentration, while the experimental data show the relation between promoter activity in units of lacLM activity and total CI concentration given in units of [CI]\textsubscript{total} in a system containing a plasmid with the wild type switch devoted to the immune state (WIU). To address this discrepancy we calculate the total CI concentrations by Eq. 4.14 and P\textsubscript{R} activities by Eq. 4.21 with p\textsubscript{R}0 in units of the LacLM activity from the bare promoter activity, for a range of free CI concentrations. We then compare P\textsubscript{R}([CI]\textsubscript{free}) vs [CI]\textsubscript{total}([CI]\textsubscript{free}) to the experimental data set. For each tested parameter set we quantify the comparison by an unweighted sum of squared residues (SSR). Fig. 6.2 shows the result of an automated parameter testing, scanning each of the six energies in the range [−15; −3] kcal/mol. Included in the figure are simulations based on four different parameter sets taken from the scan to illustrate the clear decrease in fitness between data and simulations as SSR increases.

A note should be made on the units employed in testing the model of the full system against data. Apart from scaling the predicted activity probabilities by p\textsubscript{R}0 (the basal promoter activity level in units of LacLM activity), we assign an absolute value to the units of total protein concentrations. The CI concentration in cells containing the wild-type switch in the immune state is set to approximately 10\textsuperscript{−6} M, based on a preliminary rough estimate of TP901-1 CI in vivo concentration made by Margit Pedersen (personal correspondence).
Results from modelling the full switch system

Figure 6.2: Fit between parameters and data. Left column show SSR values as function of two parameters of the 6-dimensional sets. Each of the six energies where randomly chosen in the range $[-15; -3]$ kcal/mol. $[\text{CI}]_{\text{total}}$ were fixed to $10^{-6}$M at full $P_R$ activity. Blue dots are from parameters chosen independently, only parameter sets of $SSR < 140$ are included. The darker the blue the lower the SSR value and hence the better the fit. Red dots show parameters fulfilling the restrictions $\Delta G_R = \Delta G_L + \ln(2000/28) \cdot k_B T \pm 0.5$ kcal/mol and $SSR < 140$. Stars are included to provide a mean of following individual parameter sets and their resulting simulations in the right side graphs: Black star: Best fit, $SSR = 61$. Gray star: $SSR = 140$. White star: $SSR = 200$. Magenta star: $SSR = 62$, fit fulfils restrictions similar to the red dots. Yellow star: values of plot in Fig. 6.3. Orange star: values calculated based on experimental values of Johansen et al., 2003 [45]. Full orange line indicates minimum value predicted by Model A with $(m, c) = (2, 1)$ and dashed orange lines indicate outer boundaries predicted by Model B with $(m, c) = (2, 1)$ from previous chapter [76]. $\Delta \tilde{G}_R$, $\Delta \tilde{G}_L$, and $\Delta \tilde{G}_D$ are the effective binding energies, defined as $\Delta \tilde{G}_{R,L,D} = (\Delta G_{R,L,D} + \Delta G_{C12})/2$. 

48
6.2 MOR independent repression of $P_R$

Figure 6.3: $P_R$ activity vs $[CI]_{total}$ in the absence of MOR. A) Experimental results and simulated (red) curve based on the parameter set indicated by a yellow star in Fig. 6.2 ($\Delta G_{CI2} = -3.6$ kcal/mol, $\Delta G_{OR} = -13.0$ kcal/mol, $\Delta G_{OL} = -14.4$ kcal/mol, $\Delta G_{OD} = -13.1$ kcal/mol, $\Delta G_2 = -8.5$ kcal/mol, and $\Delta G_3 = -8.9$ kcal/mol). B) Binding configuration fractions constituting the $P_R$ activity curve presented in A). Sketches below figure illustrate the involved binding configurations between CI (black rods) and operator sites.

This concentration is comparable to the previous estimate of CI total concentration in cells containing the modified switch system in the immune state when keeping in mind the 100 fold decrease of $P_R$ activity in the presence of $O_R$. We use the estimate to address the biological plausible range of binding energies to be tested. A lysogenic total CI concentration of $10^{-6}$M is match by a CI binding energy of order $\sim k_B T \cdot \ln(10^{-6}) \sim -8$ kcal/mol. Binding energies that are weaker than this ($\Delta G > -8$ kcal/mol) imply that the corresponding binding rarely occurs at standard lysogenic concentrations. Similar to the previous model approach the concentration estimate simply helps us in validating the model, all qualitative results taken from the model are independent on the estimates.

A visual inspection of simulations based on parameters of increasing SSR values sets a $SSR \sim 90$ cut-off for reproducing the plateau indicated in the data of $P_R$ activity vs $[CI]_{total}$ in the absence of MOR. Compare plots with black and gray curves in Fig. 6.2. Inflicting this cut-off restrict $\Delta G_{CI2} > -9$, where as the remaining parameters show the same dispersion as for the $SSR < 140$ cut-off showing that the plateau is readily reproduced for a wide range of pa-
Results from modelling the full switch system

rameter values. Considering the binding configurations at play at increasing CI concentrations a general pattern arises. At low CI concentrations, the promoter is open because the species of no bound CI dominates. With increasing concentration, the species of two cooperatively bound CI dimers at OR-OL or OR-OD dominate together with the species of two dimers cooperatively bound at OL-OD (species C, D, and E in Fig. 6.3). Together, these binding species constitute the approximately half-repressed level of PR because only species E allows activity from PR. At high CI concentrations, all of the operator sites are covered by bound CI dimers in either pair- or triple-wise cooperative bindings, resulting in full repression of PR. Whether species C or D will balance species E depends on the relation between the probabilities of the three states. This relation is effectively given by the binding strengths assigned to the three sites OR, OL, and OD.

We find that the two CI operator sites assigned the weakest bindings are always similar in binding energy, and OR is always found to be one of the weaker sites. Thus, even without restricting the dissociation constants of CI binding at OR and OL, the model predicts a relative weakness of CI binding to the OR site.

6.3 Experimental affinities and previous parameter restrictions

The binding sequences for CI and the affinities of CI to the operator sites were determined by Johansen et al. [45]. Apart from the previously mentioned, 28 nM reported for the binding affinity of CI to OL, the affinity between CI and OR was determined to > 2000 nM. The experimental dissociation constants include both dimerization and DNA binding [45], however, they do relate the binding energy parameters that describe CI dimer binding to OR and OL. Hence, we explore the tested parameters under the restriction that \( \Delta G_R \sim \Delta G_L + \ln(2000/28) \cdot kBT \).

In addition we note that the sequence of the OD site is similar to the consensus CI-operator site and that preliminary gel mobility shift assays have shown that CI binds with the same affinity to OD as to OL (M. Pedersen, unpublished data). Lending us the possible restriction \( \Delta G_L \sim \Delta G_D \).

In Fig. 6.2 the parameter sets fulfilling the relation between \( \Delta G_R \) and \( \Delta G_L \) are plotted in increasing intensities of red corresponding to their SSR. It is clear that this restriction does not conflict with the models ability to fit the data. However, with the additional restriction that the affinity of CI binding at OD should be similar to CI binding at OL, the fit between experimental data and simulations largely reduces and the reproduction of the indicated plateau vanishes, even though the simulations still presents acceptable data fit, see green dots in Fig 6.4.

The parameter restrictions predicted in modelling the modified system are included in the plots of \( \Delta G_{CI} \) and \( \Delta G_{OL} \) (the correspondents to \( K_{CI} \) and \( K_{OL} \)), in Figs 6.2 and 6.4, using the relation in Eq. 4.23. The parameter predictions of the two approaches show acceptable agreement, especially noting that the estimate of [CI]total \( \sim 10^5 \text{nM} \) in the immune state of the modified system is considered a minimum value, indicating that the restrictions could shift towards lower values.
6.3 Experimental affinities and previous parameter restrictions

Figure 6.4: Fit between parameters and Data. Left column show SSR values as function of two parameters of the 6-dimensional sets. Each of the six energies were randomly chosen in the range [-15,-3] kcal/mol. [Cl]_{total} were fixed to $10^{-6}$M at full PR activity. Blue dots are from parameters chosen independently, Only parameter sets of SSR < 140 are included. The darker the blue the lower the SSR value and hence the better the fit. Green dots show parameters fulfilling the restrictions $\Delta G_R = \Delta G_L + \ln(2000/28) \cdot k_B T \pm 0.5$ kcal/mol, $\Delta G_D = \Delta G_L \pm 1$ kcal/mol, and SSR < 200. Stars are included to provide a mean of following individual parameter sets and their resulting simulations in the right side graphs: Black star: Best fit, SSR = 61. Gray star: SSR = 140. white star: SSR = 200. Green star: SSR = 141, fit fulfils restrictions similar to the green dots. Yellow star: values of plot in Fig. 6.3. Orange star: values calculated based on experimental values of Johansen et al., 2003. Full orange line indicates minimum value predicted by Model A with $(m,c) = (2,1)$ and dashed orange lines indicate outer boundaries predicted by Model B with $(m,c) = (2,1)$ from previous chapter [76]. $\Delta \tilde{G}_R$, $\Delta \tilde{G}_L$, and $\Delta \tilde{G}_D$ are the effective binding energies, defined as $\Delta \tilde{G}_{R,L,D} = (\Delta G_{R,L,D} + \Delta G_{Cl_I})/2$. 

51
6.4 MOR dependent repression of P\textsubscript{R}

In tread with the models studied in the context of the modified switch we consider two scenarios of MOR dependent repression of P\textsubscript{L}; with and without formation of a MOR:CI heteromer in the cytoplasm. When testing the models we fix the parameters obtained from the best fit between the model on MOR independent repression of P\textsubscript{R} and the experimental data (Fig 6.2).

6.5 Model A - Discarding direct binding at O\textsubscript{M}

Initially we consider Model A where repression of P\textsubscript{R} in the anti-immune state is simulated by direct binding of CI and MOR at a speculated operator site, O\textsubscript{M}. This model does not allow CI and MOR to interact before DNA binding occurs and the binding assigned to the CI:MOR:O\textsubscript{M} association thus represents both binding and cooperativity. We found that such a model cannot concurrently fit the two data sets in Fig. 6.1A and B.

For all CI concentrations, the curve of model A falls at or below the P\textsubscript{R} activity curve predicted by the model of P\textsubscript{R} repression by CI dimers alone. This is illustrated by the best fit of Model A to the experimental data shown as the dashed gray curve in Fig. 6.5A. In fact, the dismissal of Model A can be interpolated directly from the experimental data without the aid of simulations. The presence of MOR renders P\textsubscript{R} open at CI concentrations that repress P\textsubscript{R} in the absence of MOR. Hence, simply including MOR as an extra P\textsubscript{R} repression effect, as in Model A, naturally will only result in further repression, not the observed release of repression.

6.6 Model B - Formation of a heteromer in the cytoplasm

Instead we focus on Model B that includes sequestration of CI into a complex with MOR before P\textsubscript{R} repression by DNA binding. Modelling the sequestration of CI in a complex with MOR reveals a hysteresis in the system related to the multivalued function of [CI\textsubscript{total}] seen for the modified system. The hysteresis is clearly illustrated in Fig. 6.5B in the relation between free CI and total CI concentration: High initial concentrations of MOR, as is the case at low [CI\textsubscript{total}], lead to sequestration of CI and thus a low free CI concentrations is maintained at increasing [CI\textsubscript{total}] (free CI concentrations follows the red arrow of Fig. 6.5B). In contrast, starting at high initial [CI\textsubscript{total}] the initial MOR concentration should be low due to P\textsubscript{L} repression by CI\textsubscript{2} and free CI is only depleted by dimer formation. The continued repression of MOR expression by the CI dimers result in high levels of free CI even at lower [CI\textsubscript{total}] (free CI concentration follows the blue arrow of Fig. 6.5B). In the experimental strains, the steady state is established from a state of fully active P\textsubscript{L} promoter and consequently a high MOR expression level. To make the calculations of the unweighed SSR we therefore choose to consider the distance to the fully drawn branch of the P\textsubscript{R}
curve in Fig. 6.5A corresponding to the free CI following the red arrow in Fig. 6.5B.

For the four simulations with acceptable experimental data fit presented in Fig. 6.6 the binding energy between MOR and CI is \( < -11.2 \text{ kcal/mol} \) corresponding to a dissociation constant of \( \lesssim 10 \text{ nM} \). This imply that, within the model, CI is sequestered in the complex at concentrations 100-fold lower than the assumed lysogenic CI concentration. In fact this is the case for all acceptable parameters of Model B. The upper limit of \( \Delta G_{MOR,CI} \) by Model B with \((m,c) = (2,1)\) falls
below the range predicted based on the modified system, included as dashed orange lines in the top plots of Fig. 6.6. However, the boundaries were based on an assumed upper limit on the P_L repression fold, and bistability was observed for a wider range of K_{MOR,C1} values. Compare Fig. 6.6 to 5.7. Inspection of similar plots of acceptable pairs of \( \Delta G_{MOR,C1} \) and \( \Delta G_{OM} \) for Model B, but with \( pL_0 = 10 \cdot pR_0 \), reveals clearer correspondence between parameter ranges of the two theoretical approaches. With the relative stronger P_L promoter, the model produces acceptable fit for weaker binding between CI and MOR.

The binding energy of the complex to DNA is bounded by a lower limit of \( \sim -9 \) kcal/mol, indicating that this binding could be omitted from the model without largely affecting the fit. Instead, the best of the simulations assign the observed P_R repression to CI dimer binding in both the absence and presence of MOR. However, the experimental data show a 20-fold stronger P_R repression level in the presence of MOR than in its absence (see insets in Fig. 6.1 A and B). This inconsistency between the simulations and the data reveals the incompleteness of the biological model.
6.6 Model B - Formation of a heteromer in the cytoplasm

Figure 6.6: Distribution of possible $G_{OM}$ and $G_{MOR:CI}$ values for Model B with remaining parameters fixed to the best fit values of Fig. 6.2. The darker the blue the lower the SSR value of the corresponding parameters and hence the better the fit. Only parameter sets of $SSR < 1300$ are included. Left column present the case of $(m,c) = (1,1)$, in the right column $(m,c) = (2,1)$. Dashed orange line in left plot is upper limit of $G_{MOR:CI}$ from modelling the simplified system by Model B $(m,c) = (1,1)$. In right column dashed orange lines are corresponding outer boundaries predicted by Model B $(m,c) = (2,1)$. Lower graphs present examples of $P_R$ activity as function of total CI concentration alongside the experimental data for $G_{OM}$ and $G_{MOR:CI}$ pairs indicated by stars in upper plot. Simulations are presented with decreasing fit to the experimental data from top to bottom. Left side: Black star: Best fit, $SSR \sim 620$. Gray star: $SSR \sim 800$. White star: $SSR \sim 1230$. Right side: Black star: Best fit, $SSR \sim 840$. Gray star: $SSR \sim 1000$. White star: $SSR \sim 1270$. 
Results from modelling the full switch system
Chapter 7

Thermodynamic parameters and experimental attempts

The statistical mechanical approach to modelling the switch of phage λ, taken by Shea and Ackers [1, 110] calls for considerable knowledge on binding energies between involved proteins and proteins and DNA. In the framework of the λ phage many of these energies are given in the literature and thus provide valuable limits on the parameters of the model. For TP901-1 a very limited range of binding constants have been measured [45]. Moreover, only preliminary experiments have been made to determine the intracellular concentration of CI and MOR. All together this greatly reduce the constrains we may impose on the parameters of the TP901-1 model.

Experimental work

The large range of binding energy parameters, all resulting in acceptable outputs from the models, show that even though our modelling do allow us to make conceptual statements, additional restrictions on binding affinities or stoichiometry of protein complexes would very nicely supplement the appliance of models in confirming or disregarding biological models of the switch mechanisms. To this end I should rightfully mention my attempts on determining binding constant for CI biding at Od, and on verifying the MOR:CI interaction. I made these experimental attempts in the laboratory of Christian Cambillau at Architecture et Fonction des Macromolécules Biologiques, Université Aix-Marseille I & II, Campus de Luminy. In collaboration with Mariella Tegoni I tested CI binding at the known operator sites in a series of surface plasmon resonance (SPR) experiments and attempted to capture an interaction between CI and MOR in solution using online High-performance liquid chromatography (HPLC) and Multiple Angle Laser light scattering (MALS).

Unfortunately we had based the latter experiments on a genetically engineered MOR species, with a tag of six histones. It turned out that this his-tag renders the protein unstable and most probably non-functional in its purified form, though able to restore bistability in the switch in vivo. The SPR experiments
showed a high dependency on buffer contents and due to the limited time of my stay I did not succeed in replicating the known binding affinities between CI and OR and OL. Dissociation constants measured by SPR have previously been shown to be lower compared to other methods [10, 38]. In the reported cases the differences was only 10-20 fold, while my measurements differed by 600 fold compared to the gel-retardation results [45]. Consequently I did not have the time to test binding at OD. Thus, even though very educational, my experimental work in Marseille did not result in any additional implications for the models.
Chapter 8

Concluding remarks and perspectives

It is remarkable that the modified switch fragment containing just one operator site, \(O_L\), preserves the bistability of the full system considering that the modified switch system may very well contain 100 times more CI molecules in its immune state than the wild type genetic switch. Our model study suggests that the robustness of the genetic switch in TP901-1 is brought about by sequestration of CI through MOR:CI complex formation in the cytoplasm.

Extending the model to describe the switch system with all of the known operator sites strengthens the biological model on MOR independent repression of \(P_R\). Within reasonable fit to the experimental data the model predicts that CI binding at \(O_R\) should be weaker than the strongest binding at \(O_L\) or \(O_D\), consistent with the experimental established relation between dissociation constants for CI binding at \(O_R\) and \(O_L\).

In contrast a comparison between the best fit of the extended model and experimental data on \(P_R\) activity in the presence of MOR, reveal an inconsistency in the level of repression at high CI concentrations. In the experimental data the presence of MOR leads to a 10 fold stronger repression of \(P_R\) than observed in absence of MOR, whereas the model assigns the same repression level in the two states. This discrepancy between model and the experiments indicate that something is missing in the biological understanding of the switch. On a speculative note, one might imagine that the high initial pressure of MOR in the experimental strain could result in the aggregation of many MOR to each CI rendering the complex unable to bind to DNA and repress \(P_R\), while still sequestering CI and thus preventing \(P_L\) repression. Continued studies of the interaction between CI and MOR, may address the stoichiometry of an eventual complex and additionally the effect of a large excess of MOR.

While the described theoretical approaches pinpoints the need for CI sequestration the interaction between CI and MOR is yet to be verified experimentally and the switch of TP901-1 remains a puzzling twist on the established understanding of mechanisms observed in bistable temperate phages.
Part III

Olfactory Neuron Differentiation
Part III frontage illustration taken from www [93]
Chapter 9

Multistability in differentiation of olfactory neurons

The examples of bistable epigenetic traits in the previous chapters, whether being results of positive feedback between transcription factors or heritable marks on histones in the chromatin, resemble simple versions of Waddington's epigenetic landscape [135, 114]. Waddington illustrated differentiation of cells by a marble rolling down a hill with embedded cliffs leading to different valleys, see Fig 9.1. In this imagery the final valley represent a final differential state. On its way down the marble will encounter bifurcations. The chosen path will then depend on prior eventual perturbations shifting the marble from the channel base. The shifting nudge must be applied within a time span of “competence”. Applied to early it will disperse before the branching, applied to late it will have no effect on the chosen path. In the developing embryo such nudges resembles the action of embryonic inducing factors. Indicated in the theory is the irrevocable nature of the final differential state. The crests surrounding the final valley will prevent any reasonable nudge from shifting the marble to another valley. Within natural embryonic development this devotion is generally true, though recent advances have allowed scientist to reprogram the differentiation of mouse fibroblast to exhibit the properties of embryonic stem cells [123].

Though very apprehensible the epigenetic landscape most certainly can not be depicted as a simple two dimensional surface. Instead the landscape has a dimension for each of the components in the system. Collapsing this higher dimensional vector field onto two dimensions would mean that the landscape should change around the developing cell, as the branch chosen at a bifurcation could greatly influence the immediate valleys available at “downhill” developmental differentiations [136, 59]. In its full reach the epigenetic landscape describes the differentiation into multiple stable states.
Multistability in differentiation of olfactory neurons

Figure 9.1: C. H. Waddington’s epigenetic landscape from 1957 and a more recent version to emphasize the possibility of intermediate states. Each of the possible valleys represent a different final differential state of the developing cells. Figures adapted from [134, 59]

9.1 Olfactory system

In 2004 Richard Axel and Linda Buck received the Nobel prize for their discovery of the olfactory receptors (ORs) and the formulation of the pathway of odour sensing in mammals. In rodents like mice and rats, each olfactory neuron presents one out of more than 1000 highly similar olfactory receptors on its cilia in the epithelial of the nasal cavity. Interactions between odour molecules and these receptors activate a cascade of events, leading to the formation of an action potential in the neuron. This potential travels along the axon of the neuron to the olfactory bulb to transmit the information of the encountered odour. Connections between neurons and glomeruli in the bulb are receptor specific. With each neuron dedicated to expressing just one OR version, the receptor specific connection ensures the essential conversion of olfactory signals to a typographical map in the bulb, recognized by the brain as the encountered odour, see Fig. 9.2 [65, 95, 130, 20, 47, 29]. The best of human noses can detect in the order of 10,000 different odours. This vast range comes about as a combinatorial input to the brain of graded signals from activation by a given odour of different receptors to different degrees.

The olfactory epithelium (OE) in rodents is characterized by a zonal expression of receptors. Any individual receptor is confined to expression within a zone along the dorsal-ventral axis of the OE, with possible overlaps between differently segregated zones [80, 144, 128], see Fig. 9.3a-d. Throughout life the OE is constantly renewed by neuronal regeneration from stem cells at the base of the epithelium. In this process of regeneration a dying neuron is not necessarily replaced by a neuron expressing the same receptor as its dying predecessor. Instead it is believed that the choice of receptor is stochastic inflicted in each maturing cell. With approximated $\sim 5 \cdot 10^6$ sensory neurons in the rodent nose and almost 3000 allelic versions of receptor genes to choose from, mere statistics ensures that each receptor should be represented in the epithelium in an apparently random pattern. Accordingly, within a zone neurons expressing the same receptor gene form an apparently random pattern interspersed with...
9.1 Olfactory system

Figure 9.2: The olfactory system. Odour recognition is achieved through a series of steps starting with the association between odorant molecules and olfactory receptors at the cilia of the olfactory receptor cells. The olfactory receptor cell is activated by the association and sends an electrical signal up its axon to a glomeruli in the olfactory bulb [37]. Signals from many olfactory neurons expressing the same receptor are relayed in the glomeruli and transmitted to higher regions of the brain. The same odorant molecule may activate different olfactory receptor cells to different degrees. It is the combination of signals from multiple glomeruli that allow the brain to detect more than 10.000 odours. Figure adapted from Karolinska Institutet and Nobel Foundation, Stockholm, Sweden.

Olfactory receptor genes constitute a large superfamily of highly similar genes. In mammals the OR gene family comprises 900-1,400 genes and are found on almost every chromosome as shown for the mouse genome in in Fig. 9.3e [47]. Often OR genes are found in clusters of more than one gene but genes within a cluster need not be restricted to expression in the same zone and genes expressed in overlapping zones are not confined to the same locus [119].
Multistability in differentiation of olfactory neurons

Figure 9.3: Details of the olfactory epithelium and genes. a-d) Green fluorescence visualization of neuronal distribution of three OR receptors in whole mounts of mouse olfactory epithelium. Each OR is confined to expression in a zone along the dorsal-ventral axis of the OE, with possible overlaps between zones. Within each zone the expression of the individual receptors appears random in nature with no apparent correlation between neighbouring cells. Figure adapted from [131]. e) Chromosomal distribution of OR gene clusters in the mouse genome indicated in red. The large OR gene family cover ~5% of the mouse genome and clusters of OR genes are found in almost every chromosome. Figure adapted from [47].

diversity and wide chromosomal position of OR gene clusters are generally considered a result of subsequent tandem gene duplications, gene conversions and recombination events and even conversions of entire coding regions [33]. In humans different ethnic groups have between 7-11 copies of a OR cluster on chromosome 19 spread out at different chromosomes apparently as a result of variable duplication history [47].

Even though the coding region of the general OR gene is just ~1kb long [32], transgenes introduced in the chromosome of laboratory mice by transfection must comprise up to 11kb of an OR gene region in order to reproduce a stochastic expression pattern parallel to endogenous genes [131, 97, 78]. Establishment of the expression pattern and probability of endogenous mouse ORs from the MOR28 locus, even requires a cis-acting element positioned up to 200 kb upstream of the affected genes [28]. Intriguingly the OR gene family includes a considerable number of genes encoding non-functional ORs. In humans about 60% of the OR genes resembles such pseudogenes, while the number is closer to 20% in mice [141, 31]. In contrast to the exclusive expression amongst functional OR genes, pseudogenes may be co-expressed alongside a functional OR gene.

Models of OR gene devotion

Developing olfactory neurons are differentiated by the expression of a single olfactory receptor and the extension of its axon to the receptor defined glomeruli in the olfactory bulb. A crucial part of the differentiation is the continued
expression of the chosen OR gene. Potential switching to another OR gene after the axon-to-glomeruli connection at the bulb, would mess up the typographical map between odour encounters and recognition in the brain. Fig. 9.4 illustrates some of the hypotheses that have been put forth to describe this exclusive expression of a single gene out of the multiple available OR genes.

Deterministic models include activation of a single gene by OR specific transcription factors (TFs) or a combination of such. The identification of several binding sites of known TFs upstream of the olfactory genes, lends plausibility to this thesis [131, 97, 84]. However, elegant experiments with transgenic mice, transfected with two differently tagged MOR28 transgenes, demonstrated that exclusive regulation applies among transgenes identical in their coding and regulatory sequences [107]. Thus excluding trans regulation by OR specific regulators as the single source of gene selection. Also such models does not account for the seemingly stochastic nature of receptor choice. These concerns are circumvented in a "race" model where a random OR gene is activated by increasing numbers of TFs cooperatively binding at its promoter. For such a description to capture and maintain the expression of a single allele the supply of TF must be strongly limited and the cooperativity between bound TFs must be very strong [52].

Other models implement observations of enhancer elements. As mentioned for the MOR28 gene locus, such elements may reside far upstream of the OR genes and affect the expression probability of many of the OR genes in the downstream locus [24, 108, 57, 102]. It is proposed that DNA looping could bring the enhancer element and the promoter of a random OR gene from the locus close together, leading to activation of this specific gene similar to the features observed for the locus control region in human β globin [56]. However, the effect of such elements are yet to be identified for all OR genes of the super gene family [28].

DNA rearrangement is another beautiful solution to obtain singularity in gene expression as observed in the immune system where DNA recombination ensures diversity of B cells [23]. This hypothesis has been substantially tested by the generation of cloned mice from olfactory sensory neuron nuclei expressing a known OR [24, 55]. Olfactory epithelium from the cloned mice express the full OR repertoire, arguing against the DNA rearrangement hypothesis. By analogy, B-lymphocytes in animals cloned from mature B-cell nuclei are monoclonal as they express only one type of antibody [40].

Finally a line of hypotheses propose the involvement of chromatin modifications in regulation of OR gene selection [68, 47]. This may be supported by findings that OR genes are often associated with silent heterochromatic regions, such as subtelomeric regions of the human chromosome and repeat rich regions [98, 33]. Moreover, OR gene clusters have increased trimethylation of lysine 9 and 20 on histone H3 and H4, respectively. Both are histone modifications that are associated with gene silencing [64], while permissive histone modifications like methylation of lysine 4 and certain types of acetylation are only observed at low levels in OR gene clusters [71, 43]. Lastly histone de-acetylases (HDACs) are reported in immature and mature olfactory sensory neurons (OSNs) [101, 79], and in vitro cultured vomeronasal progenitor cells of adult rats only develop the adult neuronal phenotype when subjected to HDAC inhibitors [140].
Multistability in differentiation of olfactory neurons

Figure 9.4: Models of OR gene expression differentiation. The mechanisms range from deterministic activation by OR specific transcription factors to stochastic descriptions invoking DNA rearrangements, singular enhancer elements, or fuzzy feedback mechanisms. Figure adapted from [29]

T.S. McClintock encompass the epigenetic control in three subsequent levels; Allelic inactivation, Zonal exclusion, and Selection and Solidification [68]. All the OR genes are initially thought to have neutral nucleosome marks. Increasing repressive chromatin marks ensures inactivation of one allele of each OR locus and zonal dependent factors further exclude genes by chromatin silencing. Finally a random OR gene amongst the genes still carrying permissive marks, would capture the transcriptional machinery and initiate mRNA synthesis. Translation into a functional protein initiate a feedback mechanism that should ensure silencing of remaining actively marked OR genes.

