

INAUGURAL DISSERTATION

zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen Gesamtfakultät
der
Ruprecht-Karls-Universität Heidelberg

vorlegt von
Anashua Banerji, M. Sc.
aus Kolkata, India

Date of Oral Examination:

Thema

**A quantitative analysis of hepcidin promoter
regulation using mathematical modelling
techniques to reveal principles underlying
systemic iron homeostasis**

Gutachter 1: Prof. Dr. Roland Eils

Gutachter 2: Prof. Dr. Hans George Bock

*Anashua Banerji
University of Heidelberg
Doctoral Thesis
2013*

Contents

<i>Acknowledgement</i>	8
<i>Abstract</i>	12
<i>Kurzfassung</i>	13
<i>Glossary</i>	15
I. Introduction	20
I.1 A brief introduction to Systems Biology and modelling	20
<i>I.1.1 Purpose and outlook of Systems biology</i>	20
<i>I.1.2 Approaches to systems biology</i>	21
<i>I.1.3 Introduction to modelling techniques</i>	22
<i>I.1.4 Thermodynamic Model</i>	25
<i>I.1.5 Differential Equation Models</i>	28
<i>I.1.6 Boolean Models</i>	32
<i>I.1.7 Summing up the reasoning</i>	34
I.2 Hepcidin	36
<i>I.2.1 Hepcidin and regulation of systemic iron homeostasis</i>	36
<i>I.2.2 Hepcidin Control pathways</i>	40
<i>I.2.3 Hepcidin Regulation with focus on modelling</i>	40
<i>I.2.4 Hepcidin malfunction and related disease- Motivation for modelling</i>	44

II. Results: Towards modelling hepcidin expression	47
II.1 Conceptual model of systemic iron homeostasis	47
II.2 Quantification of hepcidin cross-regulation by BMP and IL6.....	55
<i>II.2.1 Experimental characterization of hepcidin promoter variants</i>	
<i>- performed by Muckenthaller group</i>	<i>55</i>
<i>II.2.2 Theoretical analysis of promoter data</i>	<i>57</i>
II.3 Modelling Regulation of Hepcidin.....	66
<i>II.3.1 Definition of model assumptions</i>	<i>66</i>
<i>a) Modelling dose-response behaviour of signalling module</i>	
<i>- without assuming signalling crosstalk</i>	<i>67</i>
<i>b) Promoter Module</i>	<i>69</i>
<i>c) Gene-expression module</i>	<i>71</i>
<i>II.3.2 Thermodynamic Model for Hepcidin Promoter.....</i>	<i>75</i>
<i>II.3.3 Model Calibration</i>	<i>93</i>
III. Discussion and analysis of model and approach.....	101
III.1 Mathematical modelling of the IL6 and BMP crosstalk at the	
signalling and promoter levels	101

III.2 Experimental validation of model-predicted mechanisms of hepcidin promoter regulation	110
III.3 Analysis of signal integration in the promoter model	120
III.4 Description of detailed homeostasis model.....	125
IV. Outlook and scope of work done	131
IV.1 Relation and context to previous studies on hepcidin promoter regulation	131
IV.2 Signal integration and implication of the hepcidin promoter description.....	133
IV.3 Limitations of current modelling approach	135
IV.4 Extension of the proposed modelling approach.....	137
IV.5 Implications for disease and therapy	139
Appendix.....	142
I. Equations list for easy reference.....	142
II. Figures list for easy reference.....	157
Reference.....	164

Acknowledgement

I am grateful to Dr. Stefan Legewie for his excellent supervision and for initiating me patiently to an area of scientific research that has been completely new to me. I would like to extend my heartiest thanks to him for his consistent encouragement and concern.

I would like to thank Prof. Dr. Roland Eils for sharing our enthusiasm for this project and supervising my overall progress with the research. I am very grateful to him for providing the excellent facilities and the vibrant and conducive environment for motivating scientific research.

Here I would also take the opportunity to thank the other members of my supervisory committee: Prof. Dr. Hans George Bock, Prof. Dr. Rainer Dahlhaus and Prof. Dr. Gerhard Reinelt for their valuable time and inputs. Thanks also to their respective secretaries Frau Margret Rothfuss, Frau Elke Carlow and Frau Catherine Proux-Wieland for their considerate handling of my changing schedule and for their ever-prompt replies. Thanks also to the Secretary of the Mathematics Faculty, Frau Dorothea Heukaeufer for her patient and efficient handling of the submission process.