In another model of OR regulation by epigenetic DNA markers, Kambere and Lane imagine a low basal level of activity from all the OR genes [47]. By chance one OR gene could reach "sub-threshold" activity levels which would initiate a feedback mechanism to permanently repress all other OR genes. It is argued that such a model would require that the selected OR locus is protected from the inhibitory process. To that end the enhancer elements that have been associated with some OR clusters are encompassed in the model as a mean of shielding the winning OR from the overall feedback repression. Both of the models combine deterministic factors such as spacial or temporal transcription factors to restrict the number of possible OR loci, but assign the final choice to a random process as in the stochastic models of Fig. 9.4.

Some of the models of OR devotion, like the limiting transcriptional complexes or singular enhancer elements, picture plausible ways of stochastically choosing a single OR gene for expression. However, the models resort to unspecified feedback mechanisms in order to maintain the memory of the chosen OR gene [131, 108, 24, 54, 112, 57, 102, 28]. The hypotheses that enrol chromatin modifications for upholding the OR gene selection still implement similar fuzzy feed-
back mechanisms at some stage in establishing the choice [47, 68]. In all cases some mechanism is expected to shield the chosen OR gene from the feedback. Only exception is the race model of cooperative clustering of a limited pool of TFs at the chosen OR. In this model the feedback acts globally by practically excluding the production of activating TF.

Model of self-sustained OR selection

The singular selection of OR genes during development of olfactory sensory neurons presents a first class example of the valleys of Waddington's epigenetic landscape. Inspired by the self-sustained bistability of DNA regions with certain nucleosome modifications and in thread with the descriptive models of McClintock, and Kambere and Lane, we propose a novel model of OR gene choice that captures the multistability of the system and encompass or reproduce all the characteristics of OR gene selection.

Like previous proposals our model includes feedback as it combines a cis-acting positive feedback mediated by nucleosomes and associated read-write enzymes, with a trans-acting negative feedback. The feedback mechanism that ensures the expression of a functional OR acts globally with no required shielding of the winning OR. Though based on experimental observations all previous models remain descriptive and lack a mean for reliably testing if the dynamics of the system would in fact be as envisioned. This novel model of OR gene selection is formulated as a stochastic simulation enabling a detailed study of the mechanisms generating multistability.
Multistability in differentiation of olfactory neurons
Chapter 10

Modelling the differentiation of olfactory neurons

The olfactory system integrates signals from receptors expressed in olfactory sensory neurons. Each sensory neuron expresses only one of many similar olfactory receptors. The choice of receptor is made stochastically early in the differentiation process and is maintained throughout the life of the neuron. With thousand of OR genes to choose from, the system strictly requires a local mechanism that couples expression of each gene to itself while still leaving similar genes non-activated. On top, activation of a gene must inflict a feedback that keeps other OR genes from being activated.

Our model of OR gene choice is inspired by the proposed involvement of histone modifications as the epigenetic trait. We revisit a theoretical approach on epigenetic cell memory by nucleosome modification, initially taken by Dodd et al. in 2007 [104, 116] and explored in the context of vernalization in plants by Angel et al. [3]. The model combines nucleosomes and associated read-write enzymes as mediators of a cis-acting positive feedback with a trans-acting negative feedback, thereby coupling the local epigenetic landscapes of the individual OR genes in a way that allow one and only one gene to be active at any time.

10.1 Bistable gene regulation by nucleosome modifications

The theoretical model on histone modification mediated epigenetics in eukaryotes was inspired by the bistable switching in fission yeast mutants, briefly described in section 2.2. In the two state version of this theoretical analysis a DNA region containing $L$ nucleosomes is considered [70, 21]. The histones of each nucleosome may carry transcription permissive or restrictive marks, categorizing the nucleosome as being in one of two states; silent, or active. The
Figure 10.1: Models of epigenetic memory through nucleosome modifications. The models employ a compressed description of the multi-step complex scenario where different levels of histone modifications dictate activity of the genes. 

A) Three state model. The model includes three nucleosome types M (methylated), U (unmodified), or A (acetylated). Interconversions occur through noisy transitions or recruitment of histone-modifying enzymes by nearby M or A nucleosomes shown as dotted lines related to HMT (histone methyltransferases), HAT (histone acetyltransferases), HDM (histone demethylases), or HDAC (histone deacetylases). Cooperativity is implicitly included in the recruitment process as transformation from silent to active state requires two steps of recruitment. Figure adapted from [104].

B) Two state model. In this compressed version of the three state model, each nucleosome can be either modified (M) or Anti-modified (A). Nucleosomes may convert between the two states by noisy events or recruitment of histone modifying enzymes by two local nucleosomes. Cooperativity is explicitly included as two recruiting nucleosomes are required for each recruited transformation. Figure adapted from [70].

DNA region may then be in an overall permissive or repressed state depending on the dominating histone mark of the region.

In the context of the model, transition between nucleosome states are made randomly or by active recruitments. In a random event a nucleosome spontaneously converts to the other state. During recruitment a histone-modifying enzyme is recruited by nucleosomes of a certain modification and is assumed to modify a nearby nucleosome to match the modification of the recruiting nucleosomes. Examples of such read-write mediators are found in the enzymes PRC2 and ASHH2. When these recruitment processes include either implicit or explicit cooperativity, the state of the DNA region can be bistable [104, 70, 21], with the majority of the nucleosomes being in either the silent or active state, see Fig. 10.1. Within reasonable feedback to noise ratios a chosen DNA state can be very stable over several cell generations, thus capturing a gene in one of two distinct epigenetic states. Introduction of a transcription factor that regulates the efficiency of the read-write enzymes provides a mean for controlled switching between the stable states. With such control the model presents a feasible description of bistable gene regulation on the level of single genes.
10.2 Multistable system of communicating nucleosome regions

The model of epigenetic cell memory through nucleosome modification which allows for sensitive gene regulation by positive feedback loops [104, 116], provides us with a predefined bistable subsystem to describe each OR gene. The real OR gene family is one of the largest mammalian gene families with 900-1400 genes. In modelling we consider system sizes of 10, 100 or 500 similar genes or subsystems each consisting of \( L = 50 \) nucleosomes. Such systems may be thought to describe the random choice between potential genes within overlapping zones [68].

The model couples the individual OR genes through the histone modifying reactions, the same step that controls the switching between states in the bistable nucleosome model. In the OR model the coupling is achieved by letting the probability of a recruited event, towards the active state for any gene, decrease with a hypothesized factor, effectively governed by the activity of all OR genes, see Fig. 10.2 A. The factor is thought to prevent the conversion of silent nucleosome marks to active ones, perhaps through binding and shielding of the silent marks, or by inhibiting the activating enzymes. This global negative feedback thus favours the silent state. The model, outlined in Fig. 10.2, is formulated as a stochastic simulation as in [104, 116], and analysed in terms of deterministic equations.

**Governing equations**

In the model the global negative feedback acts through a protein factor that binds to the silenced nucleosomes and prevent their conversion to active ones. The protein factor, \( P \), is effectively governed by the gene activity of all genes:

\[
P \propto a_1^h + a_2^h + \ldots + a_N^h = \sum_j a_j^h \tag{10.1}
\]

where \( a_j = A_j/L \), with \( A_j \) being the number of active nucleosomes on gene \( j \). The “Hill coefficient”, \( h \), quantify the threshold function for activity of the individual gene. If \( h = 1 \), the activity is simply proportional to having one particular nucleosome in the active state, for example the nucleosome at the promoter. Larger \( h \) values represent an increased threshold for production from the individual genes. Within each gene the silenced nucleosome fraction, \( s = 1 - a \), fulfill

\[
s = s_p + s_{\text{free}} \tag{10.2}
\]

\[
s_p = P \cdot s_{\text{free}} = r \sum_j a_j^h \cdot s_{\text{free}}
\]

where we assume fast rates for conversion between a silenced nucleosome with \( P \) bound, \( s_p \), and a silenced nucleosome without \( P \) bound, \( s_{\text{free}} \). This gives the nucleosomes accessible for conversions

\[
s_{\text{free}} = \frac{1}{1 + r \sum_j a_j^h} \cdot s = R \cdot s \tag{10.3}
\]
Figure 10.2: Model for maintenance of selected gene expression in an olfactory neuronal cell. A) Local positive feedback and global negative feedback. \( N \) genes each covered by nucleosomes (Red for silently marked nucleosomes, Blue for actively marked). The nucleosomes on each gene form positive feedback systems that maintain the selected expression state (Red and blue arrows). The active gene expresses receptor proteins which direct enzymes that favour nucleosomes to stay in a silenced state by repressing the positive feedback towards the active nucleosome state (black arrows and blunted lines, respectively). B) Switching property of a single gene. Each gene is covered by \( L \) nucleosomes that each can be in an active state that is open for transcription (Blue) or in a silenced state (Red). The nucleosomes together with associated “read-write” enzymes form positive feedback systems that allow a gene to maintain a previously selected state (Blue and red arrows). Enzymatic activity effectuated by the activity of all active genes captures genes in the silent state (dashed black arrow and blunted line). C) Nucleosome modifications recruit read-write enzymes within each gene. For simplicity we consider only two nucleosome states, which each can modify the states of other nucleosomes cooperatively as in [116]. The feedback between genes described in A) and B) acts through protection of silenced nucleosomes against conversions to the active state with strength proportional to \( \sum_i a_i^h \) where \( a_i \) is the active nucleosome fraction of gene \( i = 1, \ldots, N \).
Multistable system of communicating nucleosome regions

which defines the reduction, $R$, of conversion activity of nucleosomes from the silent to the active state in general.

In the limit of an infinite number of nucleosomes per gene the governing equations read:

$$\frac{da_i}{dt} = (R \cdot s_i) \cdot a_i^2 - \mu \cdot a_i \cdot s_i + \beta \cdot (R \cdot s_i) - \beta \cdot a_i$$

$$\downarrow$$

$$\frac{da_i}{dt} = \frac{1}{1 + r \cdot \sum_j a_j^h} \cdot (1 - a_i) \cdot a_i^2 - \mu \cdot a_i \cdot (1 - a_i)^2$$

$$+ \frac{1}{1 + r \cdot \sum_j a_j^h} \cdot \beta \cdot (1 - a_i) - \beta \cdot a_i \quad (10.4)$$

where $a_i$ is fraction of active nucleosomes in gene $i$, $h$ is the hill coefficient for the production of each gene product in terms of active nucleosomes, and $\mu < 1$ is the bias that favour the active state of all genes. $r$ effectively counts the rate at which a gene product from any gene protects any silenced nucleosome from being activated. Binding of the protein factor also shields the nucleosomes conversion by noise, thus restricting noise conversions to active and uncovered silenced nucleosomes.

Fixed point analysis of governing equations

With Eq. 10.4 we can analyse the system with respect to the two genes with largest fractions of nucleosomes with active marks. These genes are interesting as a fixed point analysis of their states, shows how the system changes from having one to more actively marked genes. Additionally, the analysis provides an understanding of the activity level of the repressed genes.

Consider two single genes, with active nucleosome fractions, $a$ and $b$, that are freely varying in a situation where all other $N - 2$ genes are slaved to a single variable active nucleosome fraction, $c$.

$$\frac{da}{dt} = \frac{1}{1 + r(a^h + b^h + (N - 2)c^h)} \cdot (1 - a) \cdot a^2 - \mu \cdot a \cdot (1 - a)^2$$

$$+ \frac{1}{1 + r(a^h + b^h + (N - 2)c^h)} \cdot \beta \cdot (1 - a) - \beta \cdot a \quad (10.5)$$

$$\frac{db}{dt} = \frac{1}{1 + r(a^h + b^h + (N - 2)c^h)} \cdot (1 - b) \cdot b^2 - \mu \cdot b \cdot (1 - b)^2$$

$$+ \frac{1}{1 + r(a^h + b^h + (N - 2)c^h)} \cdot \beta \cdot (1 - b) - \beta \cdot b \quad (10.6)$$

$$\frac{dc}{dt} = \frac{1}{1 + r(a^h + b^h + (N - 2)c^h)} \cdot (1 - c) \cdot c^2 - \mu \cdot c \cdot (1 - c)^2$$

$$+ \frac{1}{1 + r(a^h + b^h + (N - 2)c^h)} \cdot \beta \cdot (1 - c) - \beta \cdot c \quad (10.7)$$
Modelling the differentiation of olfactory neurons

With Eq. 10.5 and 10.6 the fixed points of the system, defining a steady DNA state, can be graphically addressed in a plot of the nullclines of $\frac{da}{dt}$ and $\frac{db}{dt}$ under the assumption that $c$ takes the lowest value where $\frac{dc}{dt} = 0$. This assumption is reasonable because $\frac{dc}{dt} > 0$ for $c = 0$ and for increasing $c$ the first nullcline should therefore be stable against variations in $c$.

Stochastic implementation of model

The stochastic version of the model mimics a system of $N$ genes each covered by $L$ nucleosomes. At each time-unit $R = \frac{1}{(1 + r \cdot \sum_i a_i^b)} = \frac{1}{(1 + P)}$ where $a_i = A_i/L$ with $A_i$ being the number of active nucleosomes in gene $i$. Each step of the simulation then include:

1. Selecting 2 random nucleosomes in one random gene. If the state of both nucleosomes are active, with probability $R$ another nucleosome in this gene is selected and its state is set to active. If the two nucleosomes are in the silenced state with probability $\mu$ another nucleosome is selected and its state is set to silenced.

2. With probability $\beta$ selecting a nucleosome among all genes in the system. If the state of the chosen nucleosome is active, then it is changed to silent. If the state is silent, with probability $R$ it is changed to active.

One time-unit corresponds to one update per nucleosome in the entire system of $L \times N$ nucleosomes. Thus, in each time unit of a simulation every nucleosome is on average attempted converted by a recruitment event, with actual conversions happening much less frequently.

The olfactory epithelium is constantly renewed through neurogenesis [35, 100, 137]. In adult rats regeneration of olfactory receptor neurons (ORNs) from immature progenitors is achieved within 2 weeks [105]. Similar timescales of neurogenesis are observed in embryonic development [124, 96]. In contrast reported lifetimes for neurons in rodents are very variable, ranging from one month to almost a year [39, 61, 105]. With a turnover rate of histone modifications in the order of 10 min reported for the de-acetylation processes [22, 120, 121, 17], we want our model to choose one subsystem within 1 week $\sim 1000$ time units in the simulations and contain exclusive expression of that subsystem for at least 1 month $\sim 5000$ time units. Other modification processes may well be slower [125, 122, 142] but such change of timescale does not change our overall results. It would only move the parameter range for which we have acceptable differentiation properties.

For each simulation we hence run 5000 time-units, starting each gene from an all silenced state, and requiring that only one gene turns on within the time window, and that only one gene is active in the time window. In tread with previous nucleosome models we score a subsystem as activated when the fraction of actively marked nucleosomes exceeds $2/3$. Activation is considered lost when the fraction falls below $1/3$ [104], but in fact an insignificant fraction of the genes that switch to become the single dominating gene will switch back to the silent state within the plotted time frame.

76
Chapter 11

Results from modelling the differentiation of olfactory neurons

This chapter highlights the results in the manuscript included in appendix C. During development from immature progenitor to mature olfactory neuron each cell must choose to express only one among hundreds of OR genes. The large number of OR genes, their chromosomal spread and randomness of expression, and the large fraction of pseudogenes all adds to the complexity of how each neuron becomes devoted to the expression of a single OR gene. A mechanism that in addition should be robust to quite drastic changes in gene numbers. Our model of OR gene expression implements a local feedback mechanism centred in the cis-acting interplay of histone modifications and read-write enzymes in maintaining the activity state of each OR gene. A global acting feedback couples the systems to ensure that only one OR gene is active at any time. The model captures the basic property of the differentiation of olfactory neurons, namely the exclusive expression of just one gene out a large highly homologous gene population.

11.1 Stochastic simulations

Initially we test the model in stochastic simulations, initiating the system with all nucleosomes in the silent state consistent with the observations by Magklara et al. [64]. Within each simulation we identify the state of the individual gene, $i$, as active, when the active fraction of nucleosomes, $a_i$, exceeds $2/3$. Fig. 11.1 show a simulation of $N = 10$ genes covered by $L = 50$ nucleosomes each. Gene 5 achieves activation as the number of active nucleosomes exceeds 33 at time $\sim 60$. Stochastic fluctuation and internal local bias toward the active state, governed by the parameter $\mu$, move the subsystem into a dominant active state. Activation increases the globally acting negative feedback thus reinforcing the dominance of the active gene by decreasing the probability for local activation.
Results from modelling the differentiation of olfactory neurons

Figure 11.1: Simulation of $N = 10$ genes, each covered by $L = 50$ nucleosomes. The simulation shows that one gene is turned on quite early, while all other genes remain silenced throughout the simulation. In fact with these parameters all these genes stay silenced up to at least $t = 5000$ time units. Sketches in the right panel illustrate the OR gene state at the final time of the simulation. Red for silently marked nucleosomes, blue for actively marked. Crossed promoters indicate silent genes. Other parameters of the simulation are $\mu = 0.50$, overall repression factor $r = 1$, hill coefficient of repression $h = 2$ and noise conversion $\beta = 0.03$.

of other genes. Experiments, where one expresses an OR transgene from a promoter that is active early in the development of the OSN, show that this gene will dominate the future differentiation [78]. This is in accordance with our negative feedback hypothesis which indeed predicts that any early dominance will prevail.

Success of a simulation is assessed on three criteria. First the system needs to selectively activate a single OR gene within a given time window. Secondly, the chosen gene should remain active for a considerable time. Thirdly, no other OR genes may be activated while the initial OR gene is active. Fig. 11.2 A, D & G compactly show time courses like those of Fig. 11.1 for a system of 100 genes at three different values of the local activation bias, $\mu$. It is clear that increasing $\mu$ takes the system from defying the first criteria by switching on more than one gene, to fulfilling all criteria with a single active gene, and to failure due to lack of turn on of any genes.

11.2 Activation barriers capture genes in silent state

While simulations present a proof of concept for our model in reproducing the exclusive gene expression, the theoretical formulation of the model additionally lets us examine in detail how the dominant gene expression is achieved. Arguing that all but the two most active genes are found at the same low level of activation, three stable fix points for the states of the two genes are graphically identified. The coloured regions in Fig. 11.2 B, E & H, show where the net variation in active nucleosome fraction is positive for the two most active genes. Intersections between the regional limits are fix points of the governing equations where the net variation in the active nucleosome fraction is zero. Fix points closest to the corners of the $a, b \in [0; 1]^2$ are stable as local variation in activity in gene “$a$” and “$b$” deterministically return the system to the states of
11.2 Activation barriers capture genes in silent state

Figure 11.2: Simulation of $N = 100$ genes, each covered by $L = 50$ nucleosomes. Left panels A,D,G) show time courses of the first activated gene (blue), second activated (red) and a few examples of other genes (yellow and green). Inserts show the promoter status of the correspondingly colored genes at the final simulation time. Crossed arrows indicate silent promoters. Middle panels B,E,H) follow the trajectory of the two most expressed genes in a 2-d plane that illustrates deterministic drift of two individual genes, provided all other genes are assumed to act synchronously. In the blue region $da/dt > 0$, whereas the red region shows where $db/dt > 0$. C,F,I) show the probability (lighter colour for higher) of different numbers of active nucleosomes for the two most active genes in the system, obtained by stochastic simulation over $10^8$ time-units. The negative logarithm of this probability may be interpreted as an epigenetic landscape [134, 34], of a system that to varying degree prefer to be in the states of the corners. Parameters are $r = 1, h = 2$ and $\beta = 0.03$ constant for all figures and; $\mu = 0.30$ for A-C, $\mu = 0.50$ for D-F, and $\mu = 0.70$ for G-I.

The 2D planes in Fig. 11.2B,E & H, include trajectories of the first versus the second most active gene from the simulations to the left and show how the success relies on two barriers. At low $\mu$, examined in Fig. 11.2B, the “first” barrier stalls the system at the state where all genes are silenced until a random fluctuation causes activation of one of the genes. Subsequently, the system remains unstable against the passage of a second barrier to the state in the rightmost upper corner. At larger $\mu$ only one gene switches, as the second and larger barrier along the vertical axis in Fig. 11.2E prevents a second gene from turning on once an initial switch have been made. At even larger $\mu$ values the
Figure 11.3: Epigenetic landscapes. Examination of the probability (lighter colour for higher) for the two most active genes in the system, obtained by stochastic simulation over $10^8$ time-units. The negative logarithm of this probability may be interpreted as an epigenetic landscape [134, 34], with states that to varying degree prefer to be in the corners. The simulation uses standard parameters $N = 100$, $L = 50$, $r = 1$, $h = 2$, $\beta = 0.03$ and time delay $\tau = 0$ when nothing else is specified. Notice that both smaller $L$ and a time delay lower the barriers in the system, and favour transitions to the rightmost corner where two genes are active simultaneously.
11.3 Parameter sensitivity

first barrier is so high that no gene may pass it, essentially stalling the system with all OR genes in the silent state (see Fig. 11.2H). Thus, for a range of $\mu$ values the system is successfully stalled in the lower right corner of the $[0; 1]^2$ phase space with just one activated OR gene.

Interestingly, when the system is stalled in the lower right corner of Fig. 11.2 B, or E, the repressed genes have a small fraction of their nucleosomes in the active state, indicating a basal activity level of the silent OR genes. This “nucleosomal noise” could cause a residual production of receptors from each repressed gene and consequently a residual production of feedback protein factor. From Eq. 10.7 we can deduce the fraction of active nucleosomes in repressed genes, as this is reflected in the lowest $c$ value where $dc/dt = 0$. Using Eq. 10.7 the “nucleosomal noise” is approximately $c = \beta/(\mu(1 + r)) \sim \beta$. By Eq. 10.1, the minority fraction of active nucleosomes is associated to a minority production of receptor or feedback factor from each repressed gene of about $\beta^h$.

For comparison, the production from the chosen active gene is $a^h \sim 1$ where $a$ is determined from largest $a$ that fulfils Eq. 10.5. This value of $a$ corresponds to the “$a \sim 1$” solution in the lowest right corners of Fig. 11.2B,E. In the face of a possible “background” expression of receptors from the silent OR genes, we require that the cell response is dominated by the chosen gene, implying that the activity of the chosen gene should be in excess of all the remaining N-1 genes together: $a^h \sim 1 > (N - 1) \cdot \beta^h$. This demand is fulfilled for all examined parameters. On a biological note, it might not be catastrophic for the cell to have a single receptor miss-expressed as we imagine that more than one receptor in the membrane is needed to initiate a real odour signal.

The differentiation of OR genes is also examined in terms of epigenetic landscapes generated from stochastic simulations. Using $L = 50$, the rightmost panels in Fig. 11.2C,F,I examine the probability for the two most active genes in the system to be covered by actively marked nucleosomes. Light colours are associated with preferred states, whereas blue and black represent disfavoured combinations of activity. In Fig. 11.2C one can see that both the state where one gene is active, and the state where two genes are active, represents “probability peaks”, corresponding to valleys of favoured states in Waddington’s epigenetic landscape [134]. In contrast, for the parameters of Fig. 11.2F,I, the simultaneous activity of two genes is unlikely.

The correspondence to epigenetic landscapes are emphasized in 3D visualizations in Fig. 11.3. The figure shows the robustness of the landscapes of Fig. 11.2C,F,I with regard to different parameters of the model. Passages between the tops in the landscape are modulated by both decreasing gene sizes (left column) and an eventual time delay in the negative feedback, $\tau = 100$ (right column).

11.3 Parameter sensitivity

In the Olfactory neuron the loss of OR memory after axon extension to the predefined glomeruli of the olfactory bulb would be catastrophic as it would mess up the typographical map of odours in the brain. Investigation of multiple
Results from modelling the differentiation of olfactory neurons

Figure 11.4: Parameter sensitivity for \( N = 100 \) genes, \( L = 50 \) nucleosomes system with repression factor \( r = 1 \), hill \( h = 2 \) and noise \( \beta = 0.03 \) fixed. Data is averaged over 200 simulations. **A** Orange area shows the probability that one and only one gene becomes active within 5000 time-units as function of asymmetry \( \mu \). Concurrent red area marks success with the additional constraint that one gene becomes active within the first 1000 time-units. Gray area marks the cut-off at 50% successful simulations. **B** Orange area marks the largest number of active nucleosomes within one gene during a 5000 time-unit simulation. When no genes exceeds the activation limit during the first 5000 time-units, no olfactory receptors have turned on. Cyan and dark cyan show maximal number of active nucleosomes on the second and third most active genes, respectively. Where the number of active nucleosomes on the second most active gene exceeds the activation limit, the two genes have shown simultaneous activity.

Stochastic simulations show that even for the best parameter combinations a fraction of the simulations will fail to meet the success criteria as illustrated in a plot of success probability vs different \( \mu \) values in Fig. 11.4. Simulations fail due to activation of more than one OR gene or the simple lack of any activation. However, it is reported that about 50% of neurons born in the epithelium of adult rats are lost within 5 days and 2 weeks [105, 124]. There are no explanation to this loss, but it has been argued that it could be a control mechanism to get rid of ORNs expressing more than one receptor or no receptor at all [68, 72]. Thus for the model to successfully represent differentiation of olfactory neurons only 50% of the simulations need to be successful, indicated by the gray shaded area in Fig. 11.4A. Initially one of the success criteria dictated that the first OR should activate within \( t < 1000 \) time units. This restriction is somewhat conservative and relaxing it to activation within the first 5000 time-units widens the success range of the model, shown as concurrent orange areas in Fig. 11.4A.
Additionally, the one-neuron one-receptor rule may not be quite as manifested as widely accepted within the field [72, 47, 127]. Thus Fig. 11.4B include the maximal activity of the second and third most active genes. These show that allowing a second or even third gene to activate further widens the $\mu$ range of successes. It also captures the effect of increasing $\mu$ observed in Fig. 11.2.

In Fig. 11.5 the success probability of a simulation is examined as function of $r$ and $\mu$ at different hill coefficients, noise levels and nucleosome numbers. It is evident that the model is robust to changes in all parameters. A sharper threshold for gene activity of the individual genes, parametrized by higher hill coefficient $h = 4$, allow for more robust cell differentiation, identified within the model by a larger range of $\mu$ where model succeeds. In general the model implements $h = 2$, but Fig. 11.5 emphasize that our results are robust to sharper threshold functions. Smaller noise $\beta = 0.01$, instead of $\beta = 0.03$, likewise facilitates robustness.

The mechanism of gene selection in olfactory neurons seem to be conserved between mammals with large fluctuation in the number and chromosomal positions of OR genes to choose from. The diversity and wide chromosomal position of OR gene clusters are most likely a result of subsequent tandem gene duplications, gene conversions and recombination events and even conversions of entire coding regions [33]. Accordingly the mechanism of gene selection should be robust to quite drastic changes in gene numbers [47]. The model successfully addresses this issue by selecting one OR gene, keeping the remaining genes silent, for essentially the same set of parameters when the system size is increased from 100 to 500 genes (compare Fig. 11.5F and black curve of Fig. 11.6).

11.4 Pseudogenes, delayed feedback and early transient switching.

In general developing neurons may show transient activation of different OR genes [113]. Fixing the parameters at functional values like those of Fig. 11.2D,
Results from modelling the differentiation of olfactory neurons

Figure 11.6: Robustness to increased number of subsystems. $\mu$ dependence of success probability for a system of $N = 500$ genes, with $L = 50$ nucleosomes and $r = 1$. Remaining parameters are as indicated in corresponding colours. Comparing the black curve to Fig. 11.5F, it is clear that the model succeeds for the same set of parameters as system size is increased from $N = 100$ to $N = 500$.

Initial transient activation of more OR genes is possible if we introduce a time delay, $\tau$, between OR activation and feedback production, see Fig. 11.7B. The time delay, $\tau$, is included by allowing for degradation of the protein $P$ that facilitates the negative feedback:

$$\frac{dP}{dt} = \frac{1}{\tau}(r \cdot \sum_j a^h_j - P). \quad (11.1)$$

$P$ is updated in each step of a simulation and replace $r \cdot \sum_j a^h_j$ in the expression for $R$. The standard model is thus appropriate when the factors facilitating the negative feedback have a life time that is shorter than it takes one gene to switch from a silent to active state. In case the degradation is slower, the early activity of the OR system is often altered, with several genes turning on at very early times, for then subsequently to loose their activity when the full effect of the, then overproduced, negative feedback comes into play. As a result the qualitative behaviour from Fig. 11.4 is reproduced with all time-delays that is substantially smaller than the maturation time of 1000 time-units. Our model treats allelic genes individually but still encompasses mono-allelic expression as only one gene is activated. Initial transient activation of both alleles is thus possible, in thread with experimental observations of switching between differently marked alleles of the same gene in early neurons [113].

Within the model settings that restrict the stable activation to a single OR, we considered the concept of pseudogenes. In this context pseudogenes are subsystems affected by feedback as previously described but lacking the ability to produce the feedback. It will occasionally happen that first a pseudogene is activated, however, as the active pseudogene does not contribute to the negative feedback, $R$, another subsystem will eventually be activated and retain the dominant stable position, see Fig. 11.7A. The frequency of initial activation of a pseudogene is higher in systems without time delay due to the lack of feedback production from pseudogenes. This elegantly illustrates the effect of timedelay.
Figure 11.7: Pseudogenes and time-delay: A) Activity of genes as function of time for our standard model with $N = 100$ genes, including $N_{\text{pseudo}} = 40$ pseudogenes each covered by $L = 50$ nucleosomes, noise $\beta = 0.03$, asymmetry $\mu = 0.50$, overall repression factor $r = 1$, and hill $h = 2$. The dotted trajectories are pseudogenes, and illustrate that several pseudogenes may become activated early. The small promoter pictures show temporal activity status of the correspondingly colored genes, with pseudogenes shown in gray shaded box. B) As above, but with a time delay, $\tau$, between OR activation and the feedback $P$. The time delay opens for transient activation of several OR genes. Once $P$ accumulates the active ORs are repressed and only one OR gene remains active. As genes only sense their own activity after some time, the turn-on frequency for real genes are the same as for the pseudogenes.

When included, the real genes do not counteract their own activity while turning on, due to the delay of feedback production, and therefore switch on as often as the pseudogenes, see Fig. 11.7B.

11.5 Enhancers and modulation by transcription factors

Enhancer elements with features similar to the locus control region of the visual pigment genes have been identified for some subfamilies of OR genes [108, 109, 28]. However, such elements are yet to be identified for all OR genes of the super gene family [28]. In the context of our standard model we envision enhancer elements as a mean of disposing an OR gene for epigenetic activation. Decreasing the probability, $\mu$, of a recruited event towards the silent state for one subsystem by as little as 15% greatly increase the frequency of that subsystem being activated, Fig. 11.8A. Such an increase in frequency for H-element associated genes has been reported [28, 96, 49].

Conserved regions identified in alignments of OR genes, include binding sites of known TFs. [32, 131, 41, 97, 69]. So far no single TF, or TF binding site, have been associated with the zonal exclusivity, even though TFs as Emx2 do alter the expression frequency of a large part of the OR genes [69]. In our model TFs may be incorporated, like enhancer elements, as histone modification control mechanisms [116], enhancing or reducing the probability of activation of their
Results from modelling the differentiation of olfactory neurons

Figure 11.8: Predictions of sensitivity to enhancers and to double gene dosage:

**A)** Effect of small alterations in OR specific $\mu$ on the probability of switching on for the associated OR gene. Individual decrease (orange) or zonal increase (gray) in the activation asymmetry $\mu$ might represent the effect of an enhancer element or the removal of zonal specific TFs. Each data set from 1000 simulations, with $N = 10$ genes, each covered by $L = 50$ nucleosomes, $\mu = 0.50$, $\beta = 0.03$, $r = 1$, and $h = 2$.

**B)** Comparison between wild type case, and an engineered situation where one gene contributed with double gene dosage to the feedback $P$. The likelihood of turning on the “doubled” gene shown with orange bars is smaller than for a normal gene to an extent that depend on $\mu$. In case the “doubled” gene is constructed as a tandem repeat, possibly with two reporter proteins inserted, the two identical promoters should become active together, but with smaller probability than the wild type system. A reduction in probability that will pinpoint the effective value of $\mu$. Each data set from 1000 simulations, with $N = 100$ genes, each covered by $L = 50$ nucleosomes, noise $\beta = 0.03$, $r = 1$, and $h = 2$.

associated genes, see Fig. 11.8A. The model shows that simple TF knockout experiments will not necessarily allow for a clear identification of such TFs as full exclusion of associated OR genes. Instead the frequency by which the genes are chosen will decrease while unrelated OR genes increase as reported for the Emx2 knockout mice [69].

### 11.6 Identification of regulatory elements

Reports on the size of the regulatory elements needed for a recapitulation of the punctuate gene expression, of OR genes not associated with enhancer elements, varies from $\sim 1$kb to more than 10kb [32, 131, 97, 78], with varying intrinsic levels of spatial restriction. The short and highly variable OR coding sequences may thus be sufficient to mark a transgene for the same feedback regulation as endogenous genes. This is seen in experiments where Olfactory Marker Protein (OMP) promoter driven expression of an OR gene exclude expression of other OR genes, at least in the case of OR gene MOR28. As OMP is a marker of mature olfactory neurons, its promoter should not be regulated in the same
manner as the OR genes [78]. In contrast a mini gene containing only a small segment of the OlfR16 loci flanking the transcription start site of the gene show a mosaic expression pattern on the OE that closely resembles that of endogenous genes [84]. In our model we initially assume a regulatory size of $\sim 10$kb or at least large enough to encompass 20-50 nucleosomes.

In more general terms the model presents a way of activating a single patch of the genome, including the feedback producing element, which is responsible for keeping the remaining patches silent. In fact, the average OR genes of a locus are separated by 29kb [143], allowing for separation between activated and silent regions, and implying that the activity of their respective nucleosomes is separated by substantial barriers.

In transgene experiments one may modify distances between genes to examine the spreading mechanisms associated to the local positive feedback. In particular one may insert two genes just after each other, with identical promoters, and investigate whether such an architecture facilitate non-synchronous turn-on of the genes, or bound synchronous activation as our model suggests. With standard promoter strengths of the two coupled genes, the globally acting negative feedback will counteract any activity from such a coupled pair of OR genes as shown in Fig. 11.8B. Thus the promoter strengths of the genes should be reduced to up the chance of observing the synchronous activation. Co-expression of two functional OR genes is also predicted by the model in an experiment where the negative feedback effect is halved. In both scenarios co-expression requires that no control mechanism, apart from those of the gene selection, excludes cells that express more than one OR.
Results from modelling the differentiation of olfactory neurons
Chapter 12

Concluding remarks and perspectives

Our model for exclusive expression of a single OR gene, describes each OR gene as a region of nucleosomes, with the activity of the gene determined by the transcriptional marks of the core histones. The dynamics are based on a cis-acting positive feedback between nucleosomes effectuated by read-write enzymes, ensuring bistability within each OR gene, and a trans-acting negative feedback factor coupling the activity among genes. Alternatively, the model could describe a system where the states of the individual genes is governed by DNA methylation marks. In fact the OR genes generally carries repressive DNA methylation [60].

The model pinpoints the need for localizing the dependence of olfactory co-regulation on the size and status of chromosomal modifications of the individual olfactory gene locus. It also identify the need for exposing the common factor that facilitates the interplay between the different olfactory genes. Inspired by the model future search for this later globally acting feedback factor should be directed within pathways associated with histone modifying enzymes.

Important experiments to test our model could address the former role of the local nucleosome mediated feedback. Such experiments may be initiated without explicit knowledge of the players. Placing two olfactory receptor genes close on the DNA should result in co-expression, provided a reduction in the strength of the associated promoters. Our model does not address the possibility of secondary regulatory mechanism within the neuron that could oppose such co-expression of two functional ORs. A way to experimentally circumvent the need for reduced promoters and any intracellular control against co-expression, would be replacing one of the proximate genes with a pseudogene. Lack of consistent co-expression would then falsify our model. In fact an experiment that very closely resemble this setup, was reported by Pyrski and colleagues in 2001 [90]. They showed the matching expression patterns of a lacZ reporter fused to the olfactory marker protein promoter and an OR gene identified upstream of the insertion site of the reporter, and attributed the match to locus dependent gene expression.
A final speculation springs from the fact that olfactory neuronal cells and the sense of smell is associated to the oldest part of the cortex and the observations of OR like seven trans-membrane receptor families that have been reported in other parts of the neural system [26, 99]. Thus the genetic regulatory design of the OR genes may be found more widespread and differentiation of olfactory neurons could give insight into the way we store our memories. In particular, Fig. 11.4 opens for the possibility to activate two receptors in a cell, thus largely increasing the number of epigenetic states. As ORs are believed to direct the axon-to-glomerulus association at the olfactory bulb, receptors may also modulate physical connections between other neuronal cells. Thereby receptors could couple the intrinsic epigenetic state of cells to the neural network-architecture opening for a new perspective on memory. A perspective where the epigenetic state of individual neuronal cells may be coupled to the real memory storage in our cortex.
Bibliography


Appendix


Appendix A


Modeling of the Genetic Switch of Bacteriophage TP901-1: A Heteromer of CI and MOR Ensures Robust Bistability

Hiizu Nakanishi¹*, Margit Pedersen², Anne K. Alsing² and Kim Sneppen²

¹Department of Physics, Kyushu University 33, Fukuoka 812-8581, Japan
²Center for Models of Life, Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2100 Copenhagen, Denmark

Received 28 April 2009; received in revised form 23 August 2009; accepted 27 August 2009
Available online 9 September 2009

The lytic–lysogenic switch of the temperate lactococcal phage TP901-1 is fundamentally different from that of phage lambda. In phage TP901-1, the lytic promoter PR is repressed by CI, whereas repression of the lysogenic promoter PL requires the presence of both of the antagonistic regulator proteins, MOR and CI. We model the central part of the switch and compare the two cases for PR repression: the one where the two regulators interact only on the DNA and the other where the two regulators form a heteromer complex in the cytoplasm prior to DNA binding. The models are analyzed for bistability, and the predicted promoter repression folds are compared to experimental data. We conclude that the experimental data are best reproduced the latter case, where a heteromer complex forms in solution. We further find that CI sequestration by the formation of MOR:CI complexes in cytoplasm makes the genetic switch robust.

© 2009 Elsevier Ltd. All rights reserved.

Keywords: temperate bacteriophage TP901-1; genetic switch; mixed feedback loop; sequestration; protein interaction

Introduction

Phenotypic variability under homogeneous conditions can be readily obtained by interlinking multiple gene regulatory pathways. Several well-characterized examples of phenotypic variations are known to be important for the different developmental processes of bacteria, such as the presence of persister cells in Staphylococcus aureus and Escherichia coli, development of natural competence and sporulation in Bacillus subtilis, and the choice between lytic or lysogenic growth of temperate bacteriophages.¹,² Two distinguishable phenotypes may originate from a bistable system, that is, a system that can toggle between two alternative stable steady states.³ Infec-

tion of bacteria by temperate bacteriophages provides a classical example of the possibility to choose between two alternative modes of development.

The bacteriophage lambda infecting E. coli has been subjected to decades of intensive study, making the lytic–lysogenic switch one of the best understood gene regulatory systems.⁴–⁶ The bistability of the lambda switch is obtained from a double-negative feedback mechanism, where two repressor proteins directly repress transcription of the other repressor gene. This system has a stable state with one promoter on and the other off, and vice versa for the other stable state. Once either state has been established, it would persist indefinitely or until some trigger stimulus forces the system to switch to the other state.

The genetic switch of the temperate lactococcal bacteriophage TP901-1 infecting Lactococcus lactis subsp. cremoris provides a regulatory system diverse from the lambda genetic switch. A previous study has demonstrated that a DNA fragment obtained from the temperate lactococcal phage TP901-1 shows bistability when introduced into L. lactis. The cloned DNA fragment contains the two divergently orient-
ed promoters, PR and PL, and the two promoter proximal genes cl and mor (Fig. 1a). A knockout mutation in the mor gene showed that CI ensures tight repression of the PR promoter and partial repression of the PL promoter, whereas a knockout mutation in the cl gene results in open states of both PR and PL showing that MOR by itself does not exhibit repression of either promoter.⁷,⁸ Two types of

*Corresponding author. E-mail address: nakanisi@phys.kyushu-u.ac.jp.
repression have been shown: (i) MOR-independent repression, which is responsible for repression of PL. The PL promoter is repressed by cooperative binding of CI to the three operator sites OR, OL, and OD, by the formation of a CI–DNA loop structure. (ii) MOR-dependent repression, which is responsible for repression of PR and only occurs in the presence of both MOR and CI. This repression is suggested to occur through MOR and CI binding at a putative OM operator site. Hence, the bistability of the genetic switch from phage TP901-1 may be described as a mixed feedback loop, where both of the antagonistic repressor proteins are involved in PR repression. It is still not clear how MOR and CI collectively repress transcription from PR, and so far, there is no direct experimental evidence for interaction between CI and MOR.

In order to understand the mechanism of switching in phage TP901-1, we here study a modified version of the cloned wild-type TP901-1 switch (Fig. 1b). This construct contains only one of the three CI operator sites, OL, which gives tight repression of PL and, thus, still sustains the bistable behavior of the construct. In the immune state, the PL promoter is repressed approximately 1000-fold, but high expression from PR is allowed due to the absence of OR, which autoregulates transcription from PR. In the anti-immune state, PR is repressed approximately 100-fold with transcription from PL being allowed.

Generally speaking, it is not easy to predict the behavior of a bistable switch without quantitative analysis because a bistable switch is a dynamical and highly nonlinear system. In the present case of TP901-1, even the modified version of the switch (Fig. 1b) could involve a number of mechanisms, such as cooperativity binding via homo-/hetero-dimerization, sequestration via heterodimerization, and intertwined loops of negative and positive feedback via protein interactions. In such a situation, the only way to obtain any reliable results is to perform quantitative analysis on specific models. By confronting numerical results with experimental data, we can restrict possible mechanisms with plausible parameters.

In this article, we construct mathematical models for this modified bistable system based upon statistical mechanics, examine their behavior in the steady states numerically, and compare the obtained repression folds with experimental observations. We assume that PL is repressed by the CI dimer binding to the operator OL since the operator has two inverted repeated sequences. MOR is likely to be a DNA binding protein, but the fact that MOR alone does not repress transcription from either PL or PR suggests that the DNA-binding affinity of MOR is negligible. CI alone does not repress PR in this system due to absence of OR operator. Based upon these observations, we test two different scenarios for PR repression (Fig. 2): (i) repression through direct binding of CI and MOR to O_M (Model A) and (ii) repression through binding of MOR:CI:DNA complex formation at the putative OM operator.
The MOR:CI:DNA complex formation in Model A may be regarded as an extreme case of Model B where the MOR:CI complex formation in cytoplasm is so weak that the complex is stabilized only when it binds to DNA; thus, there is no substantial presence of MOR:CI complex in the cytoplasm. However, distinguishing Model A from Model B helps us to recognize two distinct aspects in the PR repression by the MOR:CI:DNA complex, that is, cooperativity and sequestration. The latter has been studied in silico as a possible mechanism for a genetic switch and demonstrated to provide strong nonlinearity. In fact, our Model B is reminiscent of one of the bistable switches obtained by the simulated evolution (Fig. 3a in Ref. 12). We will demonstrate that CI sequestration by the MOR:CI formation in cytoplasm can make a robust bistable system and is actually a plausible switching mechanism for TP901-1.

**Theory**

The regulatory circuit in the present system consists of the two promoters $P_L$ and $P_R$, which produce MOR and CI, respectively (Fig. 1b). The promoter $P_L$ is repressed by CI binding at $O_L$; thus, the $P_L$ activity is given by a function of the CI concentration as

$$pL([CI]) = pL_0 \cdot f_{OL}([CI])$$ (1)
where \( p_{L0} \) is the bare activity of the promoter \( P_L \). The function \( f_{OL}(\text{[CI]}) \) represents the repression factor. In the absence of CI, there is no repression: \( f_{OL}(0) = 1 \). The \( P_R \) activity, on the other hand, depends on the concentrations of both CI and MOR. Accordingly, the \( P_R \) activity can be written as

\[
p_R([\text{MOR}], [\text{CI}]) = p_{R0} \cdot f_{OM}([\text{MOR}], [\text{CI}])
\]

where \( p_{R0} \) being the bare activity. The function \( f_{OM}([\text{MOR}], [\text{CI}]) \) is the repression factor due to the binding of MOR and CI at \( O_M \) and satisfies \( f_{OM}(0, [\text{CI}]) = f_{OM}(\text{[MOR]}, 0) = 1 \).

The promoter \( P_L \) produces MOR, and \( P_R \) produces CI; thus, in the modeled feedback system, the total concentration for each protein, \([\text{MOR}]_{\text{total}}\) and \([\text{CI}]_{\text{total}}\), is governed by the dynamics equations,

\[
\frac{d}{dt}[\text{MOR}]_{\text{total}} = \frac{1}{\tau_M} (p_L([\text{CI}]) - [\text{MOR}]_{\text{total}})
\]

and

\[
\frac{d}{dt}[\text{CI}]_{\text{total}} = \frac{1}{\tau_C} (p_R([\text{MOR}], [\text{CI}]) - [\text{CI}]_{\text{total}})
\]

where \( \tau_M \) and \( \tau_C \) are the degradation times for MOR and CI, respectively. To simplify the notation, we have rescaled the promoter activities, Eqs. (1) and (2), by the degradation times; that is, the promoter activities are now measured in terms of the steady-state protein concentrations.

In steady states, the production and the degradation of each protein should balance, therefore, the promoter activities and the concentrations of the expressed proteins in the cytoplasm should satisfy the steady-state condition,

\[
p_L([\text{CI}]) = [\text{MOR}]_{\text{total}}
\]

\[
p_R([\text{MOR}], [\text{CI}]) = [\text{CI}]_{\text{total}}
\]

Not all steady states are stable against small perturbations. A steady state is unstable if a perturbation drives the system out of the state. The stability should be determined by the dynamics equations [Eqs. (3) and (4)] (see Supplementary Material). If there are two stable steady states, the system shows bistability.

We assume that the repression factors \( f_{OL} \) and \( f_{OM} \) in Eqs. (1) and (2) are given by the equilibrium statistical weights with which the corresponding operators are not bound by the regulators. This approximation holds when the time that RNA polymerase needs to start elongation after binding to DNA is much shorter than the time scales of binding/unbinding of RNA polymerase and repression factors to the promoter/operator sites.\(^{16}\) The equilibrium statistical weights depend upon the repressor concentrations, and their dependence is characterized by the Hill coefficient and the affinities of the repressors to the operator sites.\(^{17-19}\)

For the MOR-independent repression of \( P_L \), we suppose that \( P_L \) is repressed by CI dimer binding at \( O_L \) and that the dimers are formed in the cytoplasm before binding. Thus, within the above approximation for the repression factor, the \( P_L \) activity is given by

\[
p_L([\text{CI}]) = p_{L0} \cdot \frac{1}{1 + [\text{CI}_2]/K_{OL}}
\]

where the affinity \( K_{OL} \) represents the \( [\text{CI}_2] \) concentration at which \( O_L \) is occupied for 50% of the time. Since the dimer concentration \([\text{CI}_2]\) is related to the monomer concentration \([\text{CI}]\) as

\[
[\text{CI}_2] = \frac{[\text{CI}]^2}{K_{CI2}}
\]

with the dissociation constant \( K_{CI2} \), Eq. (7) may be written as

\[
p_L([\text{CI}]) = p_{L0} \cdot \frac{1}{1 + [\text{CI}]^2/K_{OL}}
\]

with the effective affinity

\[
K_{OL} = \sqrt{K_{CI2} K_{OL}}
\]

for CI concentration. In Fig. 3, the activity of \( P_L \) as a function of \([\text{CI}]\) is plotted by a green line.

As for the MOR-dependent repression of \( P_R \), we will examine two models. In Model A, monomers of MOR and CI may bind cooperatively at \( O_M \), but we do not assume that any MOR:CI complexes formed in the cytoplasm before binding to DNA. In Model B, on the other hand, CI and MOR may associate in the cytoplasm before they bind at \( O_M \). For both models, \( P_R \) is repressed by the formation of the MOR:CI:DNA complex at \( O_M \). The important point in Model B is that the formation of MOR:CI

![Fig. 3. Model A. Promoter activities, \( p_L \) and \( p_R \), and \([\text{CI}]_{\text{total}}\) as a function of free \([\text{CI}]\). Curves represent \( P_L \) activity of Eq. (9) (green line), \( P_R \) activity of Eq. (14) (continuous red line), and \([\text{CI}]_{\text{total}}\) of Eq. (15) (broken red line). The filled red circle at the intersection between the \([\text{CI}]_{\text{total}}\) and \( P_R \) curve represents the steady state.](image-url)
heteromers competes with CI dimer formation by sequestering CI monomers.

Model A

We first consider a MOR:CI:DNA complex containing one MOR and one CI protein as illustrated in Fig. 2. Then, we can approximate the total concentration of MOR unit by the MOR monomer concentration,

$$[\text{MOR}]_{\text{total}} = [\text{MOR}]$$

The activity of the P_R promoter is repressed from the bare activity $pR_0$ by the statistical weight that the operator $O_M$ is not occupied by MOR and CI, $pR([\text{MOR}], [\text{CI}]) = pR_0 \frac{1}{1 + [\text{MOR}][\text{CI}]/(K_{OM})^2}$

The affinity $K_{OM}$ is the concentration $\sqrt{[\text{MOR}] \cdot [\text{CI}]}$ where $O_M$ is occupied by MOR and CI for 50% of the time. The steady state is determined from the steady state condition [Eqs. (5) and (6)] by eliminating the MOR concentrations. With the help of Eq. (11), we obtain

$$pR_p([\text{CI}]) = [\text{CI}]_{\text{total}}$$

which represents the balance between the production and the degradation of CI. This can be solved graphically by plotting both sides as a function of $[\text{CI}]$.

$$pR_p([\text{CI}]) = pR_0 \left[ 1 + \frac{pL_0 K_{OL}/(K_{OM})^2}{([\text{CI}]/K_{OL}) + (K_{OL}/[\text{CI}])^2} \right]^{-1}$$

$[\text{CI}]_{\text{total}} = [\text{CI}] + 2(\frac{[\text{CI}]^2}{K_{CI}})$

Equation (14) represents the $p_R$ activity in the system where $O_M$ is occupied by $P_L$ but $[\text{CI}]$ is controlled externally. Note that the relative strength of the bare promoters, $pL_0$ and $pR_0$, does not affect the system behaviors, such as bistability or repression folds, because there is no direct interaction between MOR and CI in this model.

Model B

In this model, a MOR CI heterodimer is formed in solution before it binds to the putative $O_M$ site to repress $P_R$ (Fig. 2). The activity of the $P_L$ promoter is again given by Eq. (9), but the $P_R$ activity is

$$pR([\text{MOR}], [\text{CI}]) = pR_0 \frac{1}{1 + [\text{MOR}][\text{CI}]/K_{OM}}$$

where $K_{OM}$ now represents the concentration of the MOR CI heterodimer at which $O_M$ is occupied for 50% of the time. The concentration of the MOR CI heterodimer is given as

$$[\text{MOR} \cdot \text{CI}] = \frac{[\text{MOR}] \cdot [\text{CI}]}{K_{\text{MOR} \cdot \text{CI}}}$$

with the dissociation constant $K_{\text{MOR} \cdot \text{CI}}$ for the heterodimer.

The formation of the heterodimers couples the monomer concentrations of CI and MOR through

$$[\text{CI}]_{\text{total}} = [\text{CI}] + [\text{CI}_2] + [\text{MOR} \cdot \text{CI}]$$

$$[\text{MOR}]_{\text{total}} = [\text{MOR}] + [\text{MOR} \cdot \text{CI}]$$

which leads to the competition between the $\text{CI}_2$ formation and the MOR CI formation. Note that $[\text{CI}_2]$ is still given by Eq. (8).

The steady state is determined as in the case of Model A. We consider the $P_R$ activity as a function of $[\text{CI}]$ when $O_M$ is occupied by $P_L$. From Eqs. (17) and (19), $[\text{MOR}]$ is expressed in terms of $[\text{CI}]$ and $[\text{MOR}]_{\text{total}}$

$$[\text{MOR}] = \frac{[\text{MOR}]_{\text{total}}}{1 + [\text{CI}]/K_{\text{MOR} \cdot \text{CI}}}$$

and then, in the steady state where Eq. (5) holds, $[\text{MOR}]_{\text{total}}$ is given by the $P_L$ activity with Eq. (7). Then, the steady-state condition [Eq. (6)] for $P_R$ becomes

$$pR_p \left( \frac{pL([\text{CI}])}{1 + [\text{CI}]/K_{\text{MOR} \cdot \text{CI}}} \right) = [\text{CI}]_{\text{total}}$$

which can be solved graphically with the explicit forms for both sides:

$$pR_p \left( \frac{pL([\text{CI}])}{1 + [\text{CI}]/K_{\text{MOR} \cdot \text{CI}}} \right) = pR_0 \left[ 1 + \frac{pL_0 K_{OL}/(K_{OM})^2}{([\text{CI}]/K_{OL}) + (K_{OL}/[\text{CI}])^2} \right]^{-1}$$

$$[\text{CI}]_{\text{total}} = [\text{CI}] + 2(\frac{[\text{CI}]^2}{K_{CI}}) + pL([\text{CI}]) \frac{[\text{CI}]/K_{\text{MOR} \cdot \text{CI}}}{1 + [\text{CI}]/K_{\text{MOR} \cdot \text{CI}}}$$

where the effective affinities are

$$K_{OM} = \sqrt{K_{\text{MOR} \cdot \text{CI}} \cdot K_{OL}}$$

$$K_{OL} = \sqrt{K_{\text{CI}} \cdot K_{OL}}$$

Equation (21) represents the balance between the production and degradation of CI in the system where MOR is provided by $P_L$. Note that Model B reduces to Model A in the limit of large $K_{\text{MOR} \cdot \text{CI}}$ with $K_{OM}$ being kept constant as has been discussed at the end of Introduction.
Results

We numerically examine the steady states for the two versions of the models we have constructed (Fig. 2).

Model A

In our first model, we study the possibility for bistability in the system where PR is repressed by binding a MOR monomer and a CI monomer to OM without direct interaction between MOR and CI in the cytoplasm. The binding affinities for each of the proteins alone at OM should be negligible because the PR repression requires both of the proteins. Hence, the affinity $K_{OM}$ in Eq. (12) may be considered as the effective binding affinity $K_{OM}$ for MOR and CI with very weak MOR CI formation or as the resulting binding affinities from CI:OM, MOR:OM, and the interaction between the bound proteins.

Figure 3 shows a typical example of the $P_1$ activity of Eq. (9) (green line), the $P_R$ activity of Eq. (14) (continuous red line), and the total concentration of CI, or degradation rate, of Eq. (15) (broken red line) as a function of free [CI] in the logarithmic scale. One can see that $P_1$ (green line) is fully active and produces a lot of MOR at low [CI], whereas its activity is monotonically decreasing with increasing [CI], due to $P_1$ repression by CI$_2$ binding to OM; $P_1$ is virtually shut down beyond [CI]$\approx K_{OM}=\sqrt{K_{CI}} \cdot K_{OL} = 6 \times 10^{-4}$.

On the other hand, the $P_R$ activity (continuous red line) shows a more complicated behavior; that is, $P_R$ is open both at very low and at high [CI] concentrations but repressed at the intermediate concentration. One can understand this behavior by noting that the continuous red line [Eq. (14)] represents the $P_R$ activity in the system where MOR is expressed from $P_1$ under CI control. At low [CI], there is plenty of MOR due to high $P_1$ activity, while at high [CI], no MOR is present. The $P_R$ activity is repressed only at intermediate [CI] because both MOR and CI are necessary for its repression by means of the MOR:CI:DNA complex formation.

In the steady state, the production and the degradation of each protein should balance. Since CI is expressed from $P_R$ and the degradation of CI is assumed to be proportional to the total concentration of CI, the steady states are identified as $P_R(pL([CI]), [CI])=[CI]_{total}$, namely, Eq. (13). We thus find the steady states at the intersection points between the $p_R$ and $[CI]_{total}$ curves represented by Eqs. (14) and (15) (continuous red line and broken red line), respectively.

In Fig. 3, only one intersection point is observed between the two curves, showing that there is only one steady-state solution. This uniqueness of steady state holds true for any given value of the parameters because the activity of $P_R$ given by Eq. (14) never increases faster than proportional to CI, whereas $[CI]_{total}$ by Eq. (15) always increases faster or proportional to CI. Therefore, bistability is never realized in Model A with the assumption of binding of one MOR and one CI for repression of $P_R$.

Variant of Model A

Bistability may be obtained in Model A if a larger number of proteins are allowed to form a complex structure at OM. Suppose $m$ MOR monomers and $c$ CI monomers bind to OM to form the MOR$_m$:CI$_c$:DNA complex that represses transcription from $P_R$, then for the expression for the $P_R$ activity, Eq. (12) should be replaced by

$$pR([MOR], [CI]) = \frac{pR_0}{1 + [MOR]^m [CI]^c / (K_{OM})^{m+c}}$$

while all the other equations remain the same. Using this for the left-hand side of the steady-state Eq. (13) with $[MOR] \ll [CI]$ and Eq. (7) for large [CI], one can see that the largest slope of the $P_R$ activity as a function of [CI] is $2m-c$ in the logarithmic scale. Since the slope of the plot of $[CI]_{total}$ given by Eq. (15) is between 1 and 2, the steady-state Eq. (13) can have more than two solutions with Eq. (25) when $2m-c \geq 2$. In the case of $m=1$, we could obtain multiple solutions with $c=1$ or 2, that is, two MOR monomers binding together with one or two CI monomers at OM. Examples for Model A with $(m,c)=(2,1)$ and $(2,2)$ are shown in Fig. 4. In each example, the intersections between the continuous red line (the $P_R$ activity) and the broken red line (the total CI concentration) represent steady-state solutions. One can see that there are three solutions for each case in Fig. 4.