Heartiest thanks to my colleagues Stefan Kallenberger, for his valuable comments and help, and Avik Sanyal for his excellent proof-reading and suggestions.

I owe heartiest thanks to the IT department at Bioquant for addressing the technical troubles that plague a theoretician's life quite regularly and for always coming up with quick solutions.

Finally and very importantly, I would like to thank the ladies who make up Roland's Secretariat: Frau Manuela Schaefer, Frau Corinna Sprengart and Frau Sabrina Wetzel. As

with everyone, they have been a wonderful support system for this foreign PhD student, helping me with academic and official things as well as making my stay in Heidelberg a more beautiful experience. Their method of working with kindness and efficiency has also been a part of the training I have received through my PhD work.

All future success owed to this thesis stands dedicated to my grandmother Mrs. Maya Biswas and to the memory of my grand-father Mr. B.N Biswas whose presence remains eternal...

Abstract

Since the development of advanced mathematical modelling techniques in biology, thermodynamics (and therefore equilibrium statistical mechanics) has played a key role in mathematically quantifying biological activities. We use this underlying notion of thermodynamic ‘micro-states’ to attempt to uncover how the hormone hepcidin under the influence of two major signalling pathways maintains systemic iron homeostasis. Systemic iron homeostasis involves a negative feedback circuit in which the expression level of the peptide hormone hepcidin depends on and controls the iron blood levels. Hepcidin expression is regulated by the BMP6/SMAD and IL6/STAT signalling cascades. Deregulation of either pathway causes iron storage diseases such as hemochromatosis or anaemia of inflammation (AI). We quantitatively analyzed how BMP and IL6 control hepcidin expression in human hepatoma (HuH7) cells. We used data from our experimental collaborators who measured transcription factor phosphorylation and reporter gene expression under co-stimulation conditions and perturbed the promoter by mutagenesis. We applied statistical data analysis and mathematical modelling to reveal possible biological mechanisms that control hepcidin expression at the promoter level. Specifically we develop a thermodynamic modelling framework that is able to simulate and predict possible molecular mechanisms that might underlie iron homeostasis by hepcidin. Our results reveal that hepcidin cross-regulation primarily occurs by combinatorial transcription factor binding to the promoter, whereas signalling crosstalk is insignificant. We find that the presence of two BMP-responsive elements in the promoter ensures high sensitivity towards the iron-sensing BMP signalling axis, which promotes iron homeostasis in vivo. IL6 stimulation reduces the promoter sensitivity to the BMP signal that may explain the disturbance of iron homeostasis in AI. We get to understand why the iron homeostasis circuit is sensitive to certain perturbations implicated in disease. Taken together, our work reveals how mathematical quantification and modelling can aid in understanding biological phenomenon that underlies gene expression.

Kurzfassung

Durch die Entwicklung fortgeschrittener mathematischer Modellierungstechniken in der Biologie nimmt die Thermodynamik und damit die statistische Mechanik von Gleichgewichtsprozessen eine bedeutsame Rolle bei der quantitativ mathematischen Beschreibung biologischer Systeme ein. Wir verwendeten den Grundgedanken thermodynamischer Mikrozustände, um zu verstehen, wie das Hormon Hepcidin unter dem Einfluss zweier grundlegender Signaltransduktionswege die Eisenhomöostase reguliert. Die systemische Eisenhomöostase unterliegt einer negativen Rückkopplung, bei der das Expressionsniveau des Peptidhormons Hepcidin von der Eisenkonzentration im Blut gesteuert wird und letztere gleichzeitig kontrolliert. Die Expression von Hepcidin wird durch die BMP6/SMAD- und IL6/STAT-Signalwege gesteuert. Die Dysregulation eines der beiden Signalwege verursacht Eisenspeichererkrankungen wie Hämochromatose oder entzündungsbedingte Anämie. Wir untersuchten quantitativ, in welchem Maße BMP und IL6 die Expression von Hepcidin in humanen Hepatomzellen (HuH7) kontrollieren. Die verwendeten Daten stammen von experimentellen Kooperationspartnern, durch welche die Phosphorylierung von Transkriptionsfaktoren sowie die Expression von Reporter genen unter Kostimulation gemessen und eine Promotermutagenese durchgeführt wurde. Wir verwendeten statistische Methoden und mathematische Modelle, um mögliche biologische Kontrollmechanismen der Hepcidinexpression auf Promoterebene aufzudecken. Insbesondere entwickelten wir eine Modellierungsumgebung für die Simulation und Vorhersage möglicher molekularer Mechanismen der Eisenhomöostase durch Hepcidin. Wir konnten zeigen, dass die rückgekoppelte Regulation von Hepcidin vorrangig durch kombinatorisches Binden von Transkriptionsfaktoren an Promoter stattfindet, wobei wechselseitige Abhängigkeiten irrelevant sind. Wir fanden heraus, dass das Vorliegen zweier BMP-abhängiger Promoterelemente für die hohe Sensitivität gegenüber des eisensensitiven BMP Signalwegs verantwortlich sind, wodurch die Eisenhomöostase in vivo gewährleistet wird. Stimulation durch IL6 reduziert die Promotersensitivität gegenüber dem BMP-Signal, wodurch die