Dynamical analysis shows that the steady state in the middle marked by an open red circle is unstable against small fluctuations, and the states at the ends marked by filled red circles are stable (see Supplementary Material for details). Full analysis requires Eqs. (3) and (4), but the stability may be understood in the following way: For the steady state in the middle, if the CI monomer concentration increases by fluctuation from the steady value of [CI], the CI production from $P_R$ will increase more than the increase in degradation given by [CI]$_{total}$ as is seen in Fig. 4, where the continuous red line of the $P_R$ activity goes above the broken red line of [CI]$_{total}$, upon increasing [CI] from the middle steady state. This means that such a fluctuation causes further increases of [CI]; consequently, the state is driven out of the steady state. On the other hand, the steady states at both ends represent stable states. A fluctuation towards larger [CI] leads to insufficient CI production in comparison with the CI degradation; that is, the continuous red line goes under the broken red line as [CI] increases. This brings the system back to the original state; therefore, they are stable. Thus, the system has two stable steady states, which leads to bistability.$^\dagger$

$^\dagger$ Note that this simplified analysis is based on the assumption that Eq. (5) always holds; that is, the response of $P_1$ is much faster than that of $P_R$. For general case, see Supplementary Material.
The promoter activities in the two stable states can be determined from the graphic representation in Fig. 4. The PR activity is read from the ordinates of the intersection points (green circles), which allows us to estimate the repression folds for PL and PR between the two stable states. The state at the right represents the immune state with open PR and repressed PL, while the one at the left represents the anti-immune state with open PL and repressed PR.

The relative activities between the two states should be compared with the promoter activities obtained from the in vivo measurements.\(^9\) The PL repression fold should be approximately 1000-fold in the immune state and PR is repressed approximately 100-fold in the anti-immune state. To reproduce these repression folds in Model A with \((m,c)=(2,1)\) and \((2,2)\), we test the three parameters, \(K_{CI1}\), \(K_{OL}\), and \(K_{OM}\), representing the dimerization constant of CI, the effective binding constant of CI monomer at OL, and the binding constant of the MOR:CI complex at OM, respectively, by setting the criterion that the PR repression fold should be in the range from 50 to 200 \([50 < pR(open)/pR(closed) < 200]\) and the PL repression fold should at least be 500 \([pL(open)/pL(closed) > 500]\).

Figure 5 shows the distributions of accepted values of parameters out of randomly chosen values in the logarithmic scale. \(K_{OL}\) and \(K_{OM}\) are narrowly distributed while \(K_{CI1}\) are larger than \(10^{-1}\) for \((m,c)=(2,1)\) and larger than \(10^{-2}\) for \((2,2)\) in the unit of CI concentration at full activity of PR. One can also see that the accepted values for \(K_{CI1}\) are much larger than those for \(K_{OL}\). This suggests that, in order for Model A to work, CI must exist as a monomer and act by cooperative binding to form CI\(_2\) at OL, when repressing PL.

Figure 6 shows the parameters that satisfy the criterion only for the PR repression fold versus resulting PL repression fold (two left columns for Model A and two right columns for Model B). The vertical green lines are drawn at the PR repression fold 500; thus, only the plots on the right side of the lines should be accepted by the repression fold criterion. From the plots for \(K_{CI1}\) for Model A, one can see that the relatively high values for \(K_{CI1}\) in this model come from the requirement for the large PL repression fold. This can be understood as follows: In order to achieve large repression fold for PL, the difference in [CI] for the two steady states should be large, which, in turn, requires smaller slope in the [CI]\(_{total}\) curve, namely, larger \(K_{CI1}\) because the slope in [CI]\(_{total}\) changes from 1 to 2 around [CI]\(_{total}\) \(\approx K_{CI1}\) (Fig. 4).

**Model B**

Now, we consider the possibility that PR is repressed by a MOR:CI complex formed in the cytoplasm before binding to DNA. Examples for Model B are shown in Fig. 7, where one can see the PL activity of Eq. (9) (green line), the PR activity of Eq. (22) (continuous red line), and the total density of CI of Eq. (23) (broken red line) as a function of [CI] in the logarithmic scale. All of them have three steady states.

A striking difference from the case of Model A is that the [CI]\(_{total}\) (broken red line) can be nonmonotonic. Therefore, for a given [CI]\(_{total}\) within a certain range, there exist three possible states with different [CI]. This suggests that the bistability could be obtained for the system with PL even if PR were not regulated, namely, even if PR would produce CI at a fixed rate within the range. This bistability is due to the MOR:CI heteromer formation in the cytoplasm; PL produces MOR, which sequesters its own repressor, that is, CI, by forming MOR:CI. For a given [CI]\(_{total}\), the state at low [CI] is the state where most of the CIs are incorporated in MOR:CI.
heterodimers due to the MOR produced by P_L, while the state at high [CI] is the state where most of the CIs are in the dimer with P_L being repressed. The state in the middle is unstable. Such bistability is, of course, not the bistability observed in the experiments, but one can see that this feature of behavior in [CI]_{total} (broken red line) makes it easier to have three intersections with the P_R activity curve (continuous red line) than in the case of Model A.

According to the stability criterion we discussed, the steady state at both ends is stable while the state in the middle is unstable even for the system where both P_R and P_L are regulated. Full analysis, however, shows that there are some cases where the states at both ends can be unstable although the stability criterion is correct for most cases (see Supplementary Material). We analyze the bistability based upon the stability criterion, ignoring the small possibility that the states at both ends could be unstable.

We also examine Model B in the case where a larger complex, MOR\textsubscript{CI\textsuperscript{c}}\textsubscript{m}, represses P_R. Detailed formalism is given in the Appendix.

Figure 8 shows the plots for the extended Model B with (m,c)=(2,1). This version of the model shows sharper transition between the MOR\textsubscript{2CI} regime and the CI\textsubscript{2} regime for the form of CI protein as one can see in the lower graphs for the CI ratio. As for the form of MOR, substantial fraction of MOR\textsubscript{2CI} appears only in the intermediate range of [CI] in contrast to the case of the MOR\textsubscript{CI} heterodimer in Fig. 7. This is because there are not enough MORs in the high [CI] region to form MOR\textsubscript{2CI} complex because it requires two MOR proteins.

In the two right columns of Fig. 6, the parameters that give the P_R repression fold in the range [50,200]...
Fig. 6 (legend on previous page)

Modeling of the Genetic Switch of TP901-1
are plotted versus the resulting PL repression fold for Model B. The $2 \times 10^5$ parameter sets are chosen randomly over the range of $[10^{-7},1]$ for $K_{OL}$, $K_{OM}$, $K_{CI_2}$, and $K_{MOR:CI}$. One can see that a broad range of parameter sets satisfy the repression fold criterion for PL, but the resulting repression folds for PR are limited to the region larger than 5000 for $(m,c)=(1,1)$ and 50 for $(m,c)=(2,1)$.

**Discussion**

**Summary**

The bacteriophage TP901-1 has provided us with a conceptually new design of a genetic switch, in which the interaction between two antagonistic regulators, CI and MOR, is essential. Bistability between the immune and the anti-immune states has been demonstrated with a genetic switch that consists of the two divergently oriented promoters $P_L$ and $P_R$, the two promoter-proximal genes, $cl$ and $mor$, and only one of the three CI operator sites $O_L$ on a low-copy-number plasmid. The repression folds in the two states have been determined by *in vivo* measurements as around 1000-fold for $P_L$ repression in the immune state and around 100-fold for $P_R$ repression in the anti-immune state.

We constructed mathematical models for this cloned bistable system, assuming a putative operator $O_M$ to regulate $P_R$ (Fig. 1b). We assumed that $P_L$ is repressed by $CI_2$ bound to $O_L$, whereas $P_R$ is repressed by the $MOR_{CI_2}$-DNA complex on $O_M$. We examined two types of models: one where MOR and CI interact only on DNA (Model A) and the other where MOR and CI form $MOR_{CI_2}$ complex in the cytoplasm first and then the complex binds to $O_M$ (Model B). For each model, we tested bistability and performed parameter scans using the criterion that the repression folds should be consistent with the experiments.

Our results are summarized as follows: For Model A, (i) the system shows bistability only when $2m - c \geq 2$, (ii) the possible values for the operator...
affinities $K_{\text{OL}}$ and $K_{\text{OM}}$ are narrowly distributed, (iii) the possible dissociation constant $K_{\text{CI}2}$ is much larger than the operator affinity $K_{\text{OL}}$ due to the large repression fold for PL, and (iv) the possible value for $K_{\text{CI}2}$ is bounded by the relatively large lower limit. The accepted ranges for the parameters are listed in Table 1. For Model B, (i) the bistability is robust due to the sequestration of CI by the MOR complex formation, (ii) large parameter regions are allowed by the repression fold criterion, and (iii) the PL repression fold is bounded by the lower limit for the parameters that are consistent with the PR repression fold: $pL_{\text{(open)}}/pL_{\text{(closed)}} > 5000$ for $(m,c) = (2,1)$ and $> 50$ for $(m,c) = (2,2)$. The accepted ranges for the parameters are listed in Table 2.

### Validity of the models

In order to assess validity of the models, we have to determine our unit for CI concentration first. We employed the unit where $[\text{CI}]$ is measured by the bare promoter activities, $pL_0$ and $pR_0$, are equal in the left panel while the bare activity $pL_0$ is 10 times stronger than $pR_0$ in the right panel. The upper graphs show examples of the promoter activities, $pL$ and $pR$, and $[\text{CI}]_{\text{total}}$ as a function of $[\text{CI}]$. The repression folds of $pL$ and $pR$ are approximately 1000 and 50, respectively, in the left panel and approximately 1000 and 60 in the right panel. In contrast to the case with $(m,c) = (1,1)$ in Fig. 7, the heteromer MOR$_2$:CI is not formed at high $[\text{CI}]$ because the $pL$ promoter is closed faster than the MOR$_2$:CI is formed. Note that the effective affinity of OM is $K_{\text{OM}} \equiv \left( \frac{K_{\text{OM}} K_{\text{MOR}} K_{\text{CI}2}}{K_{\text{MOR}} K_{\text{CI}2}} \right)^{1/3}$.

#### Table 1. Accepted ranges of parameters for Model A

<table>
<thead>
<tr>
<th>$(m,c)$</th>
<th>(2,1)</th>
<th>(2,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{\text{OL}}$</td>
<td>$\sim 10^{-2}$</td>
<td>$10^{-3} \sim 10^{-2}$</td>
</tr>
<tr>
<td>$K_{\text{OM}}$</td>
<td>$\sim 10^{-2}$</td>
<td>$10^{2} \sim 10^{-2}$</td>
</tr>
<tr>
<td>$K_{\text{CI}2}$</td>
<td>$\sim 10^{-2}$</td>
<td>$\sim 10^{-2}$</td>
</tr>
</tbody>
</table>

The values are given in the unit of the concentrations that correspond to $[\text{CI}]$ and $[\text{MOR}]$ at the full activity of the promoter $P_\text{R}$ and $P_\text{L}$, respectively.

#### Table 2. Accepted ranges of parameters for Model B

<table>
<thead>
<tr>
<th>$(m,c)$</th>
<th>(1,1)</th>
<th>(2,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{\text{OL}}$</td>
<td>$\leq 3 \times 10^{-3}$</td>
<td>$(3 \times 10^{-2}) \sim 3 \times 10^{-2}$</td>
</tr>
<tr>
<td>$K_{\text{OM}}$</td>
<td>$\leq 5 \times 10^{-3}$</td>
<td>$(10^{-2}) \sim 3 \times 10^{-2}$</td>
</tr>
<tr>
<td>$K_{\text{CI}2}$</td>
<td>$\leq 3 \times 10^{-1}$</td>
<td>$(10^{-7}) \sim 1$</td>
</tr>
</tbody>
</table>

The values are given in the unit of the concentrations that correspond to $[\text{CI}]$ and $[\text{MOR}]$ at the full activity of the promoter $P_\text{R}$ and $P_\text{L}$, respectively. The values in parentheses are the lower limits when the fold criterion for $pL_0$ is restricted to $50 < pL_{\text{(open)}}/pL_{\text{(closed)}} < 5000$. The ranges for $K_{\text{CI}2}$ cannot be set because the accepted values extend over the whole tested range.
concentration in the steady state with the full activity of \( P_L \). This concentration should be compared with \([CI]\) in the immune state of in \( \textit{vivo} \) experiment on the system with around 10 copy-number plasmid containing the modified switch. We estimate this as follows by using the value of 300 nM for the CI concentration in the lysogenic/immune state of the wild lambda phage.\(^{20}\) First, we assume that this concentration is comparable with that for wild-type TP901-1 with a single copy. Then, we multiply this by the following two factors: the factor 10 of the copy number of plasmid and the factor 100 of the relative activity of \( P_R \) in our modified system in comparison with the wild-type switch.\(^{7}\) With these factors, we estimate that \( p_{R_{\text{d}}}=1 \), that is, \([CI]\) at the full \( P_R \) activity in the present system, could be well over the \( 10^5 \)-nM scale.

**Model A**

With this unit, we might be able to rule out Model A based upon the estimated values of \( K_{OL} \) and \( K_{CI2} \). The possible value of \( K_{OL} \) in Model A is between around \( 10^{-3} \) and \( 5 \times 10^{-2} \) (Table 1), but this contradicts the \( \textit{in vitro} \) estimate of 28 nM for the CI concentration at which \( O_L \) is occupied by CI for 50% of the time.\(^{10}\) The lower limit of \( K_{OL} \sim 10^{-3} \) in Model A should correspond to 100 nM or quite possibly even larger, but it is already well above 28 nM, that is, the \( \textit{in vitro} \) estimate for repressor–DNA affinity for TP901-1.

We also found that the large repression fold of \( P_L \) entails \( K_{CL} \gg K_{OL} \) for Model A. This means that CIs exist as monomers in the cytoplasm and form CI\(_2\) when they bind to \( O_L \), but this is in contrast with many phage-encoded repressor proteins, such as those encoded by phage lambda, 434, and 186, which tend to exist as dimers or higher oligomers in solution.\(^{21-24}\) Actually, most of the 434 and lambda repressors exist in the dimeric conformation at nanomolar concentrations.\(^{21,23}\) Our Model A challenges the presumption that the formation of dimers is a prerequisite for its specific DNA binding.

**Model B**

We found broader distribution of parameter sets that satisfy the repression fold criterion for Model B. In particular, we did not find lower bounds for possible \( K_{CL} \) in contrast to the case of Model A.

In the comparison of the two variants of Model B, our results show that the model with the formation of \( \text{MOR}_2\text{CI} \) complex is more favorable than that with \( \text{MORCI} \). For the model with \( \text{MORCI} \), the \( P_L \) repression fold turned out to be always large than 5000 for the parameters that give the \( P_R \) repression fold between 50 through 200. Such a high repression fold of \( P_L \) has never been actually observed \( \textit{in vitro} \). On the other hand, for the model with \( \text{MOR}_2\text{CI} \), the lower bound for the resulting \( P_L \) repression fold can be as low as 50, which covers the observed range of the \( P_L \) repression fold. As for the parameter ranges, this variant of Model B gives \( K_{OM} \simeq 0.2 \times 10^{-4} \) and \( K_{\text{MORCI}} \simeq 10^{-3} \) when the repression fold for \( P_L \) is smaller than 5000 (Fig. 6).

**Experimental test**

One of the distinguishing consequences of Model B is that the system with uncontrolled \( P_R \) can be bistable because of the sequestration of CI by \( \text{MORCI} \) complex formation. Even if \( P_R \) produces CI at a constant rate, there can be the two stable states: one with the repressed \( P_L \) and the other with the derepressed \( P_L \). Such a mechanism of bistability has been proposed by Francois and Hakim as a theoretical possibility.\(^{12,13}\) Our study suggests that this mechanism is employed in TP901-1.

This may be tested experimentally for the genetic switch of phage TP901-1 by measuring the promoter activity of \( P_L \) in systems containing a functional \( mor \) gene and expressing CI from uncontrolled \( P_R \) promoters at constant but various rates.

Plotting the \( P_L \) activity of each system \textit{versus} the uncontrolled \( P_R \) activity, one should find a characteristic feature for bistability as in Fig. 9, where the \( P_L \) activity is double-valued for a certain range of the \( P_R \) activity.

**Concluding remark**

The genetic switching mechanism in TP901-1 is remarkably robust. The modified system studied here with only one operator \( O_L \) contains 100 times more CI molecules in its immune state than the wild-type genetic switch with all of the three operators on plasmids yet still shows bistability. Our model study suggests that the robustness of the genetic switch in

\[
(m_c)=(2,1), \quad p_{L_{\text{d}}}=1, \quad p_{R_{\text{d}}}=1
\]

\[
K_{OL}=7.69 \times 10^{-4}, \quad K_{CI2}=1.55 \times 10^{-3}
\]

\[
K_{OM}=4.18 \times 10^{3}, \quad K_{\text{MORCI}}=1.90 \times 10^{-2}
\]

![Fig. 9. The \( P_L \) activity (or \([\text{MOR}]_{\text{total}}\)) \textit{versus} the \( P_R \) activity (or \([CI]_{\text{total}}\)) for the system with uncontrolled \( P_R \). The continuous lines show the stable states and the broken line shows the unstable state.](image-url)
TP901-1 is brought about by sequestration of CI through MOR:CI complex formation in cytoplasm.

Acknowledgements

This work was supported by the Danish National Research Foundation through the Center for Models of Life.

Appendix A. Formalism for Model B with \((m,c)=(2,1)\)

Here, we present some of the formulae we used for Model B with \((m,c)=(2,1)\). For this case, the promoter \(p_R\) activity is a function of \(\text{MOR}_2\text{CI}\) concentration,

\[
p_R([\text{MOR}], [\text{CI}]) = \frac{pR_0}{1 + [\text{MOR}_2\text{CI}]/K_{\text{OM}}},
\]

and \([\text{MOR}_2\text{CI}]\) is given by

\[
[\text{MOR}_2\text{CI}] = \frac{[\text{MOR}]^2 \cdot [\text{CI}]}{K_{\text{MOR,CI}}^2},
\]

with the dissociation constant \(K_{\text{MOR,CI}}\).

The total concentrations of CI and MOR are now

\[
[\text{CI}]_{\text{total}} = [\text{CI}] + 2[\text{Cl}_2] + [\text{MOR}_2\text{CI}],
\]

\[
[\text{MOR}]_{\text{total}} = [\text{MOR}] + 2[\text{MOR}_2\text{CI}],
\]

thus, the corresponding equations with Eq. (20) in the main text becomes

\[
[pR] = \frac{pR_0}{1 + [\text{MOR}]^2/[\text{CI}]/K_{\text{OM}}}.
\]

Thus, both sides of the steady-state condition Eq. (6) are now given by

\[
pR([\text{MOR}], [\text{CI}]) = \frac{pR_0}{1 + [\text{MOR}]^2/[\text{CI}]/K_{\text{OM}}} (31)
\]

\[
[\text{CI}]_{\text{total}} = [\text{CI}] + \frac{2[\text{Cl}_2]}{K_{\text{Cl}_2}} + \frac{1}{2} ([\text{MOR}]_{\text{total}} - [\text{MOR}]) (32)
\]

with \([\text{MOR}]\) by Eq. (30) and \([\text{MOR}]_{\text{total}}\) by Eq. (5). The effective affinity \(K_{\text{OM}}\) is defined as

\[
K_{\text{OM}} = (K_{\text{OM}}K_{\text{MOR,CI}}^2)^{1/3}. (33)
\]

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.08.075

References


Supplementary material for
“Model Analysis of the Genetic Switch Isolated from the Temperate Bacteriophage TP901-1: Repressor Sequestration Effect”

Hiizu Nakanishi\textsuperscript{a}, Margit Pedersen\textsuperscript{b}, Anne K. Alsing\textsuperscript{b}, and Kim Sneppen\textsuperscript{b}

\textsuperscript{a} Department of Physics, Kyushu University 33, Fukuoka 812-8581, Japan
\textsuperscript{b} Niels Bohr Institute, Copenhagen University, Denmark

The stability criterion for the steady state used in the manuscript is examined. Dynamical analysis shows that the criterion is valid as long as the \([CI]_{\text{total}}\) curve is a monotonically increasing function. In Model B, however, the \([CI]_{\text{total}}\) curve has a part with negative slope for some parameter region, in which case the stability cannot be determined only by comparing the slopes of the curves. The steady state at the left side could be unstable when it is located in the region where \([CI]_{\text{total}}\) is decreasing.

1. Steady States

The steady states satisfy the self-consistent conditions

\[ pL([CI]) = [MOR]_{\text{tot}} ([MOR], [CI]) \]  
\[ pR([MOR],[CI]) = [CI]_{\text{tot}} ([CI],[MOR]) \]

with

\[ [CI]_{\text{tot}} ([CI],[MOR]) = [CI] + 2[CI] + [MOR \cdot CI] \]  
\[ [MOR]_{\text{tot}} ([MOR],[CI]) = [MOR] + [MOR \cdot CI]. \]

Here, \([CI]_2\), \([MOR \cdot CI]\) are equilibrium concentrations of the protein complexes. We determined the steady solutions graphically by looking for intersections of the following two curves as a function of \([CI]\), i.e. the production curve and the \([CI]_{\text{total}}\) curve that represents degradation rate:

\[ pR = pR([MOR],[CI]) \]
\[ [CI]_{\text{tot}} = [CI]_{\text{tot}} ([CI],[MOR]). \]

with \([MOR]\) being a function of \([MOR]_{\text{tot}}\) and \([CI]\) derived from the relation \([MOR]_{\text{tot}} ([MOR],[CI])\), and \([MOR]_{\text{tot}}\) being given by \(pL([CI])\),

\[ [MOR] = [MOR]([MOR]_{\text{tot}},{[CI]}) = [MOR]([pL([CI])],[CI]). \]

2. Stability criterion

In the text, the stability of the steady state is determined by the simple criterion. Let the slope of the production curve and the degradation curve be denoted by

\[ \frac{dpR}{d[CI]} \]  
\[ \frac{d[CI]_{\text{tot}}}{d[CI]} \]

respectively. Then the criterion is
the steady state is stable if \[ \frac{dpR}{d[CI]} < \frac{d[CIcon]}{d[CI]} \] (8)

the steady state is unstable if \[ \frac{dpR}{d[CI]} > \frac{d[CIcon]}{d[CI]} \]

at the corresponding intersection.

This criterion is simple and plausible, but based on the single variable picture although the system has at least two dynamical variables: [CI] and [MOR]. The full analysis for stability requires dynamical consideration.

3. Dynamical Analysis of Stability

The dynamics for the protein concentrations is given by the set of equations:

\[ \frac{d}{dt}[MOR]_{ext} = \frac{1}{\tau_M} \left( pL([CI]) - [MOR]_{ext} \right) \] (9)

\[ \frac{d}{dt}[CI]_{ext} = \frac{1}{\tau_C} \left( pR([MOR],[CI]) - [CI]_{ext} \right). \] (10)

The total concentrations \([CI]_{ext}\) and \([MOR]_{ext}\) are given by eqs.(3) and (4). Basic assumption for this is that the equilibration among protein complexes in cytoplasm is much faster than the decay rates of CI and MOR: \(1/\tau_C\) and \(1/\tau_M\).

Consider the steady solution with \([CI]^*\) and \([MOR]^*\), which satisfies eqs.(1) and (2). Suppose the steady state is perturbed by small fluctuation as

\[ [CI] = [CI]^* + \delta[CI] \]

\[ [MOR] = [MOR]^* + \delta[MOR], \]

and see if the small deviation will grow or decay in time.

By inserting these into eqs.(9) and (10), we obtain the equations for the time evolution of \(\delta[CI]\) and \(\delta[MOR]\).

\[ \left( \begin{array}{c} \delta C \\ \delta M \end{array} \right) = \left( \begin{array}{cc} \frac{\partial M}{\partial C} & \frac{\partial M}{\partial M} \\ \frac{\partial C}{\partial C} & \frac{\partial C}{\partial M} \end{array} \right) \left( \begin{array}{c} \frac{1}{\tau_M} \left( \frac{\partial L}{\partial C} - \frac{\partial M}{\partial C} \right) \\ \frac{1}{\tau_M} \left( \frac{\partial M}{\partial C} \right) \end{array} \right) \left( \begin{array}{c} \frac{1}{\tau_C} \left( \frac{\partial R}{\partial C} - \frac{\partial C}{\partial C} \right) \\ \frac{1}{\tau_C} \left( \frac{\partial C}{\partial M} - \frac{\partial C}{\partial M} \right) \end{array} \right) \delta C \] (14)

\[ \delta M \]

where we employ the abbreviated notations:
M_t \equiv [\text{MOR}]_{\text{tot}}, \quad M \equiv [\text{MOR}]_t, \quad C_t \equiv [\text{CI}]_{\text{tot}}, \quad C \equiv [\text{CI}], \quad L \equiv pL, \quad R \equiv pR.

We further abbreviate the notation as

\[ M_{t,c} \equiv \left( \frac{\partial [\text{MOR}]_{\text{tot}}}{\partial C} \right)^t, \quad L_{c} \equiv \left( \frac{\partial L}{\partial C} \right)^t = \left( \frac{\partial pL([\text{CI}])}{\partial [\text{CI}]} \right)^t, \quad \text{etc.} \]

then eq.(21) is expressed as

\[
\begin{pmatrix}
M_{t,c}, & M_{t,M}\end{pmatrix} \begin{pmatrix}
\delta \dot{C} \\
\delta \dot{M}
\end{pmatrix} = \begin{pmatrix}
\frac{1}{\tau_M} (L_c - M_{t,c}), & -\frac{1}{\tau_M} M_{t,M} \\
1(1 - C_{t,c}), & \frac{1}{\tau_M} (R_M - C_{t,M})
\end{pmatrix} \begin{pmatrix}
\delta C \\
\delta M
\end{pmatrix}.
\]

(22)

4. The criterion is always valid for Model A:

In the case of Model A, \([\text{MOR} \cdot \text{CI}] = 0\), thus we have

\[ C_{t,M} = 0, \quad M_{t,M} = 1, \quad M_{t,c} = 0, \]

then eq.(22) becomes

\[
\begin{pmatrix}
0, & 1\end{pmatrix} \begin{pmatrix}
\delta \dot{C} \\
\delta \dot{M}
\end{pmatrix} = \begin{pmatrix}
\frac{1}{\tau_M} L_c, & -\frac{1}{\tau_M} \\
1(R_c - C_{t,c}), & \frac{1}{\tau_M} R_M
\end{pmatrix} \begin{pmatrix}
\delta C \\
\delta M
\end{pmatrix}.
\]

(23)

Now we assume the solution as

\[ \delta C, \quad \delta M \propto e^{\omega t}, \]

then we have

\[
\begin{pmatrix}
\frac{1}{\tau_M} L_c, & -\omega \\
\frac{1}{\tau_M} (R_c - C_{t,c}) - C_{t,M} \omega, & \frac{1}{\tau_M} R_M
\end{pmatrix} \begin{pmatrix}
\delta C \\
\delta M
\end{pmatrix} = 0.
\]

(24)

The condition that this equation has non-zero solution gives

\[
-C_{t,c} \omega^2 + \frac{1}{\tau_M} (R_c - C_{t,c}) - \frac{1}{\tau_M} C_{t,c} \omega + \frac{1}{\tau_M} [L_c R_M + R_c - C_{t,c} ] = 0
\]

(25)

If all the solutions \(\omega\) have a negative real part, the state is stable, whereas the state is unstable if there is a solution with a positive real part.

Note that the slopes of the production curve and the degradation curve are given by

\[
\frac{d pR}{dC} \equiv \left( \frac{d pR([\text{MOR}]_t[\text{CI}])}{d[\text{CI}]} \right) = L_c R_M + R_c
\]

(26)
Using these, eq.(25) becomes

\[
-\frac{d\text{Cl}_{\text{tot}}}{d\text{Cl}} \omega^2 + \left[ \frac{1}{\tau_C} \left( \frac{d\text{pR}}{d\text{Cl}} - \frac{d\text{Cl}_{\text{tot}}}{d\text{Cl}} - L_C R_M \right) - \frac{1}{\tau_M} \frac{d\text{Cl}_{\text{tot}}}{d\text{Cl}} \right] \omega \\
+ \frac{1}{\tau_C \tau_M} \left( \frac{d\text{pR}}{d\text{Cl}} - \frac{d\text{Cl}_{\text{tot}}}{d\text{Cl}} \right) = 0
\]  

(28)

Note that

\[ L_C, R_C, R_M < 0. \]

1. (1) The case of \( \tau_M \ll \tau_C \)

One solution is of order \( 1/\tau_M \) and the other is of order \( 1/\tau_C \).

\[
\omega = \left\{ \begin{array}{c}
\frac{-1}{\tau_M} \\
\frac{1}{\tau_C} \left[ \frac{d\text{Cl}_{\text{tot}}}{d\text{Cl}} \right]^{-1} \left[ \frac{d\text{pR}}{d\text{Cl}} - \frac{d\text{Cl}_{\text{tot}}}{d\text{Cl}} - L_C R_M \right] \\
\end{array} \right.
\]

This shows the criterion (8) is valid.