gestörte Eisenhomöostase bei der entzündungsbedingten Anämie erklärt werden kann. Wir konnten in Erfahrung bringen, weshalb der Regelkreis der Eisenhomöostase sensitiv gegenüber bestimmten krankheitsbedingten Beeinträchtigungen ist. In Zusammenschau unserer Ergebnisse zeigt sich, wie durch quantitativ mathematische Beschreibung und Modellierung ein tieferes Verständnis biologischer Phänomene der Genexpression erlangt werden kann.

Glossary:

Actin: A globular multi-functional protein that participates in cell-signalling along with many other functions

BMP: Bone morphogenic protein (a group of growth factors with the ability to induce formation of bone and cartilage)

BRE1: First BMP responsive element (portion of the hepcidin promoter [in this case an area near the transcription start site or “proximal”] that is essential for response to BMP stimulation)

BRE2: Second BMP responsive element (portion of the hepcidin promoter [in this case an area farther from the transcription start site or “distal”] that is essential for response to BMP stimulation)

HH: Hereditary hemochromatosis (a genetic disorder characterized by excessive internal absorption of dietary iron resulting in pathological increase in body’s iron store)

HJV: Hemojuvelin (a membrane bound protein in mammals that is responsible for the severe iron overload condition or hemochromatosis in humans)

HFE: the HFE gene codes the Human hemochromatosis protein

HuH7: an immortal cell-line derived from cancer-infected liver cells

IL6: Interleukin 6 (small signalling molecules released by white blood cells and act as response to infection)

kb: Kilo-base (Base pairs are building blocks of the DNA double helix and contribute to the folding of DNA and RNA. The size of an individual gene is measured in base pairs since the DNA is double-stranded. Total number of base pairs is therefore equal to the total number of nucleotides in one of the strands. A Kilobase (kb) is a unit of measurement equal to 1000 base pairs of DNA or RNA.)

Specific binding: Binding to the receptor of interest. (Binding to other receptors is known as **non-specific binding**)

pSMAD/pSTAT: phosphorylated proteins [*see SMAD/STAT for explanation*] (in the context of this work, the purpose of phosphorylation is to activate and increase energy of the proteins so that they can take part in initiating gene-expression)

RNAP: RNA (Ribonucleic Acid) Polymerase (an enzyme that produces RNA, necessary in cells for constructing RNA chains using DNA genes as templates in transcription)

SMAD: Intracellular proteins that transduce/transfer signals from growth factors (like BMP) when they activate gene transcription

STAT: Intracellular proteins that transduce/transfer signals from cytokines (like IL6) and growth factors (like BMP)

STATBS: STAT binding site (special region in the hepcidin promoter that selectively binds to phosphorylated STAT proteins)

TFR2: Transferrin receptor protein 2 (this protein mediates cellular uptake of of transferrin bound iron and mutations in the gene TFR2 have been associated with hereditary hemochromatosis)

TGF- β : Transforming growth factor beta is a protein that controls cellular differentiation and other functions in most cells and acts majorly through the SMAD pathway

TSS: Transcription start site is the location on the genetic sequence where transcription is initiated

WT: Wild type (phenotype of the typical form of a species as it occurs in nature and considered as the standard for comparison in experiments)

**A quantitative analysis of hepcidin promoter regulation
using mathematical modelling techniques to reveal
principles underlying systemic iron homeostasis**

* Unless otherwise mentioned, all work addressed as '*we*' has been performed by Anashua Banerji in direct discussion with Dr. Stefan Legewie

** Parts of this thesis has been submitted for publication in Nucleic Acids Research and is under review

I. Introduction

I.1 A brief introduction to Systems Biology and modelling

The 21st century has seen a rapid development of technology due to which scientific research is witnessing an exponential increase in the amount of information as raw data. As with other scientific disciplines, such development poses an important challenge to cell biology which is to facilitate integration of the available molecular information for improving quantitative understanding of biological units/structures at the systems level. This is where and why systems biology comes in with its typically cyclical methodology of efficient experimentation validated and updated by evolving modelling approaches.