2. (2) The case of \( \tau_C \ll \tau_M \)

One solution is of order \( 1/\tau_M \) and the other is of order \( 1/\tau_C \).

\[
\omega = \left\{ \begin{array}{c}
\frac{1}{\tau_C} \left[ \frac{d\text{Cl}_{\text{tot}}}{d\text{Cl}} \right]^{-1} \left[ \frac{d\text{pR}}{d\text{Cl}} - \frac{d\text{Cl}_{\text{tot}}}{d\text{Cl}} - L_C R_M \right] \\
\frac{-1}{\tau_M} \left[ \frac{d\text{pR}}{d\text{Cl}} - \frac{d\text{Cl}_{\text{tot}}}{d\text{Cl}} - L_C R_M \right] \\
\end{array} \right.
\]

The criterion (8) is also valid because

\[ \frac{d\text{Cl}_{\text{tot}}}{d\text{Cl}} > 0, \quad L_C R_M > 0. \]

3. (3) General case

The criterion (8) can be shown to be valid because eq.(25) always has two real solutions, \( \omega_0 \) and \( \omega_1 \), and the sum of the two solutions is negative when the slope of the production curve is less steep than that of the degradation curve, i.e.
\( \omega_1 + \omega_2 < 0 \) when \( \frac{dpR}{dCI} \frac{d\text{Cl}_{\text{tot}}}{dCI} < 0 \).

The positivity of the discriminant \( D \) of eq.(25):

\[
D = \left[ \frac{1}{\tau_c^2} \left( R_c - C_{i,c} \right) - \frac{1}{\tau_m^2} C_{i,c} \right]^2 + 4 \frac{C_{i,c}}{\tau_m \tau_c} \left( L_c R_m + R_c - C_{i,c} \right) - \frac{1}{\tau_c \tau_m} C_{i,c} R_m L_c > 0, \]

thus eq.(25) has two real solutions.

5. The criterion is valid for Model B as long as the degradation curve has positive slope, but the state may be unstable otherwise.

In this model, the existence of \( \text{MOR} \cdot \text{Cl} \) makes the expressions for the slopes of the production curve and the degradation curve a bit more complicated:

\[
\frac{dpR}{dCI} = \frac{d\text{Cl}_{\text{tot}}}{dCI},
\]

\[
\frac{d\text{Cl}_{\text{tot}}}{dCI} = \frac{d\text{Cl}}{d\text{Cl}_{\text{tot}}},
\]

where the variables that kept constant upon partial differentiation are explicitly indicated as

\[
\left( \frac{\partial M}{\partial M_i} \right) = \frac{\partial}{\partial \text{MOR}_{\text{tot}}} \left( \text{MOR} \left( \text{MOR}_{\text{tot}}, [\text{Cl}] \right) \right)
\]

whenever it could be ambiguous. In the derivation, we have used the relations
\[
\left( \frac{\partial M}{\partial M} \right)_{jC} = \left( \frac{\partial M}{\partial M} \right)_{jC}^{-1}, \quad \left( \frac{\partial M}{\partial C} \right)_{jM} \left( \frac{\partial M}{\partial M} \right)_{jC} \left( \frac{\partial C}{\partial M} \right)_{jM} = -1.
\]

1. Stability of the system with PL with externally controlled [CI]_{tot}

First, we will examine the stability of the system without PR, CI being provided externally. The system is shown to be bistable for some parameter region. Based upon the approximation that the relaxation in the solution is much faster than the protein production rate by PL, we consider the system where [CI] and [MOR] satisfy

\[
[CI]_{tot} = [CI]_{tot}([CI],[MOR]) \quad (31)
\]

\[
\frac{d}{dt}[MOR]_{tot} = \frac{1}{\tau_M} \left( pL([CI]) - [MOR]_{tot} \right), \quad (32)
\]

thus the deviation from the steady state follows

\[
C_{i,C} \delta C + C_{i,M} \delta M = 0 \quad (33)
\]

\[
M_{i,C} \delta \dot{C} + M_{i,M} \delta \dot{M} = \frac{1}{\tau_M} \left( L_{i,C} \delta C - M_{i,M} \delta C - M_{i,M} \delta M \right), \quad (34)
\]

which results in

\[
\delta \dot{C} = \frac{1}{\tau_M} \frac{(L_{i,C} - M_{i,C})C_{i,M} + M_{i,M}C_{i,C}}{M_{i,M}C_{i,M} - M_{i,M}C_{i,C}} \delta C
\]

\[
\quad = -\frac{1}{\tau_M} \left( \frac{M_{i,M}}{A} \frac{d[CI]_{tot}}{dCI} \right) \delta C, \quad (35)
\]

with the notation

\[
A \equiv C_{i,C}M_{i,M} - C_{i,M}M_{i,C} > 0,
\]

whose inequality can be shown from the actual expressions of [CI]_{tot} and [MOR]_{tot}, (3) and (4).

Therefore, we have

\[
\frac{d[CI]_{tot}}{dCI} < 0 \text{ unstable}
\]

\[
\frac{d[CI]_{tot}}{dCI} > 0 \text{ stable}
\]

2. Stability of genetic switch with PL and PR:

The growth rate \( \omega \) is determined by the characteristic equation
\[
\frac{1}{\tau_M} (L_C - M_{r,t}) - M_{r,t} \omega_t, \quad \frac{1}{\tau_M} M_{i,t} - M_{i,M} \omega_t \]
\[
\frac{1}{\tau_C} (R_C - C_{r,C}) - C_{r,C} \omega_t, \quad \frac{1}{\tau_C} (R_M - C_{r,M}) - C_{r,M} \omega_t
\]
\[= 0, \quad (36)\]
which can be expanded as
\[
\omega^2 [M_{r,t} C_{r,M} - C_{r,t} M_{i,M}]
\]
\[+ \omega \left[ \frac{1}{\tau_C} (-M_{r,t} (R_M - C_{r,M}) + M_{r,t} (R_C - C_{r,C}) - \frac{1}{\tau_M} (C_{r,M} (L_C - M_{r,C}) + C_{r,C} M_{r,t})) \right]
\[+ \frac{1}{\tau_M \tau_C} [(L_C - M_{r,t}) (R_M - C_{r,M}) + (R_C - C_{r,C}) M_{i,t}] = 0. \quad (37)\]

This can be put in the form
\[
-\omega^2 \frac{A}{M_{r,t}} + \omega \left[ \frac{1}{\tau_C} \left( \frac{d p R}{d C I_{mm}} - \frac{d C I_{mm}}{d C I} \right) - \frac{1}{\tau_M} \frac{d C I_{mm}}{d C I} \right] - \frac{1}{\tau_D} \frac{L_C (R_M - C_{r,M})}{M_{i,t}}
\]
\[+ \frac{1}{\tau_M \tau_C} \left[ \frac{d p R}{d C I} - \frac{d C I_{mm}}{d C I} \right] \frac{d C I_{mm}}{d C I} = 0. \quad (38)\]

This is almost the same with the corresponding equation for Model A eq.(28).

4. (1) The case of \(\tau_M \ll \tau_C\)

One solution is of order \(1/\tau_M\) and the other of order \(1/\tau_C\):
\[
\omega = \begin{cases} \frac{1}{\tau_M} M_{r,t} \left( \frac{d C I_{mm}}{d C I} \right) \\ \frac{1}{\tau_C} \left[ \frac{d p R}{d C I} - \frac{d C I_{mm}}{d C I} \right] \left( \frac{d C I_{mm}}{d C I} \right)^{-1} \end{cases}
\quad (39)\]

Therefore, the criterion (8) is valid as long as the slope of degradation curve is positive, i.e. \((d C I_{mm}/d C I) > 0\), but the state is always unstable for the negative slope for the degradation curve, i.e. \((d C I_{mm}/d C I) < 0\).

5. (2) The case of \(\tau_C \ll \tau_M\)

\[
\omega = \begin{cases} \frac{1}{\tau_M M_{r,t}} \left[ \frac{d p R}{d C I} - \frac{d C I_{mm}}{d C I} \right] \frac{L_C (R_M - C_{r,M})}{M_{i,t}} \\ -\frac{1}{\tau_M} \left[ \frac{d p R}{d C I} - \frac{d C I_{mm}}{d C I} \right] \left( \frac{d C I_{mm}}{d C I} \right)^{-1} \left[ \frac{d p R}{d C I} - \frac{d C I_{mm}}{d C I} \right] \end{cases}
\quad (40)\]
In this case, the criterion (8) is always valid because
\[
\frac{L_c(R_M - C_{t,M})}{M_{t,M}} > 0, \quad \frac{M_{t,M}}{A} > 0.
\]

6. (3) General case

We can show eq.(38) always has two real solutions, \( \omega_1 \) and \( \omega_2 \), and the steady state stability is determined from the sign of \( \omega_1 + \omega_2 \):

\[
\frac{d pR}{dCI} - \frac{d CL_{un}}{dCI} < 0, \quad \frac{d CL_{un}}{dCI} > 0 \text{ then stable}
\]

\[
\frac{d pR}{dCI} - \frac{d CL_{un}}{dCI} > 0, \quad \text{then unstable}
\]

but

\[
\frac{d pR}{dCI} - \frac{d CL_{un}}{dCI} < 0, \quad \frac{d CL_{un}}{dCI} < 0 \text{ stable unstable.}
\]

This means that the stability cannot be determined only from the slopes of the production and the degradation curves in the case the degradation curve has a negative slope.
Appendix: The positivity of the discriminant $D$ of eq. (37):

\[
D = \frac{1}{\tau_c^2} \left( -M_{i,C}(R_{i,M} - C_{i,M}) + M_{i,M}(R_{i,C} - C_{i,C}) \right)
- \frac{1}{\tau_M^2} \left( C_{i,M}(L_{i,C} - M_{i,C}) + C_{i,C}M_{i,M} \right)^2
- \frac{4}{\tau_M^2 \tau_c} \left[ M_{i,C}C_{i,M} - C_{i,C}M_{i,M} \right] \left[ (L_{i,C} - M_{i,C})(R_{i,M} - C_{i,M}) + (R_{i,C} - C_{i,C})M_{i,M} \right]
= \left\{ \begin{array}{l}
\frac{1}{\tau_c} \left( -A - R_{i,M}M_{i,C} + R_{i,C}M_{i,M} \right) - \frac{1}{\tau_M} \left( A + L_{i,C}C_{i,M} \right) \right\}^2
+ \frac{4}{\tau_M^2 \tau_c} A \left[ -A + L_{i,C}R_{i,M} - L_{i,C}C_{i,M} - R_{i,M}M_{i,C} - R_{i,C}M_{i,M} \right]
= \left\{ -A \left( \frac{1}{\tau_c} + \frac{1}{\tau_M} \right) + \left( \frac{1}{\tau_c} \left( -R_{i,M}M_{i,C} + R_{i,C}M_{i,M} \right) - \frac{1}{\tau_M} L_{i,C}C_{i,M} \right) \right\}^2
+ \frac{4}{\tau_M^2 \tau_c} A \left[ -A + L_{i,C}R_{i,M} - L_{i,C}C_{i,M} - R_{i,M}M_{i,C} + R_{i,C}M_{i,M} \right]
= A^2 \left[ \frac{1}{\tau_c} - \frac{1}{\tau_M} \right]^2 + 2A \left( \frac{1}{\tau_c} + \frac{1}{\tau_M} \right) \left( \frac{1}{\tau_c} \left( R_{i,M}M_{i,C} - R_{i,C}M_{i,M} \right) + \frac{1}{\tau_M} L_{i,C}C_{i,M} \right)
+ \left( \frac{1}{\tau_c} \left( -R_{i,M}M_{i,C} + R_{i,C}M_{i,M} \right) - \frac{1}{\tau_M} L_{i,C}C_{i,M} \right)^2
\]

The coefficient of $2A$ is

\[
(R_{i,M}M_{i,C} - R_{i,C}M_{i,M}) \left( \frac{1}{\tau_c} - \frac{1}{\tau_M} \right) \left( \frac{1}{\tau_c} \right) + L_{i,C}C_{i,M} \left( \frac{1}{\tau_c} + \frac{1}{\tau_M} \right) \frac{1}{\tau_M} + L_{i,C}R_{i,M} \frac{2}{\tau_M^2 \tau_c}.
\]

Thus, $D$ is
$$D = A \left[ \frac{1}{\tau_c} - \frac{1}{\tau_m} \right]^2$$

$$+ 2A \left[ \frac{1}{\tau_c} - \frac{1}{\tau_m} \right] \left( \frac{1}{\tau_c} (R_M M_{i,C} - R_c M_{i,M}) - \frac{1}{\tau_m} L_c C_{i,M} \right) + \frac{2}{\tau_M} L_c R_M \right]$$

$$+ \left( \frac{1}{\tau_c} - \frac{1}{\tau_m} \right) A + \left( \frac{1}{\tau_c} (R_M M_{i,C} - R_c M_{i,M}) - \frac{1}{\tau_m} L_c C_{i,M} \right)^2$$

$$- \left( \frac{1}{\tau_c} (R_M M_{i,C} - R_c M_{i,M}) - \frac{1}{\tau_m} L_c C_{i,M} \right)^2 + 2A \frac{2}{\tau_m} L_c R_M$$

$$+ \left( \frac{1}{\tau_c} (R_M M_{i,C} - R_c M_{i,M}) + \frac{1}{\tau_m} L_c C_{i,m} \right)^2$$

$$\left[ \frac{1}{\tau_c} - \frac{1}{\tau_m} \right] A + \left( \frac{1}{\tau_c} (R_M M_{i,C} - R_c M_{i,M}) + \frac{1}{\tau_m} L_c C_{i,m} \right)^2$$

$$+ \frac{4}{\tau_m} L_c ((R_M M_{i,C} - R_c M_{i,M}) C_{i,m} + (C_{i,C} M_{i,M} - C_{i,m} M_{i,C}) R_M)$$

The second line of the last expression is shown to be positive as follows:

Note that

$$C_{i,C} = 1 + C_{2,C} + MC_{C}, \quad C_{i,M} = MC_{M}, \quad M_{i,C} = MC_{C}, \quad M_{i,M} = 1 + MC_{C},$$

$$R_c = R' \cdot MC_{C}, \quad R_M = R' \cdot MC_{M},$$

where the abbreviated notations are used:

$$C_{2,C} = \frac{\partial [Cl]}{\partial [Cl]}, \quad MC_{M} = \frac{\partial [MOR \cdot Cl]}{\partial [MOR]}, \text{ etc. and } R' = \frac{\partial pR}{\partial [MOR - Cl]}.$$

Then the second line of $D$ is

$$L_c ((R_M M_{i,C} - R_c M_{i,M}) C_{i,m} + (C_{i,C} M_{i,M} - C_{i,m} M_{i,C}) R_M)$$

$$= L_c (-R_c M_{i,M} C_{i,M} + C_{i,C} M_{i,M} R_M)$$

$$= L_c M_{i,M} (-R_c C_{i,M} + R_M C_{i,C})$$

$$= L_c M_{i,M} (-R' \cdot MC_{C} MC_{M} + R' \cdot MC_{M} (1 + C_{2,C} + MC_{C}))$$

$$= L_c R' \cdot MC_{M} (1 + C_{2,C}) > 0$$

Therefore,

$$D > 0.$$
Appendix B


Key Players in the Genetic Switch of Bacteriophage TP901-1

Anne Alsing,† Margit Pedersen,†* Kim Sneppen,† and Karin Hammer‡
†Center for Models of Life, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark; and ‡Department of Systems Biology, Technical University of Denmark, Kongens Lyngby, Denmark

ABSTRACT After infection of a sensitive host temperate phages may enter either a lytic or a lysogenic pathway leading to new phage assembly or silencing as a prophage, respectively. The decision about which pathway to enter is centered in the genetic switch of the phage. In this work, we explore the bistable genetic switch of bacteriophage TP901-1 through experiments and statistical mechanical modeling. We examine the activity of the lysogenic promoter Pr at different concentrations of the phage repressor, CI, and compare the effect of CI on Pr in the presence or absence of the phage-encoded MOR protein expressed from the lytic promoter Pl. We find that the presence of large amounts of MOR prevents repression of the Pr promoter, verifying that MOR works as an antirepressor. We compare our experimental data with simulations based on previous mathematical formulations of this switch. Good agreement between data and simulations verify the model of CI repression of Pr. By including MOR in the simulations, we are able to discard a model that assumes CI and MOR do not interact before binding together at the DNA to repress Pr. The second model of Pr repression assumes the formation of a CI:MOR complex in the cytoplasm. We suggest that a CI:MOR complex may exist in different forms that either prevent or invoke Pr repression, introducing a new twist on mixed feedback systems.

INTRODUCTION

Epigenetic differentiation is found in a wide range of biological areas. Perhaps most appreciable is the differentiation of human cells, which carry the same genetic material but present truly different phenotypes. Sustained differentiation relies on the existence of multiple stable states. Bistability is the simplest example of this multistability, and is readily seen in investigations of the reciprocal transcription repression exemplified by the genetic switch of Escherichia coli phage λ (Fig. 1A) (1). However, bistability may be acquired by motif designs that are quite different from the classical motif as shown in in silico studies (2,3). One of these is the mixed feedback loop (MFL), in which protein A regulates the expression of protein B and the less-abundant protein is sequestered into an inactive heteromer with the other protein (Fig. 1B). Such sequestration combined with repression invoked by the heteromer has been proposed to be part of the genetic switch mechanisms of the temperate Lactococcal phage TP901-1 (Fig. 1, C and D) (4,5). In line with earlier work, we define the immune state of this genetic switch as when the early lytic promoter Pl is repressed, and accordingly call the state of an open Pl promoter the antiimmune state (Fig. 1, E and F). In the immune state, the phage-encoded repressor, CI, represses transcription from Pr through cooperative binding at the operator sites Or and Ot (6). Further cooperative binding of CI to a distant operator site Oo tightens this repression (4). CI binding at Oo also represses the early lysogenic promoter Pr, though only partly, because Oo is a relatively weak CI-binding site. The binding sequences for CI and the affinities of CI to the DNA were determined by Johansen et al. (6), who reported binding affinities of CI to Ot and Oo of 28 nM and >2000 nM, respectively. Since the sequence of the Oo site is similar to the consensus CI-operator site and since preliminary gel mobility shift assays have shown that CI binds with the same affinity to Oo as to Ot. (M. Pedersen, unpublished data) the binding constant of CI at Oo is considered as equal to CI binding at Ot. (6). MOR is known to function as an antirepressor, and tight repression of Pr in the antiimmune state is only obtained in the presence of both CI and the phage-encoded protein MOR, and is apparently independent of the known CI operator sites (7,4). In vivo studies have shown that MOR alone does not repress any of the two promoters (4). Furthermore, gel mobility shift assays using DNA fragments covering the region from +70 for the Pr promoter to +140 for the Pl promoter and purified MOR protein did not reveal any interaction of MOR with DNA in vitro (data not shown).

The tight repression of Pr in the antiimmune state involves CI, MOR, and supposedly the DNA in the vicinity of the Pr promoter, but this interaction has not been verified experimentally (8). However, all temperate phages that encode MOR homologs also hold a conserved region in CI. This conserved region was proposed to be important for the CI-MOR interaction (9). A previous in silico study confirmed that interactions among CI, MOR, and DNA could play a role in the repression of Pr. However, it was not possible to determine exclusively whether the CI:MOR complex should form directly on the DNA or act partly through sequestration before DNA binding (5).

In this work we consider two genetic systems based on the minimal switch fragment isolated by Pedersen and Hammer (4) with CI introduced in trans from a synthetic promoter.
library (10,11). We find that the presence of MOR allows PR to be active at CI concentrations where PR is otherwise repressed by CI, and deduce that if MOR works through interaction with CI in a heteromer protein complex, this complex must form in the cell cytoplasm before binding at the DNA. For each model, all possible binding conformations between CI and its operator sites are listed. Each of these binding species is assigned a statistical weight that is normalized such that the species of all free operator sites is given a weight of one. The promoter activity probability is the sum of weights of binding species that allow promoter activity (see Table S1). The total concentration of CI and MOR is expressed as the sum of components using equilibrium kinetics.

We based our investigation of the cooperativity in CI binding on the general assumptions that CI binds as a dimer, and that binding of CI at Oo blocks activity of both Pn and Pr, whereas CI binding at Os only blocks activity from Pr. It is assumed that CI binding at Os alone does not influence the promoter activities.

A survey of models of increasing cooperativity between bound CI dimers led us to a model that assigns individual

\[
\Delta G_s = \begin{cases} 0 & \text{Species } i(s) \text{ and } j(s) \text{ refer to the number of bound CI and MOR, respectively.} \\
\sum_s \exp(-\Delta G_s/k_B T) & \text{Speculated motif of the } \text{TP901-1 switch adopted from Pedersen and Hammer (4).} \\
\end{cases}
\]

for a simplified TP901-1 switch by Nakanishi et al. (5), and include CI binding at all known CI-binding sites and levels of cooperativity.

An elaborate description of the models is given in the Supporting Material, and the main equations are replicated in Table 1. Briefly, the activity probability of each promoter is represented as a function of the concentration of free CI. For each model, all possible binding conformations between CI and its operator sites are listed. Each of these binding species is assigned a statistical weight that is normalized such that the species of all free operator sites is given a weight of one. The promoter activity probability is the sum of weights of binding species that allow promoter activity (see Table S1). The total concentration of CI and MOR is expressed as the sum of components using equilibrium kinetics.

We based our investigation of the cooperativity in CI binding on the general assumptions that CI binds as a dimer, and that binding of CI at Os blocks activity of both Pn and Pr, whereas CI binding at Os only blocks activity from Pr. It is assumed that CI binding at Os alone does not influence the promoter activities.

A survey of models of increasing cooperativity between bound CI dimers led us to a model that assigns individual

\[
\Delta G_s = \begin{cases} 0 & \text{Species } i(s) \text{ and } j(s) \text{ refer to the number of bound CI and MOR, respectively.} \\
\sum_s \exp(-\Delta G_s/k_B T) & \text{Speculated motif of the } \text{TP901-1 switch adopted from Pedersen and Hammer (4).} \\
\end{cases}
\]

for a simplified TP901-1 switch by Nakanishi et al. (5), and include CI binding at all known CI-binding sites and levels of cooperativity.

An elaborate description of the models is given in the Supporting Material, and the main equations are replicated in Table 1. Briefly, the activity probability of each promoter is represented as a function of the concentration of free CI. For each model, all possible binding conformations between CI and its operator sites are listed. Each of these binding species is assigned a statistical weight that is normalized such that the species of all free operator sites is given a weight of one. The promoter activity probability is the sum of weights of binding species that allow promoter activity (see Table S1). The total concentration of CI and MOR is expressed as the sum of components using equilibrium kinetics.

We based our investigation of the cooperativity in CI binding on the general assumptions that CI binds as a dimer, and that binding of CI at Os blocks activity of both Pn and Pr, whereas CI binding at Os only blocks activity from Pr. It is assumed that CI binding at Os alone does not influence the promoter activities.

A survey of models of increasing cooperativity between bound CI dimers led us to a model that assigns individual

\[
\Delta G_s = \begin{cases} 0 & \text{Species } i(s) \text{ and } j(s) \text{ refer to the number of bound CI and MOR, respectively.} \\
\sum_s \exp(-\Delta G_s/k_B T) & \text{Speculated motif of the } \text{TP901-1 switch adopted from Pedersen and Hammer (4).} \\
\end{cases}
\]

for a simplified TP901-1 switch by Nakanishi et al. (5), and include CI binding at all known CI-binding sites and levels of cooperativity.

An elaborate description of the models is given in the Supporting Material, and the main equations are replicated in Table 1. Briefly, the activity probability of each promoter is represented as a function of the concentration of free CI. For each model, all possible binding conformations between CI and its operator sites are listed. Each of these binding species is assigned a statistical weight that is normalized such that the species of all free operator sites is given a weight of one. The promoter activity probability is the sum of weights of binding species that allow promoter activity (see Table S1). The total concentration of CI and MOR is expressed as the sum of components using equilibrium kinetics.

We based our investigation of the cooperativity in CI binding on the general assumptions that CI binds as a dimer, and that binding of CI at Os blocks activity of both Pn and Pr, whereas CI binding at Os only blocks activity from Pr. It is assumed that CI binding at Os alone does not influence the promoter activities.

A survey of models of increasing cooperativity between bound CI dimers led us to a model that assigns individual

\[
\Delta G_s = \begin{cases} 0 & \text{Species } i(s) \text{ and } j(s) \text{ refer to the number of bound CI and MOR, respectively.} \\
\sum_s \exp(-\Delta G_s/k_B T) & \text{Speculated motif of the } \text{TP901-1 switch adopted from Pedersen and Hammer (4).} \\
\end{cases}
\]

for a simplified TP901-1 switch by Nakanishi et al. (5), and include CI binding at all known CI-binding sites and levels of cooperativity.

An elaborate description of the models is given in the Supporting Material, and the main equations are replicated in Table 1. Briefly, the activity probability of each promoter is represented as a function of the concentration of free CI. For each model, all possible binding conformations between CI and its operator sites are listed. Each of these binding species is assigned a statistical weight that is normalized such that the species of all free operator sites is given a weight of one. The promoter activity probability is the sum of weights of binding species that allow promoter activity (see Table S1). The total concentration of CI and MOR is expressed as the sum of components using equilibrium kinetics.

We based our investigation of the cooperativity in CI binding on the general assumptions that CI binds as a dimer, and that binding of CI at Os blocks activity of both Pn and Pr, whereas CI binding at Os only blocks activity from Pr. It is assumed that CI binding at Os alone does not influence the promoter activities.

A survey of models of increasing cooperativity between bound CI dimers led us to a model that assigns individual

\[
\Delta G_s = \begin{cases} 0 & \text{Species } i(s) \text{ and } j(s) \text{ refer to the number of bound CI and MOR, respectively.} \\
\sum_s \exp(-\Delta G_s/k_B T) & \text{Speculated motif of the } \text{TP901-1 switch adopted from Pedersen and Hammer (4).} \\
\end{cases}
\]
energies to pair- and triplewise cooperative binding. The model ignores the energy associated with the DNA loop of Or-Or or Ol-Or cooperative binding, and was chosen to keep the number of model parameters limited. It is justified because, to date, no experimental determinations of binding energies allow us to make distinctions between pairwise cooperativity with or without loop formation. The 14 species included in the model are listed in Table S1.

To include MOR in the repression of Pr, we focused on two main models (5): model A, in which the CI:MOR complex associates at the DNA, and model B, in which the CI:MOR complex forms before binding at the DNA. In both models, it is assumed that a complex that is bound at a putative Om site prevents activity from the Pr promoter without disturbing the activity from Pr. This should be consistent with the experimental indications that MOR abolishes CI repression of Pr while interfering with a tighter repression of Pr in concert with CI (4). In addition, we assume that complex binding does not hinder CI dimer binding at its known operator sites and vice versa. Including MOR in the model thus effectively doubles the number of states to 28.

MATERIALS AND METHODS

Bacterial strains, transformation, and growth conditions

Table 2 lists the bacterial strains used in this work. E. coli was propagated in Luria-Bertani medium at 37°C with agitation (14). In solid medium, 1.5% (wt/Vol) Bacto agar was used. When appropriate, the medium was supplemented with Ampicillin (100 μg/mL) or Tetracycline (4 μg/mL). Plasmid DNA was introduced into E. coli by electroporation as described previously (8). Lactococcus lactis subsp. cremoris MG1363 was grown at 30°C in GM17 medium (M17 broth supplied with 0.5% glucose) containing Erythromycin (5 μg/mL), Chloramphenicol (5 μg/mL), or Tetracycline (2 μg/mL) when appropriate. Lactococcal cells were made electrocompetent and transformed by electroporation as previously described (15). Screening on 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) plates was performed at an X-gal concentration of 90 μg/mL.

DNA techniques

Plasmid isolation from E. coli was done using Qiaprep Spin Miniprep Kit 250 (Qiagen, Duesseldorf, Germany). Conventional restriction enzymes and T4 DNA ligase were obtained from Fermentas Life Sciences (St. Leon-Rot, Germany) and used according to protocol. A GFX PCR II lustra DNA and GelBand Purification Kit (GE Healthcare, Brondby, Denmark) was used for purification of DNA fragments. PCR amplifications were performed for verification of inserts and products intended for cloning with Taq or ExTaq polymerase, respectively. Both polymerses and buffers were acquired from New England BioLabs (Ipswich, MA) and used as described by the supplier.

Construction of plasmids and strains

The plasmids and primers used in this study are listed in Table 2. Plasmid pAKA5 was constructed in E. coli by ligation of a purified 0.8 kb HindIII-PstI fragment obtained from plasmid pPV40 (provided by P. Varmanen, Faculty of Veterinary Medicine, University of Helsinki) and a purified 6.6 kb fragment obtained by digestion of pCS374 (16) with HindIII and PstI. The orientation of pepI was verified by sequencing.

Plasmid pMAP86 was constructed as follows: A DNA fragment containing cIOr "mor1Or" was constructed by ligation of two Xbd digested DNA fragments obtained by PCR amplification using plasmid pMAP53 (cI Or "mor 1Or") (4) as template DNA and primers Clmut for with pAK80term and primer Clmutrev with pAK80, respectively. The primers introduced a Xbd restriction site and a frame-shift mutation at the beginning of cI, similar to the cl mutation (cI1) introduced in pPM131 (7). The DNA fragment was cloned using the TOPO TA cloning kit from Invitrogen (San Diego, CA). The insert was transferred as a HindIII fragment to the promoter probe plasmid pAK80, and the orientation of the insert was verified by sequencing.

Strain LB436 (17) was transformed with plasmid pMAP86(mor1, cI1, Pc-lacLM), pMAP88(mor1, cI1, Pc-lacLM), or pMAP87(mor1, cI1, Pc-lacLM) (4) to construct strain LAKA2, LAKA8, or LAKA9, respectively. Strains SPLAKA1, SPLAKA3, and SPLAKA4, which contain synthetic promoter libraries (P promin) of cI-pepI, were constructed using the Solem-Jensen approach (11). PCR products containing cl transcribed by synthetic promoters were constructed using plasmid MAP80 (8) as template, primer cIPst.rev, and one of the synthetic promoter primers. Upon digestion of the PCR products and plasmid pAK5 by PstI and XbaI, a ligation mixture of the two was introduced into each of the three strains (LAKA2, LAKA8, and LAKA9). Transformants were chosen based on a decrease in the blue color of the colonies when grown on X-Gal-supplemented agar as compared with the parent strains. Integration of a synthetic promoter cI fragment upstream of the pepI reporter gene on the chromosome was subsequently verified by altered PepI activity compared with strains transformed with pAK5 containing the promoter less pepI.

Enzyme activity assays

Cell extracts for the enzyme assays were prepared in the following way: L. lactis cells were grown in 100 mL GM17 medium with appropriate antibiotics at 30°C to an optical density of OD600 = 0.5. Then 90 mL of each sample were harvested, washed once in 4 mL 50 mM HEPES buffer pH 7, and resuspended in 980 μL of the same buffer, all performed on ice. The samples were subsequently stored in stable volumes at −80°C. On the day of the assay, the samples were thawed on ice and the cells were ruptured using a FastPrep bead beater. Cell debris and FastPrep beads were removed by centrifugation and the resulting supernatant was used as the protein extract in the assays.

Peptidase activity assays were performed in a manner similar to that described by Varmanen et al. (18) on 1.5–5 mg total protein, in 50 mM HEPES buffer, pH 7, using 10 μL L-Proline-para-nitroanilide (10 mg/mL) as substrate in a final assay volume of 1 mL. The release of nitroanilide was followed at 410 nm (Zeiss thermostat spectrophotometer, Zeiss, Jena, Germany) at 37°C.

The β-galactosidase activity was assayed as described previously (19) with some modifications. Assays were performed at 30°C on 0.3–4 mg total protein diluted in Z-buffer to a total volume of 650 μL. After the samples were preheated, the assay was initiated by addition of 100 μL Ortho-Nitrophenyl-β-galactoside (4 mg/mL) and stopped by addition of 250 μL of Na2CO3 when the sample had turned slightly yellow, resulting in a final assay volume of 1 mL. The release of o-nitrophenol was subsequently determined as the change in optical density at 422 nm between the assay sample and a similar sample in which no reaction of β-galactosidase had been allowed.

Using the molar extinction coefficients of 8800 M−1 cm−1 and 4500 M−1 cm−1 for nitroanilide at 410 nm (19) and o-nitrophenol at 422 nm (18), respectively, the specific enzyme activity was calculated in units of nanomole substrate processed per minute per milliliter of cell extract per 280 nm optical density.
Western blotting

The protein concentration was determined by measuring OD$_{280}$ (1 OD$_{280}$ = 1 mg/mL). Generally, samples containing 200 µg protein were prepared for analysis by sodium dodecyl sulfate polyacrylamide gel electrophoresis using 15% gels. After separation, the proteins were transferred onto Hybond P (PVDF) membranes (GE Healthcare Life Sciences) by electroblotting with a semidry blotter (BioRad, Hercules, CA). Western blotting was performed as described previously (4).

Fitting models to data

We quantified the goodness of fit between the models and data by calculating an unweighted sum of squared residues (SSR); $10^6$ parameter sets were tested in each fitting procedure. Parameters were randomly chosen such that the basal activity of the P$_{S}$ promoter, $pR_S$, lay between LacLM activities of 23 and 30, in similarity to the basal level in the data sets.

The six energy parameters were taken between −15 and 0 kcal/mol corresponding to binding constants between $10^{-11}$ and $10^{-1}$ M. In the presence of MOR, we fixed the parameter sets found in the model describing MOR-independent repression of P$_S$ and scanned the two additional binding energies.

The CI concentration in cells containing the wild-type switch in the immune state was estimated to $10^{-6}$ M corresponding to 1 wild-type immune-state unit (WIU). This is based on a preliminary rough estimate of TF901-1 CI concentration in vivo and is comparable to previous estimates based on lysogenic CI concentrations in phase λ (5). In the simulations that did not include MOR, the estimated relation between WIU and concentration set a reference for all of the binding energies. Changing this relation would rescale all binding energies proportionally without changing any other results. The concentration of MOR is upheld by the relation would rescale all binding energies proportionally without changing any other results. The concentration of MOR is upheld by the

<table>
<thead>
<tr>
<th>Strain, plasmid, or primer</th>
<th>Description or sequence 5'-3'</th>
<th>Reference or construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. lactis subsp. cremoris</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MG1363</td>
<td>Plasmid-free derivative of NCDOO712</td>
<td>(28)</td>
</tr>
<tr>
<td>LB436</td>
<td>MG1363/pLB65</td>
<td>(17)</td>
</tr>
<tr>
<td>MP401</td>
<td>MG1363/pMAP50</td>
<td>(8)</td>
</tr>
<tr>
<td>LAKA2</td>
<td>LB436/pMAP86</td>
<td>This study</td>
</tr>
<tr>
<td>LAKA8</td>
<td>LB436/pMAP88</td>
<td>This study</td>
</tr>
<tr>
<td>LAKA9</td>
<td>LB436/pMAP87</td>
<td>This study</td>
</tr>
<tr>
<td>SPLAKA1</td>
<td>LAKA2/pAKA5::P$_{R}$-cl-pepI integrated in attB</td>
<td>This study</td>
</tr>
<tr>
<td>SPLAKA3</td>
<td>LAKA8/pAKA5::P$_{R}$-cl-pepI integrated in attB</td>
<td>This study</td>
</tr>
<tr>
<td>SPLAKA4</td>
<td>LAKA9/pAKA5::P$_{R}$-cl-pepI integrated in attB</td>
<td>This study</td>
</tr>
<tr>
<td>E. coli</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH5α</td>
<td>ϕ80lacZ Δ(lacZYA-arg F)U169 rec A1</td>
<td>Stratagene</td>
</tr>
<tr>
<td>Plasmids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pLB65</td>
<td>pCI372::1.9-kb Eco RI-Sal I pLB61 orf 1Prom1-1, oriE.coli, Cam$R$</td>
<td>(17)</td>
</tr>
<tr>
<td>pPV40</td>
<td>pLB85 ΔgusA-pepI, Erm$R$ oriE.coli</td>
<td>!</td>
</tr>
<tr>
<td>pAK80</td>
<td>Promoter probe vector containing lacLM, Erm$R$</td>
<td>(29)</td>
</tr>
<tr>
<td>pMAP50</td>
<td>pAK80::cl+ mor+ P$_L$ -lacLM, oriL.lactis, Erm$R$</td>
<td>(8)</td>
</tr>
<tr>
<td>pMAP53</td>
<td>pAK80::cl -mor1 P$_L$ -lacLM, oriL.lactis, Erm$R$</td>
<td>(4)</td>
</tr>
<tr>
<td>pMAP86</td>
<td>pAK80::cl 1 mor1 P$_L$ -lacLM, oriL.lactis, Erm$R$</td>
<td>This study</td>
</tr>
<tr>
<td>pMAP87</td>
<td>pAK80::cl 1 mor- P$_L$ -lacLM, oriL.lactis, Erm$R$</td>
<td>(4)</td>
</tr>
<tr>
<td>pMAP88</td>
<td>pAK80::cl 1 mor- PR -lacLM, oriL.lactis, Erm$R$</td>
<td>(4)</td>
</tr>
<tr>
<td>pCSS74</td>
<td>pLB85 derivative, gusA, attP,P901-1, oriE.coli, Bla$R$, Tel$R$</td>
<td>(16)</td>
</tr>
<tr>
<td>pAKA5</td>
<td>pCSS74 ΔgusA-P$_{R}$, attP,P901-1, Tel$R$</td>
<td>This study</td>
</tr>
<tr>
<td>Prima1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI promoters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAKA1.sp</td>
<td>GGGGGGCTATGAAGCTT NNNNAGTTAT TCTTGCANN</td>
<td>CI library</td>
</tr>
<tr>
<td>prAKA1.sp</td>
<td>GGGGGGCTATGAAGCTT NNNNAGTTAT TCTTGCANN</td>
<td>CI library</td>
</tr>
<tr>
<td>prAKA4.sp</td>
<td>GGGGGGCTATGAAGCTT NNNNAGTTAT TCTTGCANN</td>
<td>CI library</td>
</tr>
<tr>
<td>Clmut.rev</td>
<td>GGGGGGCTATGAAGCTT NNNNAGTTAT TCTTGCANN</td>
<td>CI library</td>
</tr>
<tr>
<td>pAK80erm</td>
<td>GGGGGGCTATGAAGCTT NNNNAGTTAT TCTTGCANN</td>
<td>pMAP86</td>
</tr>
<tr>
<td>CImut.for</td>
<td>GGGGGGCTATGAAGCTT NNNNAGTTAT TCTTGCANN</td>
<td>pMAP86</td>
</tr>
<tr>
<td>pAK80</td>
<td>GGGGGGCTATGAAGCTT NNNNAGTTAT TCTTGCANN</td>
<td>pMAP86</td>
</tr>
</tbody>
</table>

*Library of synthetic promoters.

1Supplied by P. Varmanen.

2Enzyme restriction sites are underlined. N is any of the four bases, and R is 50% A and 50% G.
RESULTS

Pr is repressed 25-fold by a 600-fold increase in CI

A 979 bp DNA fragment containing the divergently oriented early promoters (Pr and Pm), the three known CI operator sites, and the genes mor and cI (both of which are required for bistability) was isolated from the genome of phage TP901-1 and shown to be bistable when cloned in a low-copy-number plasmid and introduced in a compatible host (8).

To measure the activity of the Pr promoter at different CI concentrations in the absence of MOR, we used plasmid pMAP86, which holds the genetic switch fragment with frame-shift mutations in cI and mor and a LacLM reporter gene downstream of the Pr promoter (Fig. 2, A, left). The bare activity of Pr in the absence of any CI protein was measured as 30 ± 7 nmol processed Ortho-Nitrophenol-\(\beta\)-galactoside per minute per milliliter of cell extract per OD\(_{280}\). We then introduced CI in trans using a library of synthetic constitutive promoters integrated on the bacterial chromosome (11,20). To monitor the level of CI, we placed the reporter gene pepI downstream of cI. Each strain thus allowed us to determine the Pr activity reported by LacLM at a constitutive level of CI reported by PepI. The promoter activity given by the LacLM activity is plotted against the CI expression determined by the PepI activity in Fig. 2 A. We observe a 25-fold decrease in Pr activity over an ~600-fold change in CI expression. The decrease in activity apparently contains a plateau stretching over a 10-fold increase in CI concentration where the Pr activity is held at approximately half of its full activity.

MOR acts as an antirepressor of Pr

To measure the activity of Pr in the presence of MOR at various CI concentrations, we constructed a number of strains that contained plasmid pMAP88 holding a frame-shift mutation only in cI and with Pr activity reported by lacLM (8). Again CI was introduced from a single copy on the chromosome and monitored by the reporter PepI (Fig. 2 B, left).

The resulting data points of LacLM activity (Pr activity) versus PepI activity (CI concentration) are shown in Fig. 2 B. Pr is repressed ~500-fold as CI concentration increases 600-fold, but the repression is not observed until the last 10-fold increase in CI expression occurs. Thus, the presence of MOR renders the Pr promoter open, even at the CI concentration at which the promoter was partially repressed in the absence of MOR (Fig. 2, A and B).

In addition to the studies on Pr promoter activity, we performed a smaller sampling to examine the response of the Pr promoter to different levels of CI in the presence of MOR. Plotting the Pr activity (LacLM activity) against the CI expression (PepI Activity) confirms that mor is indeed transcribed in the strains for CI expression beyond levels at which the Pr promoter is closed (Fig. 2 C, and insets in A and B).

To verify a linear correlation between the PepI activity level and CI concentration in our systems, we performed Western blotting on a selection of the assay samples (Fig. 3). The plot shows the PepI activity versus CI concentration in units of the CI concentration in a system containing a wild-type switch in the immune state (i.e., the WIU). The linear relation is clear from the least-square fit and was used to rescale the experimental PepI activity units to CI concentrations before model fitting was conducted.

Models predict the relative weakness of CI binding at Or

To investigate the CI-invoked Pr repression, we used the model presented in Table 1 to fit the data set on Pr response to increasing CI concentrations (Fig. 2 A). For each parameter set, we calculated an unweighted SSR.
The experimental dissociation constants of CI at O\textsubscript{L} and O\textsubscript{R} include both dimerization and DNA binding (6). However, they relate the binding energy parameters that describe CI dimer binding to O\textsubscript{L} and O\textsubscript{R}. Hence, we initially restricted the parameters such that $\Delta G_i \sim \Delta G_i + \ln(2000/28) k_B T$. Additionally, $\Delta G_i \sim \Delta G_L$ due to the conserved sequence of O\textsubscript{R} to the consensus sequence of O\textsubscript{L} (6). The best fit found with these constraints on the model parameters is plotted together with the experimental data in Fig. 4A (dashed curve, SSR $\sim 130$).

An additional fitting procedure was set up without any constraints on the relations between parameters. From this fit, all simulations of SSR < 100 reproduced the plateau indicated in the data and were subjected to further investigation. The best-fitting repression curve generated by the unconstrained fitting is plotted in Fig. 4A (solid line, SSR $\sim 60$).

Both of the simulations presented in Fig. 4A show an acceptable fit to the data. The hypothetical thermodynamical parameters predicted by the simulations are linked to the lysogenic CI level, which we assume to be $10^{-6} \text{ M}$. This concentration would be matched by a CI-binding energy of order $-k_B T \ln 10^{-6} = \sim 8 \text{ kcal/mol}$. Binding energies that are weaker than this ($\Delta G > \sim 8 \text{ kcal/mol}$) imply that the corresponding binding rarely occurs at standard lysogenic concentrations.

In our unconstrained simulation, the obtained CI dimerization energy is $\Delta G_{\text{DIM}} = \sim 3.6 \text{ kcal/mol}$, implying that there are almost no free dimers in a lysogen. Thus, dimers only form on the DNA. The obtained weak dimerization therefore reflects a cooperative effect associated with CI dimerization on the operators.

Our constrained simulation represents the opposite scenario, where the CI dimerization is so strong that CI instantly forms dimers. In this case, $\Delta G_i$ (and $\Delta G_{\text{DIM}}$) of the simulation may be related to the measured dissociation constant for CI on O\textsubscript{R}, $K_D = 2 \exp(\Delta G_i/k_B T) \sim 2600 \text{ nM}$ (where the factor 2 is associated with the strong dimerization limit; for further details, see Supporting Material). In the constrained simulation, the cooperativity is reflected in weaker CI dimer binding to single operators aided by $\Delta G_{\text{DIM}} \sim \sim \sim 3 \text{ kcal/mol}$ (compare $\Delta G_i$, $\Delta G_L$, between unconstrained and constrained parameters in Fig. 4, D and E).

In the remainder of this text, we focus on the results of the unconstrained simulations because they provide some predictions that validate the model. All of the implications of the model when MOR is included later are similar for the two simulations (see Fig. S2 and Fig. S3).

From the unconstrained fitting, all parameter sets subjected to further investigation show the same pattern of binding species. Eleven of the 14 binding species in the model (Fig. 4B) constitute the repression of P\textsubscript{R} and invoke their individual effects over three intervals. At low CI concentrations, the promoter is open because the species of no bound CI dominates. With increasing concentration, the species of two cooperatively bound CI dimers at O\textsubscript{R}-O\textsubscript{L} or O\textsubscript{R}-O\textsubscript{O} dominate together with the species of two dimers cooperatively bound at O\textsubscript{R}-O\textsubscript{O} (species C, D, and E in Fig. 4C). Together, these binding species constitute the approximately half-repressed level of P\textsubscript{R} because only species E allows activity from P\textsubscript{R}. At high CI concentrations, all of the operator sites are covered by bound CI dimers in either pair- or triplewise cooperative bindings, resulting in full repression of P\textsubscript{R}.

Whether species C or D will balance species E depends on the relation between the probabilities of the three states. This relation is effectively given by the assigned binding strengths of the three sites (O\textsubscript{R}, O\textsubscript{L}, and O\textsubscript{O}). We find that the two CI operator sites assigned the weakest bindings are always similar in binding energy, and O\textsubscript{R} is always found to be one of the weaker sites. Thus, even without restricting the CI dissociation constants to O\textsubscript{R} and O\textsubscript{L}, the model predicts a relative weakness of CI binding to the O\textsubscript{R} site.

Simulations discard the model of direct binding of CI and MOR at DNA

Initially, we simulated the repression of P\textsubscript{R} in the anti-immune state by direct binding of CI and MOR at a speculated operator site, O\textsubscript{M}. This model does not allow CI and MOR to interact before DNA binding occurs. We find that such
a model cannot concurrently fit the two data sets in Fig. 2, A and B. The dashed gray curve in Fig. 5A shows the best fit of this model to the experimental data. For all CI concentrations, the curve falls at or below the $P_R$ activity curve predicted by the model of $P_R$ repression by CI dimers alone (solid gray line). (Refer to Fig. S2 for examples of simulations with direct binding of CI and MOR to DNA.)

Therefore, we instead focus on the model that includes sequestration of CI into a complex with MOR before DNA binding. An example of $P_R$ repression simulated by this model is shown in Fig. 5A (black curve).

The sequestration of CI into a complex with MOR introduces a hysteresis effect into the system. High initial concentrations of MOR compared with CI leads to full initial sequestration of CI, leaving only the complex to repress $P_R$. In contrast, at low initial MOR concentrations, the CI forms homodimers, leading first to repression of the $P_L$ promoter and a further decrease in MOR.

**FIGURE 4** Model of $P_R$ repression by CI. (A) Simulation of $P_R$ activity (constrained simulation in dashed line, unconstrained simulation in solid line) as a function of total CI concentration together with experimental data (solid circles). Repressor concentration is rescaled to units of immune state CI level (WIU) according to Fig. 3. (B) Species fractions constituting the $P_R$ repression curve of the unconstrained simulation in A as a function of total CI concentration. Species represented by dashed curves allow activity from $P_R$, whereas the binding species indicated by black curves exclude promoter activity. The gray line indicates the $P_R$ activity probability. Species F1–3 are assigned the same total free energy in the model and thus have identical species fraction curves (summed here for clarity). Refer to Fig. S1 for a species fraction plot for the constrained fit. (C) Illustrations of the binding species corresponding to the curves in B. Each black figure represents a CI dimer. (D) Parameters of the constrained simulation: $[CI]_{total} = 10^{-6}$ M at full $P_R$ activity. The 50/50 plateau in the $P_R$ repression is a result of a balance between the weights of species C and E. For other parameter sets, species D takes the place of species C in the balance. In a similar way, species G and F1–3 may contribute together or solely at high CI total concentrations to invoke the full repression of $P_R$. Species B1–3 may contribute at low CI concentrations for other sets of parameters. (E) Parameters of the unconstrained simulation: $[CI]_{total} = 10^{-6}$ M at full $P_R$ activity.

**FIGURE 5** Model of $P_R$ repression in systems containing both CI and MOR. (A) Simulated $P_R$ activity by the model with CI sequestration for systems with high (black curve) or low (dashed black curve) initial concentrations of MOR. Solid circles represent experimental data from Fig. 2B rescaled according to Fig. 3. The gray curve is the unconstrained simulation of $P_R$ activity in the absence of MOR, equivalent to the curve in Fig. 4A. The dashed gray curve is the simulated $P_R$ activity by the model without CI sequestration. (B) Relation between concentration of free CI and total CI, clearly showing the hysteresis effect. High initial concentrations of MOR lead to CI sequestration and thus low free CI concentrations (solid black curve). In contrast, at low initial MOR concentrations, free CI is only depleted by dimer formation, resulting in high levels of free CI (dashed black curve). The simulation is based on the parameters of the unconstrained simulation given in Fig. 4D, with $\Delta G_{M} = -10.5$ kcal/mol, $\Delta G_{CI} = -6.4$ kcal/mol, and $[MOR]_{total} = 10^{-5}$ M at full $P_R$ activity.
concentration. Fig. 5B illustrates the hysteresis in the relation between free CI and total CI concentration. Note that this hysteresis effect is similar in nature to the characteristic feature of bistability predicted for plots of Pr activity versus uncontrolled Pr activity by Nakanishi et al. (5).

In the experimental strains, the steady state is established from a state of fully active Pr, promoter and consequently a high MOR expression level. The unweighted SSR was thus calculated for the fully drawn branch of the Pr curve in Fig. 5A. The best scoring parameter sets all presented similar Pr activity trajectories (SSR = 1580). The energy of CI-MOR binding is $-10.6$ kcal/mol for the simulation in Fig. 5A. This corresponds to a dissociation constant of $\sim 20$ nM, implying that CI is sequestered in the complex at concentrations 50-fold lower than the assumed lysogenic CI concentration. In contrast, the binding energy of the complex to DNA is only $-6.4$ kcal/mol in the simulation, indicating that this binding could be omitted from the model. Instead, the simulations assign the observed Pr repression to CI dimer binding in both the absence and presence of MOR. However, the experimental data show a 20-fold stronger Pr repression level in the presence of MOR than in its absence (see insets in Fig. 2, A and B). This inconsistency between the simulations and the data reveals the incompleteness of the biological model.

**DISCUSSION**

Experimental data show that upon a 600-fold increase in CI repressor, the lysogenic promoter Pr is repressed 25-fold in the absence of MOR, compared with a 500-fold repression of Pr in the presence of MOR. In the repression of Pr by CI in the absence of MOR, a stepwise manner is indicated, with a plateau held at approximately half of the full Pr activity upon a 10-fold increase in CI concentration.

To simulate the repression profile of Pr, we use a mathematical model that includes pair- and triplewise cooperative binding of CI dimers. In an unconstrained fitting between the model and the data, the model predicts that two of the known CI-binding sites should bind CI more weakly than the third, and that the Oo site will always be one of the weaker sites. This is consistent with the binding-strength determinations of Johansen et al. (6), and strengthens our acceptance of the model. The plateau in these simulations is realized by competition for occupation of the two weaker sites, both supported by cooperative binding to the stronger site. At CI concentrations just below 1 WIU, this competition would allow for some activity from Pr while keeping Pr repressed. We thus speculate that the Oo site not only stabilizes the fully repressed state but also secures a controlled replenishing of CI.

In the presence of MOR, Pr is held open over a 200-fold range of CI concentrations before being repressed. This shows that binding of CI:MOR at DNA is not the sole action of MOR, and that MOR evidently functions by inhibiting CI binding at DNA. Somewhat similar mechanisms were reported for antirepressors of other bacteriophages, including the Tum protein of phage 186 (21), and Coi of phage P1 (22).

The two models of Pr repression in the presence of both CI and MOR considered here originate from the models presented by Nakanishi et al. (5), who verified that sequestration of CI into a CI:MOR complex could be part of the switch mechanism. However, although the other model, in which CI and MOR bind together directly at the DNA, was proved to be theoretically unlikely, it could not be discarded. We compared the results from simulations by the two models with our experimental data and found that the model that does not allow interaction between CI and MOR before DNA binding cannot concurrently fit the two data sets on Pr activity. This allows us to conclusively discard the model of direct binding.

The two-state behavior observed in our data, with high and low activity of Pr, respectively, does not correspond to the known immune and antiimmune states, as high Pr activity is associated with a high level of MOR. In contrast, TP901-1 is known to have an antiimmune state characterized by substantial repression of Pr and full activity of Pr (7). The model of CI sequestration may fit the data but accordingly fails to identify a classic antiimmune state. Indeed, one can omit repression of Pr by the CI:MOR complex from the model without significantly changing the fit. The model assigns the observed repression to CI dimer binding in both the absence and presence of MOR. However, our experiments revealed a 20-fold stronger repression of Pr at high CI concentrations in the presence of MOR than in its absence. This shows that the same mechanism cannot account for the repression in the two situations. Previous studies showed that repression of Pr by the CI:MOR complex was needed to reproduce the known difference in Pr activity between the immune and antiimmune states (5).

This inconsistency between the biological model and the data became apparent in the mathematical simulations, and shows that a revised model of the switch mechanisms is needed. Inspired by the RelB-RelE system of *E. coli*, we propose that a large, inactive CI:MOR complex may form at unnaturally high MOR/CI ratios (23,24). The inability of such a large complex to repress Pr could explain the absence of the antiimmune state in our constructs, based on the high level of MOR present in the strains when CI is introduced. It also offers an explanation for previous findings indicating that MOR in large excess inhibits the formation of plaques (7). In contrast, lower MOR/CI ratios may lead to an active complex that represses Pr.

As with phage λ (25–27), the illusive nature of antiimmune repressors makes it difficult to determine their full range of capabilities. However, our data should establish a reliable model for CI repression of Pr, because immunity against superinfection requires that CI alone should be able to act in trans.
SUPPORTING MATERIAL
Two tables, three figures, additional equations, and references are available at http://www.biophysj.org/biophysj-supplemental/S0006-3495(10)05191-X.

We thank P. Varmanen (Faculty of Veterinary Medicine, University of Helsinki) for supplying the plasmid pPV40.

This work was supported by the Danish National Research Foundation through the Center for Models of Life.

REFERENCES
Supporting material

Biophysical Journal.
The Key Players of the Genetic Switch of Bacteriophage TP901-1
Anne Alsing,† Margit Pedersen,‡ Kim Sneppen,† and Karin Hammer†

Document S1 - Models

Our present switch systems include all of the three known operator sites of
the TP901-1 switch and presumably the proposed binding site responsible
for the MOR dependent repression of $P_R$ (1). The MOR independent re-
pression of the promoters in the immune state is treated first followed by
an exploration of the model on MOR dependent repression advertised by
Nakanishi et al (2).

To compare the promoter activities predicted by the models with the
experimental data rescaling is required. The experimental data was thus
rescaled in units of the CI concentration in a system containing plasmid
pMAP50 carrying a wild type switch in the immune state (wild type immune
state units; WIU) while the predicted activity probabilities are multiplied
by the basal promoter activity level in units of LacLM activity.

The CI mediated repression of $P_L$ in the immune state, and the partly
repression of $P_R$ associated with CI binding at $O_R$, is relatively well under-
stood. To probe the level of cooperativity in CI binding, we considered four
models of increasing complexity; 1) The minimal model where no coopera-
tivity is included; $P_R$ is repressed when a CI dimer binds at $O_R$. 2) The model
of single cooperative energy, $\Delta G_{\text{triplet}}$; assigned when CI dimers are bound at
all of the three operator sites corresponding to hexamer formation. 3) The
model of pair- and triple-wise cooperative binding of CI dimers, with the
energy $\Delta G_{\text{pair}}$ and $\Delta G_{\text{triplet}}$, respectively. 4) Full model including individual
energies for each pair- and triple-wise cooperative binding of CI dimers (i.e.,
$\Delta G_{RL}$, $\Delta G_{RD}$, $\Delta G_{LD}$ for pairs of CI dimers bound at $O_R$ and $O_L$, $O_R$ and
$O_O$, $O_L$ and $O_D$, respectively, and $\Delta G_{\text{triplet}}$ for triplet of CI dimers bound at
$O_R$, $O_L$, $O_D$).

†Center for Models of Life, Niels Bohr Institute, University of Copenhagen, Copen-
hagen, Denmark
*Corresponding author. Address: Center for Models of Life, Niels Bohr Institute,
Blegdamsvej 17, 2100 Copenhagen, Denmark. Tel.: +45 353 25373, Fax: +45 353 25016,
E-mail: margit.pedersen@nbi.dk
‡Department of Systems Biology, Technical University of Denmark, Kongens Lyngby,
Denmark
Modeling the phage TP901-1 switch

The microscopic configurations and associated energy contributions of CI repressor bound to the operators O_R, O_L and O_D in the four models are listed in Table S1. The probability of a configuration species \(s\), with \(i_s\) CI monomers bound is given as

\[
f_s = \frac{[CI]^{i_s}e^{-\Delta G_s/k_B T}}{\sum_s [CI]^{i_s}e^{-\Delta G_s/k_B T}},
\]

(S1)

following the notation of (3, 4).

In accordance with the known features of the TP901-1 switch, we let binding of a CI dimer at O_R exclude RNA polymerase binding at both promoters, while CI dimer binding at O_L will only exclude RNA polymerase from binding at the P_L promoter (5, 6). CI binding at O_D alone is assumed not to influence the transcription from either of the promoters. In the models where cooperativity is included binding at O_D may tighten binding at O_R or O_L through the extra energy associated with the cooperativity (6). No activation of the promoters are observed and the probability of P_R promoter activity is approximated as:

\[
p_R([CI]) = p_{R0} \times (f_0 + f_2 + f_3 + f_6 + f_{10}).
\]

(S2)

Where \(p_{R0}\) is the basal promoter activity of the promoter. The sum resembles the probability of finding the promoter unoccupied by CI repressor and amounts to 1 for \([CI]=0\). For discussion of possible mistakes due to dynamics of RNA polymerase initiation see (7).

Eq. S2 was fitted to the experimental data using Eq. S1 for each state probability, \(f_s\), with the individual \(\Delta G_s\) found in Table S1 for each model. Note that the probability is a function of the free CI concentration, but plotted against the total CI concentration given as:

\[
[CI]_{\text{total}} = [CI] + 2 \times [CI_2] = [CI] + 2 \times \frac{[CI]^2}{K_{\text{dim}}},
\]

(S3)

With \(K_{\text{dim}}\) being the dimerization constant:

\[
CI + CI \rightleftharpoons CI_2, \quad \text{with} \quad K_{\text{dim}} = \frac{[CI]^2}{[CI_2]},
\]

(S4)

related to the dimerization free energy, \(\Delta G_{\text{dim}}\) trough \(K_{\text{dim}} = e^{\Delta G_{\text{dim}}/k_B T}\), with a prefactor measured in Molar.

We explore the use of models to increase our understanding of the switch mechanisms of TP901-1. For a model to hold biological relevance it should
include and explain all or most of what is known about the system from previous experimental work. Thus despite a relative good fit between experimental data and the minimal model with no Cooperativity in CI dimer binding and the model including only cooperativity as CI is bound at all known operator sites, neither of these models are appropriate as they do not explain the importance of the O_L and O_D sites as observed in the mutant study of Pedersen et al. where the presence of the O_R site alone is not enough for the establishment of the immune state (6).

Of the two remaining models, we choose the model including just two cooperative energies; one for pair binding and one for triple binding of CI dimers, to keep the number of parameters at a minimum. Moreover this model returns the rough relations that $\Delta G_R \gtrsim \Delta G_L$ that are not straightforwardly deduced from the full model including separate cooperative energies for each of the CI dimer pair- and triple-wise bindings.

In consistency with experimental results the function of MOR should be to abolish the CI repression of P_L, while inflicting a tighter repression of P_R (6). We focused on two main models for the MOR dependent P_R repression; Model A where the CI:MOR complex associates at the DNA, and Model B where we assume complex formation in solution and subsequent binding at the DNA. In both models we assume that the complex bound at O_M prevents activity from the P_R promoter without disturbing the activity from P_L, and that binding at O_M is independent from CI binding at the known operator sites. The presence of MOR thus effectively doubles the number of binding states listed in Table S1.

Based on the model of pair- and triple-wise cooperative binding of CI dimers, model A simply includes additional binding of CI together with MOR at a putative DNA site. This binding species is given the statistical weight

$$f_s = [CI]^c[MOR]^m e^{-(c+m) \times \Delta G_M/k_B T}, \quad (S5)$$

where $\Delta G_M$ resembles the energy associated with both protein binding to DNA and possible cooperative binding between proteins. With previous results in mind, model A was tested with both $c = m = 1$, and $c = 1, m = 2$ (2). With Eq. S5 the activity of promoter P_R now becomes dependent on both CI and MOR, however, Eq. S3 remains unaltered and the direct MOR dependence can be eliminated by following steps similar to those described below for model B.

For model B we introduce the formation a heteromer complexes; CI:MOR.
Table S1: Microscopic configurations and associated energy contributions of CI repressor for the four Pr repression models investigated for the TP901-1 operator system.
The probability of each new state is now given by:

\[ f_s = \frac{[CI]^{i_s}MOR^{j_s}e^{-\Delta G_s/k_BT}}{\sum_s[i_s][MOR]^{j_s}e^{-\Delta G_s/k_BT}}, \]  

(S6)

with \( \Delta G_s \) now resembling the sum of free energies associated with multimer formation and operator site binding of the state \( s \) with a total of \( i_s \) CI monomers and \( j_s \) MOR monomers bound at the DNA. The \( P_R \) promoter activity probability thus becomes a function of both the free CI and free MOR concentrations, but the dependency on MOR can be eliminated based on the assumption that the complex binding will not affect activity from the \( P_L \) promoter.

In the experiments promoter activities are measured at an established steady state of the switch feedback system. The activity from the synthetic promoter upholds the concentration of CI, while the total concentration of MOR is a result of \( P_L \) activity (\( p_L \)) as governed by the dynamic equation:

\[ \frac{d}{dt}[MOR]_{\text{total}} = p_L([CI]) - [MOR]_{\text{total}} \times \frac{1}{\tau_M} \]  

(S7)

Rescaling the promoter activity in terms of the degradation rate of the protein (\( \tau \)) the steady state condition, \( \frac{d}{dt}[MOR]_{\text{total}} = 0 \), becomes:

\[ p_L([CI]) = [MOR]_{\text{total}} \]  

(S8)

Total MOR concentration is related to the free MOR concentration through

\[ [MOR]_{\text{total}} = [MOR] + m \times [MOR_{m:CI_c}] \]

\[ = [MOR] + m \times \frac{[MOR]^m[CI]^c}{K_{\text{het}}^{m+c-1}}, \text{ assuming} \]  

(S9)

\( c \times CI + m \times \text{MOR} \equiv \text{MOR}_{m:CI_c} \), with \( K_{\text{het}}^{m+c-1} = \frac{[MOR]^m[CI]^c}{[MOR_{m:CI_c}]} \) 

related to the multimerization energy through \( K_{\text{het}} = \exp(\Delta G_{\text{het}}/k_BT) \).

Finally expressing the probability of \( P_L \) activity in a similar manner as for the \( P_R \) promoter activity

\[ p_L([CI]) = p_{L0} \times (f_0 + f_3 + f_{0X} + f_{3X}) \]  

(S11)

where \( f_{sX} \) resembles the probability of the state with a \( \text{MOR}_{m:CI_c} \) complex bound in addition to the CI dimers bound in state \( f_s \), the \( P_L \) promoter.
activity probability may be given as a function of free CI concentration alone. However, as in the case of CI repression alone we still graphically present $pL$ plotted against the total CI concentration, which in the presence of MOR is given as:

$$[\text{CI}]_{\text{total}} = [\text{CI}] + 2 \times [\text{CI}_2] + c \times [\text{MOR}_m : \text{CI}_c]$$

$$= [\text{CI}] + 2 \times \frac{[\text{CI}]^2}{K_{\text{d}1}} + c \times \frac{[\text{MOR}]^m [\text{CI}]^c}{K_{\text{het}}^{m+c-1}} \quad (S12)$$

In the present case of $c = m = 1$ and the total CI concentration, $[\text{CI}]_{\text{total}}$, is easily expressed as a function of CI alone using Eq. S11 and S8 in determining $[\text{MOR}]$ as function of $[\text{CI}]$ and $pL([\text{CI}])$. However, introducing MOR means that $[\text{CI}]_{\text{total}}([\text{CI}])$ is no longer invertible. Thus each experimental value of total CI concentration may not be mapped onto a single value of free CI concentration as needed in determining the promoter activity probabilities with Eq. S2 and S11. Instead a range of CI free concentrations are used for simultaneous calculation of $[\text{CI}]_{\text{total}}$, $P_a$ and $P_L$.

In this study we focused on the case $m = c = 1$ as our data does not allow us to distinguish between complexes of different stoichiometry. However the restriction increase the focus on the effects of CI sequestration.
Document S2 - Theoretical relation between $\Delta G_{\text{dim}}$, $\Delta G_R$ and the measured dissociation constants

This section is based on chapter 7 in ref. (8). In our model of $\text{Pr}$ repression in the absence of MOR, CI binding to the known operator sites is assumed to proceed in two steps:

CI dimerization:

$$\text{CI} + \text{CI} \rightleftharpoons \text{CI}_2, \quad \text{with} \quad K_{\text{dim}} = \frac{[\text{CI}]^2}{[\text{CI}_2]} \quad (S13)$$

and CI dimer binding at an operator site, O;

$$\text{CI}_2 + O \rightleftharpoons \text{CI}_2 : O, \quad \text{with} \quad K_O = \frac{[\text{CI}_2] * [O]_{\text{free}}}{[\text{CI}_2 : O]} \quad (S14)$$

The measured CI-operator dissociation constants for $O_L$ (28 nM) and $O_R$ (~2000 nM) represent both of these processes. The bound fraction of CI dimers may then be expressed as:

$$\frac{[\text{CI}_2 : O]}{[O]_{\text{total}}} = \frac{[\text{CI}_2]}{K_O + [\text{CI}_2]} \quad (S15)$$

In terms of the free CI monomer concentration this becomes

$$\frac{[\text{CI}_2 : O]}{[O]_{\text{total}}} = \frac{[\text{CI}]^2}{K_O * K_{\text{dim}} + [\text{CI}]^2} \quad (S16)$$

With half occupancy of the operator site at a free CI monomer concentration of $[\text{CI}]_{\text{free}} \sim \sqrt{K_O * K_{\text{dim}}}$. To compare to the experimental dissociation values we express the bound fraction in terms of total CI concentration. Using Eq. S3 and S13 and solving for the free CI monomer concentration:

$$K_{\text{dim}} = \frac{[\text{CI}]^2}{[\text{CI}_2]} = \frac{[\text{CI}]^2}{1/2 * ([\text{CI}]_{\text{total}} - [\text{CI}])} \quad \Downarrow$$

$$[\text{CI}] = - \frac{K_{\text{dim}}}{4} + \frac{K_{\text{dim}}}{4} \sqrt{1 + \frac{8}{K_{\text{dim}}} * [\text{CI}]_{\text{total}}.} \quad (S18)$$

Depending on the value of the total CI concentration the fractional occupancy of an operator may be estimated as:
\[ \frac{[\text{Cl}_2 : \text{O}]}{[\text{O}]_{\text{total}}} \approx \frac{[\text{Cl}]^2_{\text{total}}}{K_O K_{\text{dim}} + [\text{Cl}]^2_{\text{total}}} \quad \text{for } [\text{Cl}]_{\text{total}} \ll K_{\text{dim}} \quad (S19) \]

\[ \frac{[\text{Cl}_2 : \text{O}]}{[\text{O}]_{\text{total}}} \approx \frac{[\text{Cl}]_{\text{total}}/2}{K_O + [\text{Cl}]_{\text{total}}/2} \quad \text{for } [\text{Cl}]_{\text{total}} \gg K_{\text{dim}} \quad (S20) \]

Thus at predicted \( K_{\text{dim}} \) values well above \( [\text{Cl}]_{\text{total}} = 10^{-6} \) M the measured \( O_R \) dissociation constant of \( \sim 2000 \) nM should represent approximately \( \sqrt{K_R * K_{\text{dim}}} \). In our simulations low \( K_{\text{dim}} \) values are associated with high \( K_R \) values resulting in \( \sqrt{K_R * K_{\text{dim}}} \approx 700 \) kcal/mol (see Table S2). Same considerations can be made for relating parameter sets and the measured \( O_L \) dissociation constant.

In the opposite limit, where predicted \( K_{\text{dim}} \) values are well below \( [\text{Cl}]_{\text{total}} = 10^{-6} \) M the measured dissociation constant should represent approximately \( 2*K_R \). In the simulations of the constrained fit the dimerization is indeed strong resulting in \( 2*K_R \) values between 2000 nM - 200 \( \mu \)M. Parameters resulting in the high \( \mu \)M regime thus does not represent thermodynamic parameters relevant for this system. However, they are included in Table S2 for completeness and illustrate how the model allow these states by compensating with stronger pair- and triple-wise cooperativity between bound dimers.
Table S2: Parameters of best fits. Final four columns are calculations based on the parameters for comparison to the measured dissociation constants of CI binding at O_R (∼2000 nM, column 9-10) and O_L (28 nM, column 11-12). Refer to text for details.

<table>
<thead>
<tr>
<th>Unconstrained fits</th>
<th>Constrained fits</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSR* Pm max</td>
<td>ΔG_{Pm}</td>
</tr>
<tr>
<td>[I/μL]</td>
<td>[kcal/mol]</td>
</tr>
<tr>
<td>60.8</td>
<td>26.3</td>
</tr>
<tr>
<td>61.7</td>
<td>27.2</td>
</tr>
<tr>
<td>62.5</td>
<td>27.4</td>
</tr>
<tr>
<td>62.7</td>
<td>25.3</td>
</tr>
<tr>
<td>63.2</td>
<td>28.8</td>
</tr>
</tbody>
</table>

*: Sum of squared differences between simulation and data
†: K_{R}=[1M]exp(ΔG_{R}/k_{B}T), K_{dim}=[1M]exp(ΔG_{dim}/k_{B}T)
‡: K_{L}=[1M]exp(ΔG_{L}/k_{B}T), K_{dim}=[1M]exp(ΔG_{dim}/k_{B}T)
Appendix C


Paper submitted to Genes and Development in August 2012. Alsing and Sneppen 2012
Differentiation of developing olfactory neurons analyzed in terms of coupled epigenetic landscapes.

Anne Katrine Alsing 1 and Kim Sneppen 1∗

1Center for Models of Life, Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark

Received November 1, 2012; Revised Month date, 20XX; Accepted Month date, 20YY

ABSTRACT

The olfactory system integrates signals from receptors expressed in olfactory sensory neurons. Each sensory neuron expresses only one of many similar olfactory receptors (ORs). The choice of receptor is made stochastically early in the differentiation process and is maintained throughout the life of the neuron. The underlying mechanism of this stochastic commitment to one of multiple similar OR genes remains elusive. We present a theoretical analysis of a mechanism that invokes important epigenetic properties of the system. The proposed model combines nucleosomes and associated read-write enzymes as mediators of a cis-acting positive feedback with a trans-acting negative feedback, thereby coupling the local epigenetic landscape of the individual OR genes in a way that allow one and only one gene to be active at any time. The model pinpoint that singular gene selection does not require transient mechanisms, enhancer elements or transcription factors to separate choice from maintenance. In addition our hypothesis allow us to combine all reported characteristics of singular OR gene selection, in particular that OR genes are silenced from OR transgenes. Intriguingly it predicts that OR transgenes placed in close proximity should always be expressed simultaneously, though rarely.

INTRODUCTION

The first step in odor reception in mammals is effectuated in the almost canonical one neuron-one receptor rule (1, 2). Each olfactory neuron expresses only one allele out of a large and highly homologous gene family, comprising almost 1400 olfactory receptor (OR) genes in mice. (3, 4, 5, 6). Millions of olfactory neurons comprise the olfactory epithelium (OE). Olfactory neurons expressing a particular OR gene are confined to zones along the dorsal-ventral axis of the OE in mice, with possible overlaps between differently segregated zones (7, 8, 9). The coding region of a general OR gene is just 1kb. However, transgenes must include a much larger part of the OR gene region in order to reproduce expression patterns parallel to endogenous genes (10, 11, 12). During development the neuron extends its axon to a receptor defined glomerulus of the olfactory bulb, ensuring the essential conversion of olfactory signals to a typographical map in the bulb (13, 14, 15). Main characteristics of the olfactory neuron differentiation are summarized in table 1 (reviewed recently by (4, 16, 17)).

In choosing to express just one amongst thousand of OR genes each olfactory neuron represents a system with multiple stable states. Numerous mechanisms have been put forth to describe the underlying nature of this seemingly stochastic multi-stability. So far all proposals fail to fully encompass both the choice and the maintenance of the expression of a single OR gene. Most description like limiting transcriptional complexes or singular enhancer elements pictures plausible ways of stochastically choosing a single OR gene for expression but resort to an unspecified feedback mechanism in order to maintain the memory of the chosen OR gene (11, 27, 28, 31, 35, 36, 37, 38). Recent reviews that hypothesize the involvement of chromatin re-modeling in upholding OR gene selection implement similar fussy feedback mechanisms in establishing the choice (4, 39). Moreover, though based on experimental observations, all proposals remain descriptive and lack a mean for reliably testing if the dynamics of the system would in fact be as envisioned.

In this article we present a quantitative analysis of a model of OR gene choice inspired by the proposed involvement of epigenetic modifications. We revisit a theoretical approach on epigenetic cell memory by nucleosome modification, initially taken by Dodd et al. (40, 41) and explored in the context of vernalization in plants by Angel et al. (42). In the two state version of this theoretical analysis a DNA region containing L nucleosomes is considered (43, 44). Each nucleosome may be in one of two states; silent, or active. Transition between states are made randomly or by active recruitments. In a random event a nucleosome spontaneously converts to the other state. During recruitment a histone-modifying enzyme is recruited by nearby nucleosomes and is thought to modify the nucleosome at hand to match the modification of the recruiting nucleosomes. When these recruitment processes include either
implicit or explicit cooperativity, the state of the DNA region can be bistable (40, 44), with the majority of the nucleosomes being in either the silent or active state, see Fig. 1 B.

One may associate the two nucleosome states of the model with permissive and repressive methylation marks as recently reported for the active and inactive OR gene regions by Magklara et al (34). Alternatively, the states could be DNA methylation marks (45) or increasing levels of histone acetylation. Histone de-acetylases (HDACs) are reported in immature and mature olfactory sensory neurons (OSNs) (46, 47), and in vitro cultured vomeronasal progenitor cells of adult rats only develop the adult neuronal phenotype when subjected to HDAC inhibitors (48). We employ a compressed description of these multi-step scenarios where different levels of histone modifications dictate activity of the genes.

We study whether olfactory differentiation can be described by combining several subsystems with internal architectures similar to those of the nucleosome model. Each OR gene should then be in one of two general states; either activated or silenced. To mimic the mutual exclusion of expression amongst OR genes we couple the subsystems through a hypothesized factor, effectively governed by active OR genes, see Fig. 1 A. The factor is thought to prevent the conversion of silent nucleosome marks to active ones, perhaps through binding and shielding of the silent marks, or by inhibiting the activating enzymes. Thus effectively the factor represses the positive feedback towards active nucleosomes, see Fig. 1 C.

Formulating our model of OR gene selection as a stochastic simulation we examine how the model captures the properties of the system and identify the effect of the involved mechanisms. We demonstrate that combining a cis-acting positive feedback mediated by nucleosomes and associated read-write enzymes, with a transacting negative feedback, encompass or reproduce all the characteristics of OR gene selection summarized in the table.

### METHODS

The model of epigenetic cell memory through nucleosome modification allows for sensitive gene regulation by positive feedback loops (40, 41). Thus providing us with a predefined bistable subsystem to describe each OR gene. The real OR gene family is one of the largest mammalian gene families.

In modeling we consider system sizes of 10, 100 or 500 similar genes or subsystems consisting of \( L = 50 \) nucleosomes. With this simplification the system may be thought to describe the random choice between potential genes after zonal exclusion (39). The individual OR genes are coupled through modification reactions that increase with the activity of all OR genes. This global negative feedback favors the silent state. The model, outlined in Fig. 1, is formulated as a stochastic simulation as in (40, 41), and analyzed in terms of deterministic equations.

#### Governing equations

The negative feedback act through some protein factors that effectively are governed by the gene activity of all genes:

\[
P \propto a_1^h + a_2^h + \ldots + a_N^h = \sum_j a_j^h \quad (1)
\]

where \( a_j = A_j / L \), with \( A_j \) being the number of active nucleosomes of gene \( j \). P binds to silenced nucleosomes and

<table>
<thead>
<tr>
<th>Table 1. Characteristics of olfactory neuron differentiation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>- The OR gene family of mice includes more than 1000 genes spread out over most chromosomes. About 85% of these can express functional receptors (18). In humans only ( \sim 40% ) are functional (19, 20).</td>
</tr>
<tr>
<td>- The OE is divided into 4 zones and an overlapping patch more or less restricted to a distinct group of ORs (21, 22, 23, 24).</td>
</tr>
<tr>
<td>- The majority of all mature neurons express only one receptor (2, 25, 26).</td>
</tr>
<tr>
<td>- Expression of an OR protein enforce a feedback that (normally) keeps the neuron from expressing a second OR (23, 25, 27, 28).</td>
</tr>
<tr>
<td>- Immature neurons can switch between expressing different ORs, including the two allelic versions, prior to maturation (25).</td>
</tr>
<tr>
<td>- At least 40% of neurons die before they fully mature (29, 30).</td>
</tr>
<tr>
<td>- Pseudogenes are OR genes that do not result in a functional receptor and may not implement the presumed feedback. Pseudogenes can be co-expressed with a functional OR (27, 28).</td>
</tr>
<tr>
<td>- Transgene expression of an OR gene from a promoter associated with early transcription results in OR expression over the full OE (12).</td>
</tr>
<tr>
<td>- Enhancer elements upstream of an OR locus alter the probability of the associated OR genes being chosen for expression (26, 27, 31).</td>
</tr>
<tr>
<td>- Promoter and coding region of ORs contain TF binding sites (10, 32, 33). However, identical transgenes are not co-expressed and thus OR gene choice can not be fully governed by TFs (2).</td>
</tr>
<tr>
<td>- Silenced OR genes are covered by nucleosomes marked with H3K9me3 and H4K20me3 methylation, whereas the nucleosomes associated to active OR-genes are marked with H3K4me3 (34).</td>
</tr>
</tbody>
</table>
Nucleic Acids Research, 20XX, Vol. YY, No. ZZ

RESULTS

Stable activity of one and only one subsystem.

Our model of OR gene expression captures the basic property of the system, namely the exclusive expression of just one gene out a large highly homologous gene population. We initiate the system with all nucleosomes in the silent state consistent with the observations by Magklara et al. (34). Within the simulations we identify the state of the individual gene, i, as active, when the active fraction of nucleosomes, \(a_i\), exceeds 2/3. Fig. 2 show a simulation of \(N=10\) genes covered by \(L=50\) nucleosomes each. Gene 5 achieves activation as the number of active nucleosomes exceeds 33 at time \(\sim 60\). Stochastic fluctuation and internal local bias toward the active state move the subsystem into a dominant

1. Select 2 random nucleosomes in one random gene. If both are active, one with probability \(R\) select another nucleosome in this gene and set its state to active. If the two nucleosomes are in the silenced state one with probability \(\mu\) select another nucleosome and set its state to silenced.

2. Select with probability \(\beta\) a nucleosome among all genes in the system. If the chosen nucleosome is active, then change it to silenced. If the chosen nucleosome is silenced then change it with probability \(R\) to active.

One time-unit corresponds to one update per nucleosome in the entire system of \(L \times N\) nucleosomes. Thus, in each time unit of a simulation every nucleosome is on average attempted converted towards the active state, with actual conversions happening much less frequently.

The OE is constantly renewed through neurogenesis (49, 50, 51). In adult rats regeneration of olfactory receptor neurons (ORNs) from immature progenitors is achieved within 2 weeks (29). Similar timescales of neurogenesis are observed in embryonic development. (30, 52). In contrast reported lifetimes for neurons in rodents are very variable, ranging from one month to almost a year (29, 53, 54). With a turnover rate of histone modifications in the order of 10 min reported for the de-acetylation processes (55, 56, 57, 58), we thus want our model to choose one subsystem within 1 week \(\sim 1000\) time units in the simulations and contain exclusive expression of that subsystem for at least 1 month \(\sim 5000\) time units. Other modification processes may well be slower (59, 60, 61) but such change of timescale does not change our overall results. It would only move the parameter range for which we have acceptable differentiation properties.

For each simulation we hence run 5000 time-units, starting from a all silenced state, and requiring that only one gene turns on within this time window and that only one gene is active in the time-window.

In thread with previous nucleosome models we score a subsystem as activated when the fraction of actively marked nucleosomes exceeds 2/3. Activation is considered lost when the fraction falls below 1/3 (40), but in fact an insignificant fraction of the genes that switch to become the single dominating gene will switch back to the silent state within the plotted time frame. The model was implemented in both C++ and fortran, and are available upon request.
active state. Activation increases the globally acting negative feedback thus reinforcing the dominant gene by decreasing the probability for local activation of other genes.

Success of a simulation may be accessed on three criteria. First the system needs to selectively activate a single OR gene within a given time window. Secondly, the chosen gene should remain active for a considerable time. Thirdly, no other OR genes may be activated while the initial OR gene is active. Fig. 3 A,D & G compactly show time courses like those of Fig. 2 for a system of 100 genes at three different values of the local activation bias parameter $\mu$. It is clear that increasing $\mu$ takes the system from defying the first criteria by switching on more than one gene, to fulfilling all criteria with a single active gene, and to failure due to lack of turn on of any genes.

### Activation barriers capture genes in silent state

While simulations present a proof of concept for our model in reproducing the exclusive gene expression, the theoretical formulation of the model also let us examine in detail how the dominant gene expression is achieved. Arguing that all but the two most active genes are found at the same low level of activation, we graphically identify three stable fix points for the states of the two genes (see methods). The colored regions in Fig. 3 B,E & H, show where the net variation in gene activity fraction is positive for the two most active genes. Intersections between the regional limits are fix points in gene activity fraction is zero. Fix points closest to the corners of the $a,b \in [0;1]^2$ are stable as local variation in activity in gene “a” and “b” deterministically will return the system to the states of the fix points (see methods and supplementary section S1).

The 2-d planes in Fig. 3 B,E & H, include trajectories of the first versus the second most active gene from the simulations to the left and show how the success relies on two barriers. At low $\mu$, examined in Fig. 3 B, the “first” barrier stalls the system at the state where all genes are silenced until a random fluctuation causes activation of one of the genes. Subsequently, the system remains unstable against the passage of a second barrier to the state in the rightmost upper corner of Fig. 3 B. At larger $\mu$ only one gene switches, as the second and larger barrier along the vertical axis in Fig. 3 E prevents a second gene from turning on once an initial switch have been made. At even larger $\mu$ values the first barrier is so high that no gene may pass it, essentially stalling the system with all OR genes in the silent state (see Fig. 3 H). Thus, for a range of $\mu$ values the system is successfully stalled in the lower right corner of the $[0;1]^2$ phase space with just one activated OR gene.

Interestingly, when the system is stalled in the lower right corner of Fig. 3 B, E or H, the repressed genes have a small fraction of their nucleosomes in the active state. This “nucleosomal noise” should cause a residual production of receptors from each repressed gene of about $\beta^3$ (see supplementary section S1). Sensitivity of the cell sensing system to the chosen gene, on the other hand, provide information on the threshold for protein production, here parametrized by the hill coefficient $h$.

### Parameter sensitivity

Even for the best of parameter combinations a fraction of the simulations will fail to fulfill the success criteria, see Fig. 4. In the Olfactory neuron the loss of OR memory after axon extension to the predefined glomerulus of the olfactory bulb would be catastrophic. However, it is seen that about 50% of neurons born in the epithelium of adult rats are lost within 5 days and 2 weeks (29, 30). There are no explanation to this loss, but it has been argued that it could be a control mechanism to get rid of ORN expressing more than one receptor (39). Thus for the model to successfully represent differentiation of olfactory neurons only 50% of the simulations need to be successful, indicated by the gray shaded area in Fig. 4 A. We initially set the criteria that the first OR should activate within $t<1000$ time units (see methods). However, this restriction is somewhat conservative and relaxing it widens the success rate of the model, shown as concurrent orange areas in Fig. 4 A.

Additionally, the one-neuron one-receptor rule may not be quite as manifested as widely accepted within the field (6, 64, 65). To this end, we include in figure 4 B, the maximal activity of the second and third most active genes. These show that allowing a second or even third gene to activate further widens the $\mu$ range of successes. However, we note that the model does not address how such tandem expression would affect axon guidance.

### Robustness to gene copy number.

The diversity and wide chromosomal position of OR gene clusters are generally considered a result of subsequent tandem gene duplications, gene conversions and recombination events and even conversions of entire coding regions. Accordingly the mechanism of gene selection should be robust to quite drastic changes in gene numbers (4, 19). The model successfully addresses this issue by selecting one OR gene, keeping the remaining genes silent, for essentially the same set of parameters when the system size is increased from 100 to 500 genes, see supplementary Fig. S2 A.
and supplementary material). In the blue region for the two most active genes in the system, obtained by stochastic simulation over an epigenetic landscape (62, 63), with states that to varying degree prefer to be in the corners, see also Fig. S1. Parameters are a fraction $\beta = 0$ time. Crossed arrows indicate silent promoters. Middle panels activated (red) and a few examples of other genes (yellow and green). Inserts show the promoter status of the correspondingly coloured genes at the final simulation.

Figure 3. Simulation of $N = 100$ genes, each covered by $L = 50$ nucleosomes. Left panels A,D,G show time course of the first activated gene (blue), second activated (red) and a few examples of other genes (yellow and green). Inserts show the promoter status of the correspondingly coloured genes at the final simulation time. Crossed arrows indicate silent promoters. Middle panels B,E,H follow the trajectory of the two most expressed genes, identified by their active nucleosome fraction $a$ and $b$, in a 2-d plane that illustrates deterministic drift of two individual genes, provided all other genes are assumed to act synchronously (see methods and supplementary material). In the blue region $da/dt > 0$, whereas the red region shows where $db/dt > 0$. C,F,I show the probability (lighter colour for higher) for the two most active genes in the system, obtained by stochastic simulation over $10^6$ time-units. The negative logarithm of this probability may be interpreted as an epigenetic landscape (62, 63), with states that to varying degree prefer to be in the corners, see also Fig. S1. Parameters are A-C $\mu = 0.30$, $r = 1$, $h = 2$ and $\beta = 0.03$. D-F $\mu = 0.50$, $r = 1$, $h = 2$ and $\beta = 0.03$. G-I $\mu = 0.70$, $r = 1$, $h = 2$ and $\beta = 0.03$.

Robustness to model variations.

Remarkably, the overall behaviour of the model does not depend on details on how the recruitment processes and global feedback are organized. In the supplement we examine 4 variants of a cis-acting positive feedback based on nucleosomes, see supplementary Fig.s S4-S7. In particular we consider the 3-state nucleosome model from (40), and various ways that the negative feedback can modulate the recruitment processes. In all cases we persistently reproduce a working olfactory system, provided 1) effective cooperativity in the recruitment processes, and 2) that the global negative feedback acts through a production term that at least grows with the square of the active nucleosome fraction of each gene. In principle this last requirement of cooperativity in the negative feedback could be relaxed, but $h > 1$ greatly facilitate the contrast between the active and silenced genes, see supplementary Fig. S2 B-G.

Pseudogenes, delayed feedback and early transient switching.

Within the model settings that restrict the stable activation to a single OR, we considered the concept of pseudogenes. In this context these are subsystems affected by feedback as previously described but lacking the ability to produce the feedback. It will occasionally happen that first a pseudogene is activated, however, as the active pseudogene does not contribute to the negative feedback, $R$, another subsystem will eventually be activated and retain the dominant stable position, see Fig. S1 A. Experiments, where one expresses an OR transgene from a promoter that is active early in the development of the OSN, show that this gene will dominate the future differentiation (12). This is in accordance with our negative feedback hypothesis which indeed predicts that any early dominance will prevail. In general developing neurons may show transient
activation of different OR genes (25). Fixing the parameters at functional values like those of Fig. 3 D, initial transient activation of more OR genes is possible if we introduce a time delay, $\tau$, between OR activation and feedback production (See Fig. 5 B and methods).

Our standard model is appropriate when the factors facilitating the negative feedback have a life time that is shorter than it takes one gene to switch from silent to active state. In case the degradation is slower, the early activity of the OR system is often altered, with several genes turning on at very early times, for then subsequently to loose their activity when the full effect of the, then overproduced, negative feedback comes into play. As a result the qualitative behaviour from Fig. 4 is reproduced with all time-delays that is substantially smaller than the maturation time of 1000 time-units (see supplementary Fig. S3).

The frequency of initial activation of a pseudogene is higher in systems without time delay due to the lack of feedback production from pseudogenes. In the case of time delay, real genes do not counteract their own activity while turning on and therefore switch as often as the pseudogenes (see Fig. 5 B).

Enhancers and modulation by transcription factors

Enhancer elements with features similar to the locus control region of the visual pigment genes have been identified for some subfamilies of OR genes (26, 27, 31). However, such elements are yet to be identified for all OR genes of the super gene family (31). In the context of our standard olfactory nucleosome modification model we envision enhancer elements as a mean of disposing the OR gene for epigenetic activation. Decreasing the probability, $\mu$, of a recruited event towards the silent state for one subsystem by as little as 15% greatly increase the frequency of that subsystem being activated, Fig. 6 A. Such an increase in frequency for H-element associated genes has been reported (31, 52, 66).

Conserved regions identified in alignments of OR genes, include binding sites of known transcription factors (TFs), (10, 32, 33). To our knowledge, so far no single TF or TF binding sites have been associated with the zonal exclusivity, even though TFs as Emx2 do alter the expression frequency of a large part of the OR genes (33). In our model TFs may be incorporated, like enhancer elements, as chromatin modification control mechanisms (41), enhancing or reducing the probability of activation of their associated genes, see Fig. 6 A. The model shows that simple TF knockout experiments will not necessarily allow for a clear identification of such TFs as full exclusion of associated OR genes. Instead the frequency by which the genes are chosen will decrease while unrelated OR genes increase as reported for the Emx2 knockout mice (33).

Identification of regulatory elements

Reports on the size of the regulatory elements needed for a recapitulation of the punctuate OR gene expression in the OE, varies from ~1kb to more than 10kb (10, 11, 12), with varying intrinsic levels of spatial restriction. The short and highly variable OR coding sequences may be sufficient to mark a transgene for the same feedback regulation as endogenous genes (12). In our model we initially assume a regulatory size of ~10kb or at least large enough to encompass 20-50 nucleosomes. However, in more general terms the model presents a way of activating a single patch of the genome, including the feedback producing element, which is responsible for keeping the remaining patches silent. In fact, the average OR genes of a locus are separated by 29kb (3), allowing for separation between activated and silent regions, and implying that the activity of their respective nucleosomes is separated by substantial barriers. In transgene experiments one may modify distances between genes to examine the spreading mechanisms associated to the local positive feedback. In particular one may insert two genes just after each other, with identical promoters, and investigate whether such an architecture facilitate non-synchronous turn-on of the genes, or synchronous activation as our model suggests. Notice that with standard promoter strength, the globally acting negative feedback will counteract any activity from such a coupled pair of genes (see Fig. 6 B) and thus it is recommended to reduce the strength of the two promoters.

**DISCUSSION**

Early in our approach to the olfactory system we recapitulated the need for multiple stable states in the system. Relying on mechanisms like TF or enhancer element binding to the chosen OR gene inevitably couples the stability of the choice to these associations. As a result such models often incorporate an unspecified “feedback mechanism” to lock the system to one OR (16). In contrast, within a considerable time window, our nucleosome modification model of the olfactory receptor gene regulation, does not require solidifying agents. The model pinpoints the need for localizing the dependence of olfactory co-regulation on the size and status of chromosomal modifications of the individual olfactory gene locus. It also identify the need for finding the common factor that facilitates the interplay between the different olfactory genes. In particular we aim to direct the search for this later globally acting feedback factor within pathways associated with histone modifying enzymes. The model presented explicitly points at histone modifiers that favour the marks associated to protecting the silenced state. In the supplementary section S3 we examine other variants of the model, noting the possibility for other feedback pathways (see supplementary Fig. S7).

Important experiments addressing the former role of the local nucleosome mediated feedback in the model could be initiated without explicit knowledge of the players. As previously described, placing two olfactory receptor genes close on the DNA should result in co-expression, provided a reduction in the strength of the associated promoters. A way to experimentally circumvent this need for reduced promoters would be replacing one of the proximate genes with a pseudogene. Lack of consistent co-expression would falsify our model. Alternatively it should be possible to express two olfactory receptors from non-olfactory regions, provided that these regions of the genome in some way are protected from silenced nucleosomes. We again emphasize that our model does not address how the neuron will cope with such eventual co-expression.
CONCLUSION

Perspectives for olfactory differentiation as a model system for coupled epigenetic landscapes

A final speculation springs from the fact that olfactory neuronal cells and the sense of smell is associated to the oldest part of the cortex. Thus their genetic regulatory design may be found more widespread. In fact the olfactory bulb may well be the predecessor of the enlarged cortex of mammals (67), and differentiation of olfactory neurons could give insight into the way we store memories in the brain. In particular, Fig. 4 open for the possibility to activate two receptors in a cell, a possibility that open for a large increase in the number of epigenetic states. As ORs are believed to direct the axon-to-glomerulus association at the olfactory bulb, receptors may also modulate physical connections between other neuronal cells. Thereby receptors could couple the intrinsic epigenetic state of cells to the neural network-architecture. An architecture where neurons which share one receptor could be coupled, and build branching signalling pathways when cells are allowed to have more than one receptor. Accordingly, the ability for individual neuronal cells to differentiate their behaviour on the basis of expression of membrane bound receptors opens for a new perspective on memory. A perspective where the epigenetic state of individual neuronal cells may be coupled to the real memory storage in our cortex.

ACKNOWLEDGEMENTS

We would like to thank Ian Dodd and Danesh Moazed for discussions and Szabolcs Semsey for valued comments on the manuscript. The work was supported by Danish National Research Foundation through the Center for Models of Life.

Conflict of interest statement. None declared.

Supplementary Data are available at NAR online: Supplementary methods with Supplementary figures S1-S7, all referred to at appropriate points.

REFERENCES

8 Nucleic Acids Research, 20XX, Vol. YY, No. ZZ


Figure 1. Model for maintenance of selected gene expression in an olfactory neuronal cell. A Local positive feedback and global negative feedback. N genes each covered by nucleosomes (Red for silently marked nucleosomes, Blue for actively marked). The nucleosomes on each gene form positive feedback systems that maintain the selected expression state (Red and blue arrows). The active gene expresses receptor proteins which direct enzymes that favour nucleosomes to stay in a silenced state by repressing the positive feedback towards the active nucleosome state (black arrow and blunted lines, respectively). B Switching property of a single gene. Each gene is covered by \( L \) nucleosomes that each can be in an active state that is open for transcription (Blue) or in a silenced state (Red). The nucleosomes together with associated "read-write" enzymes form positive feedback systems that allow a gene to maintain a previously selected state (Blue and red arrows). Enzymatic activity effectuated by the activity of all active genes captures genes in the silent state (dashed black arrow and blunted line). C Nucleosome modifications recruit read-write enzymes within each gene, as in ref. (40) but for simplicity we consider only two nucleosome states which each can modify the states of other nucleosomes cooperatively. The feedback between genes described in A) and B) acts through protection of silenced nucleosomes against conversions to the active state with strength proportional to \( \sum_i a_i^h \) where \( a_i \) is the active nucleosome fraction of gene \( i = 1, \ldots, N \).

Figure 2. Simulation of \( N = 10 \) genes, each covered by \( L = 50 \) nucleosomes. The simulation show that one gene is turned on quite early, while all other genes remain silenced throughout the simulation. In fact with these parameters all these genes stay silenced up to at least \( t = 5000 \) time units. Sketches in the right panel illustrate the OR gene state at the final time of the simulation. Red for silently marked nucleosomes, blue for actively marked. Crossed promoters indicate silent genes. Other parameters of the simulation are \( \mu = 0.50 \), overall repression factor \( r = 1 \), hill coefficient of repression \( h = 2 \) and noise conversion \( \beta = 0.03 \).
Figure 4. Parameter sensitivity for $N = 100$ genes, $L = 50$ nucleosomes system with repression factor $r = 1$, hill $h = 2$ and noise $\beta = 0.03$ fixed. Data is averaged over 200 simulations. A Orange area shows the probability that one and only one gene becomes active within 5000 time-units as function of asymmetry $\mu$. Concurrent dark orange area marks success with the additional constraint that one gene becomes active within the first 1000 time-units. Gray area marks the cut-off at 50% successful simulations. B Orange area marks the largest number of active nucleosomes within one gene during a 5000 time-unit simulation. When no genes reach full activity during the first 5000 time-units, no olfactory receptors have turned on. Cyan and dark cyan show maximal activity of the second and third most active genes, respectively. Where the second most active gene has many active nucleosomes, the two genes have shown simultaneous activity. See also Fig. S2.

Figure 5. Pseudogenes and time-delay: A Activity of genes as function of time for our standard model with of $N = 100$ genes, and $N_{\text{pseudo}} = 40$ pseudogenes each covered by $L = 50$ nucleosomes, noise $\beta = 0.03$, asymmetry $\mu = 0.50$, overall repression factor $r = 1$, and hill $h = 2$. The dotted trajectories are pseudogenes, and illustrate that several pseudogenes may become activated early. The small promoter pictures show activity status of the correspondingly coloured genes in the covered time, with pseudogenes shown in gray shaded box. B As above, but with a time delay, $\tau$, between OR activation and the feedback $R$. The time delay opens for transient activation of several OR genes. Once $R$ accumulates the active ORs are repressed and only one OR gene remains active. As genes only sense their own activity after some time delay, the turn-on frequency for real genes is the same as for the pseudogenes. See also Fig. S3.
Figure 6. Predictions of sensitivity to enhancers and to double gene dosage:
A Effect of small alterations in OR specific $\mu$ on the probability of switching on for the associated OR gene. Individual decrease (orange bars) or zonal increase (gray bars) in the activation asymmetry $\mu$ might represent the effect of an enhancer element or the removal of zonal specific TFs. Each data set from 1000 simulations, with $N=10$ genes, each covered by $L=50$ nucleosomes, $\mu=0.50$, $\beta=0.03$, $r=1$, and $h=2$. B Comparison between wild type case, and a engineered situation where one gene contributed with double gene dosage to the feedback $R$. The likelihood of turning on the “doubled” gene shown with orange bars is smaller than for a normal gene to an extent that depend on $\mu$. In case the “doubled” gene is constructed as a tandem repeat, possibly with two reporter proteins inserted, the two identical promoters should become active together, but with smaller probability than the wild type system. A reduction in probability that will pinpoint the effective value of $\mu$. Each data set from 1000 simulations, with $N=100$ genes, each covered by $L=50$ nucleosomes, noise $\beta=0.03$, $r=1$, and $h=2$. 
Supplementary Material for
“Differentiation of developing olfactory neurons analyzed in terms of coupled epigenetic landscapes”

Anne Katrine Alsing∗ and Kim Sneppen∗∗

Center for Models of Life, Niels Bohr Institute,
University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen Ø

∗ alsing@nbi.dk
∗∗ corresponding author; sneppen@nbi.dk

S1 - Stability analysis of the model

To analyze eq. 4 in the main text we consider two single genes, with active nucleosome fractions a and b, to be freely varying in a situation where all other N − 2 genes are slaved to a single variable active nucleosome fraction, c.

\[
\frac{da}{dt} = \frac{1}{1 + r(a^h + b^h + (N - 2)c^h)} \cdot a^2 \cdot (1 - a) - \mu \cdot a \cdot (1 - a)^2 - \frac{\beta \cdot a}{1 + r(a^h + b^h + (N - 2)c^h)}
\]

\[
\frac{db}{dt} = \frac{1}{1 + r(a^h + b^h + (N - 2)c^h)} \cdot b^2 \cdot (1 - b) - \mu \cdot b \cdot (1 - b)^2 - \frac{\beta \cdot b}{1 + r(a^h + b^h + (N - 2)c^h)}
\]

\[
\frac{dc}{dt} = \frac{1}{1 + r(a^h + b^h + (N - 2)c^h)} \cdot c^2 \cdot (1 - c) - \mu \cdot c \cdot (1 - c)^2 - \frac{\beta \cdot c}{1 + r(a^h + b^h + (N - 2)c^h)}
\]

Main text Fig. 3 B,D,F, show the sign of \(\frac{da}{dt}\) and \(\frac{db}{dt}\) assuming that c takes the lowest value where \(\frac{dc}{dt}\) = 0. This assumption is reasonable
because $dc/dt > 0$ for $c = 0$ and for increasing $c$ the first null cline should therefore be stable against variations in $c$.

From eq. 3 we can deduce the fraction of active nucleosomes in repressed genes, as this is reflected in the lowest $c$ value where $dc/dt = 0$. This “nucleosomal noise” is approximately $c = \beta/(\mu(1 + r)) \sim \beta$. Using main text eq. 1, this minority fraction of active nucleosomes is associated to a minority production from each repressed gene of about $\beta^h$. For comparison, the receptor production from the chosen active gene is $a^h \sim 1$ where $a$ is determined from largest $a$ that fulfils eq. 1. This $a$ value correspond to the “$a \sim 1$” solution in the lowest right corners of main text Fig. 3B,E,H. Requiring that the cell response is dominated by the chosen gene, imply that the activity of the chosen gene should be in excess of all the remaining N-1 genes together, $a \sim 1 > N \cdot \beta^h$, a demand that is fulfilled for all examined parameters.

S2 - Robustness to model parameters

The investigation presented in the main text uses standard parameters, with $N = 100$ genes each covered by $L = 50$ nucleosomes that interact with parameters $\beta = 0.03$, $r = 1$, hill coefficient $h = 2$ and time-delay $\tau = 0$. Here we demonstrate that varying these parameters leave our main prediction robust.

Fig. S1 show a 3-d visualization of the epigenetic landscapes for small $L$ (left column), respectively for a time-delay $\tau = 100$ (right column). These simulation can be compared to our standard parameter simulations in the middle column.

Fig. S2 A show that model predictions can be reproduced with a larger gene number $N = 500$, thus easily accounting for differentiation in a realistic size olfactory system. Also the figure show that sharper threshold for gene activity of the individual genes, parameterized by higher hill coefficient $h = 4$, allow for more robust cell differentiation (larger range of $\mu$ where model works). Also smaller noise $\beta = 0.01$, instead of $\beta = 0.03$, facilitate robustness.

Fig. S2 B-G examines the success probability of simulations as function of $r$ and $\mu$ at different hill coefficients, noise levels and nucleosome numbers. It is evident that the model is robust to changes in all parameters.

Fig. S3 extend Fig. 5 from the main text, by examining systematically the effect of time-delay $\tau$ on probability for successful differentiation. Furthermore, in the last panel of Fig. S3 we examine pseudogenes, here defined...
Fig. S 1: Related to Fig. 3 in main text. Epigenetic landscapes. Here we examine the probability (lighter colour for higher) for the two most active genes in the system, obtained by stochastic simulation over $10^8$ time-units. The negative logarithm of this probability may be interpreted as an epigenetic landscape (62,64) with states that to varying degree prefer to be in the corners. The simulation uses standard parameters $N = 100$, $L = 50$, $r = 1$, $h = 2$, $\beta = 0.03$ and time delay $\tau = 0$ when nothing else is specified. Notice that both smaller $L$ and a time delay lower barriers in the system, and favour transitions to the rightmost corner where two genes are active simultaneously.
Fig. S 2: Related to Fig. 4 in main text. Robustness to changes in number of subsystems, noise and hill coefficient. A $\mu$ dependence of success probability for a system of $N = 500$ genes, with $L = 50$ nucleosomes and $r = 1$. Remaining parameters are as indicated in corresponding colours. B-G Success probability for a system of $N = 100$ genes as function of the values of $r$ and $\mu$ calculated from 20 simulations. The simulations use standard parameters $L = 50$, $r = 1$ and time delay $\tau = 0$ with $h$ and $\beta$ as specified for each plot.
Fig. S 3: Related to Fig. 5 in main text. Investigation of time-delay for standard model with $N = 100$, $L = 50$, $\beta = 0.03$, $h = 2$, $r = 1$ and $N_{pseudo} = 40$ pseudogenes. A Success probability with and without time-delay. The orange area is standard conditions whereas blue curve is same simulation but only requiring that acceptance conditions are fulfilled after time 1000. The filled circles represent success probability with two different time delays $\tau$. B-E Example of trajectories without (left) and with time-delay (right), and for 2 different values of bias $\mu$. F Number of active pseudogenes in units of number of active normal genes. Pseudo-genes does not contribute to the negative feedback, and can more easily be turned on, cyan curve. Red curve show that with time delay, then pseudogenes are turned on exactly as much as real genes as indicated by gray area. Orange curve is same simulation as in cyan, but only counting turn-on frequency for cells that successfully turn on one and only one real gene in the simulation. Therefore measurement of pseudogene turn on rates will constrain our guess on $\mu$ and $\tau$. 
as olfactory genes that do not contribute to the negative feedback. One sees
that pseudogenes do not always become activated with same probability as
normal genes.

**S3 - Robustness to model variations**

Our model is envisioned as a compressed version of a more elaborate
but unknown sequence of recruitment events that stretch from nucleosomes
with marks permissive for transcription, to different marks that prevent
transcription. We have examined some other variants of such compressed
models, including different ways to implement the overall negative feedback
\[ P = r \sum a^h_j \]
Here we show four variant models, illustrated in top panels of
Figs S4-S7. Below we outline the stochastic rules for each of the four models,
emphasizing with **bold face letters** the part that differ from the standard
model.

*Stochastic implementation of variant model in Fig. S4:*

Here we examine a recruitment from silenced state that is sensitive to
number of silenced nucleosomes exposed to the negative feedback. Consider
a system of \( N \) genes each covered by \( L \) nucleosomes. Let at each time
\[ R = \frac{1}{1 + r \cdot \sum a^h_i} \]
where \( a_i = A_i/L \) with \( A_i \) being the number of active
nucleosomes in gene \( i \). At each step of the simulation one

1. Select 2 random nucleosomes in one random gene. If both are active,
one with probability \( R \) select another nucleosome in this gene and set
its state to active. If the two nucleosomes are in the silenced state one
with **probability** \( \mu(1-R)^2 \) select another nucleosome and set its state
to silenced.

2. Select with probability \( \beta \) a nucleosome among all genes in the system.
If the chosen nucleosome is active, then change it to silenced. If the
chosen nucleosome is silenced then change it with probability \( R \) to
active.

*Stochastic implementation of variant model in Fig. S5:*

Here we examine a noise that also acts on silenced nucleosomes exposed
to the negative feedback. Consider a system of \( N \) genes each covered by \( L \)
nucleosomes. Let at each time
\[ R = \frac{1}{1 + r \cdot \sum a^h_i} \]
where \( a_i = A_i/L \) with \( A_i \) being the number of active nucleosomes in gene \( i \). At each step of the
simulation one
Fig. S 4: Related to Fig. 1 and 4 in main text. Parameter sensitivity. A Interactions in nucleosome model where negative feedback both limits the recruitment from the active state and increases recruitment from silenced state. B Probability that one and only one gene becomes active within 5000 time units as function of asymmetry $\mu$. Parameters of the simulations were as indicated in corresponding colour, except $L = 50$, $r = 1$ and $h = 2$ for all simulations Gray area marks the cut-off at 50% successful simulations. C Orange marks largest number of active nucleosomes within one gene during a 5000 time unit simulation. When no genes reach full activity during this time-interval, no olfactory receptors have turned on. Cyan and dark cyan shows maximal activity of the second and third most active genes, respectively. Where the second most active gene have many active nucleosomes, the two genes have shown simultaneous activity. D Average time for the first gene to become active, with the restriction of a 5000 time unit simulation time. Gray area marks cut-off where average switch on time is below 1000 time.
Fig. S 5: Related to Fig. 1 and 4 in main text. Parameter sensitivity. A Interactions in feedback model where noise also acts on silenced nucleosomes exposed to the negative feedback. B Orange area shows probability that one and only one gene becomes active within 5000 time units as function of asymmetry $\mu$. Parameters of the simulations were as indicated in corresponding colour, except $L = 50$, $r = 1$ and $h = 2$ for all simulations Gray area marks the cut-off at 50% successful simulations. C Orange marks largest number of active nucleosomes within one gene during a 5000 time unit simulation. When no genes reach full activity during this time-interval, no olfactory receptors have turned on. Cyan and dark cyan shows maximal activity of the second and third most active genes, respectively. Where the second most active gene have many active nucleosomes, the two genes have shown simultaneous activity. D Average time for the first gene to become active, with the restriction of a 5000 time unit simulation time. Gray area marks cut-off where average switch on time is below 1000 time units.
1. Select 2 random nucleosomes in one random gene. If both are active, one with probability $R$ select another nucleosome in this gene and set its state to active. If the two nucleosomes are in the silenced state, one with probability $\mu$ select another nucleosome and set its state to silenced.

2. Select with probability $\beta$ a nucleosome among all genes in the system. If the chosen nucleosome is active, then change it to silenced. If the chosen nucleosome is silenced then change it to active.

Stochastic implementation of variant model in Fig. S6:

Here we examine a true 3-state model, with nucleosomes in fully silenced state, neutral silenced state or active state, similar to the model examined by Dodd et al. in 2007 & Sneppen et al. in 2008. This model includes cooperativity implicitly, and not explicitly. Furthermore the model explicitly simulates the negative feedback $P$ acting on the transition between the neutral and the fully silenced state. Again we consider a system of $N$ genes each covered by $L$ nucleosomes. Let at each time $P = r \cdot \sum_i a_i$ where $a_i = A_i/L$ with $A_i$ being the number of active nucleosomes in gene $i$. At each step of the simulation we:

1. Select one random nucleosome $j$ in one random gene. If it is active, one select another nucleosome in this gene and convert it in the direction of active state. If the nucleosome $j$ is in the silenced state one with probability $\mu$ select another nucleosome $k$. If $k$ is active its state is reset to neutral. If $k$ is neutral its state is set to be silenced.

2. In this separate step we simulating the globally acting negative feedback: a) Select a nucleosome $P$ times, and for each selection where this nucleosome is neutral, then convert its state to silenced. If $P < 1$ one with probability $P$ select a nucleosome and if it is neutral, then change it to become silenced. b) Select a nucleosome, and if it silenced, then change it to become neutral.

3. Select with probability $\beta$ a nucleosome among all genes in the system. If the chosen nucleosome is active, then change it to neutral state. If the chosen nucleosome is neutral then with equal probability set its new state to be either active or
silenced. If the chosen nucleosome is silenced, then change it to neutral.

Stochastic implementation of variant model from Fig. S7:

Here we examine a model where recruitment activity from the active state is directly modulated by the negative feedback. As two nucleosomes are chosen as recruiting, this model effectively “squares” the effect of the negative feedback. To implement the corresponding model we again consider a system of $N$ genes each covered by $L$ nucleosomes. Let at each time $R = 1/(1 + r \cdot \sum_i a_i^h)$ where $a_i = A_i/L$ with $A_i$ being the number of active nucleosomes in gene $i$. At each step of the simulation one

1. Select 2 random nucleosomes in one random gene. If both are active (A), one with probability $R^2$ select another nucleosome in this gene and set its state to active. If the two nucleosomes are in the silenced state (S) one with probability $\mu$ select another nucleosome and set its state to silenced.

2. Select with probability $\beta$ a nucleosome among all genes in the system. If the chosen nucleosome is active, then with probability $R$ change it to silenced. If the chosen nucleosome is silenced then change it with probability 1 to active.
Fig. S 6: related to Fig. 1 and 4 in main text. Parameter sensitivity. A Interactions in 3 state version of model, reflecting the Schizosaccharomyces pombe model suggested by (40,41). B Orange area shows probability that one and only one gene becomes active within 5000 time units as function of asymmetry \( \mu \). Parameters of the simulations were as indicated in corresponding colour, except \( L = 50, r = 1 \) and \( h = 2 \) for all simulations Gray area marks the cut-off at 50% successful simulations. C Orange marks largest number of active nucleosomes within one gene during a 5000 time unit simulation. When no genes reach full activity during this time-interval, no olfactory receptors have turned on. Cyan and dark cyan shows maximal activity of the second and third most active genes, respectively. Where the second most active gene have many active nucleosomes, the two genes have shown simultaneous activity. D Average time for the first gene to become active, with the restriction of a 5000 time unit simulation time. Gray area marks cut-off where average switch on time is below 1000 time units.
Fig. S 7: Related to Fig. 1 and 4 in main text. Parameter sensitivity. 

A Interaction in model where we now revert the effect of the negative feedback, such that it acts to decrease the fraction of active nucleosomes that can recruit from the silenced state. Orange area shows probability that one and only one gene becomes active within 5000 time units as function of asymmetry $\mu$. Parameters of the simulations were as indicated in corresponding colour, except $L = 50$, $r = 1$ and $h = 2$ for all simulations. Gray area marks the cut-off at 50% successful simulations.

B Orange marks largest number of active nucleosomes within one gene during a 5000 time unit simulation. When no genes reach full activity during this time-interval, no olfactory receptors have turned on. Cyan and dark cyan shows maximal activity of the second and third most active genes, respectively. Where the second most active gene have many active nucleosomes, the two genes have shown simultaneous activity.

D Average time for the first gene to become active, with the restriction of a 5000 time unit simulation time. Gray area marks cut-off where average switch on time is below 1000 time units.