I.1.1 Purpose and outlook of Systems biology

Systems Biology aims to understand how interaction between components of a living system brings about its functional properties and characteristic behaviour. As a field of study, Systems Biology is unique as its foundations lie in existing and evolving technologies and purports to analyse scientific issues that are described by a variety of existing disciplines yet remain unsolved. Its main focus is to quantitatively and predictably reveal biological processes that underlie regulation of complex cellular pathways or intercellular communication [2, 3]. It is believed by scientists that the existing paradigms for biomedical research fail for a multitude of human diseases due to the lack of achieving this understanding of living systems as a unified entity rather than a conglomerate of isolated systems. Systems

biology aims to provide a suitable approach to study these complex and multifactor diseases. Contemporary systems biology is a vigorous and expanding discipline, in many ways a successor to molecular biology and genomics on the one hand and mathematical biology and biophysics on the other [4]. It is typically an interdisciplinary field as it lies at the interface of many different scientific disciplines. However in effect it may also be seen as a 21st century culmination of all the biological research that science has witnessed till now. Systems biology aims to understand how functional properties of life and living matter might arise out of a particular organization of its molecular components or interactions between them. Models are used to describe cell types or regulation mechanisms in order to arrive at new hypotheses about the biological system. This methodology reveals approaches that cannot be explained exclusively by traditional natural sciences like biology or physics [5]. Alongside, development of the various high-throughput technologies used in genome sequencing, transcriptomics, proteomics and metabolomics have enabled comprehensive analysis of complete living systems in terms of the identity and concentration of all their components [6]. However, since these technologies do not take into account the interactions and biological processes at the molecular level, by themselves they are incapable of providing an understanding of the living cell and hence life. Systems biology, therefore, provides scientists with the techniques to decipher how biological functions arise from the interactions between components of living organisms [7-9].

1.1.2 Approaches to systems biology

Systems biology is characterized by two established approaches: Top-down and bottom-up [8]. Top-down systems biology begins with experimental data based on the behaviour of specific components in the living system as a whole. This is mostly done using high-throughput approaches to measure types and levels of (macro)molecules in the cell on a large

scale [6]. In the analysis of such data, new hypotheses on the molecular organization and functioning of the organism may be induced on the basis of correlations in the behaviour of the concentration of the molecules [10]. In contrast, the bottom-up approach of systems biology starts from the interactive properties of the molecules and proceeds to determine how these interactions may lead to functional behaviour. Typically, a biological network forms the basis of this approach. A network is a system of sub-units that link the parts to the whole and biological networks are essential in providing a ground for mathematical analysis of how the constituents are connected. Most important biological networks involve building-blocks like the DNA, RNA, proteins and metabolites and follows up interactions between them. For example, a gene regulatory network is a collection of DNA segments in a cell that can be activated or suppressed thereby controlling production of particular proteins at a given time. Such regulation can be triggered by external signals or regulatory proteins. In many cases these proteins serve to activate other genes and are called transcription factors, which are the key-players in bringing about gene-expression. Metabolic networks on the other hand represent the complete extent of metabolic and physical processes determining physiological and biochemical properties of a cell. Another form of networks is formed by the signals that are transduced within cells or between cells thereby forming complex signalling networks. At the other end, neuronal networks connect neurons (nerve cells) in the brain the activation of which defines a recognizable linear pathway.

1.1.3 Introduction to modelling techniques

One of the major goals of systems biology is to create comprehensive models that would be able to predict cellular behaviours. The prevalence of expanding genomic information and high-throughput experimental techniques with a focus on systems biology has lead to the predominance of mathematical modelling of biological systems. Gene regulation modelling is

especially of great importance because gene expression is connected to a variety of biological processes, so much so that subtle changes of regulatory protein levels or links can underlie human diseases, population differences, and the evolution of morphological novelties [11].

Till recently, the study of mechanistic gene regulation has been empirically focused on the characteristics of individual transcriptional components, rather than creating an integrated picture of the system. Of late, studies in systems biology are being substantiated by large-scale biological datasets. An example of such detailed data is complete sequences of genomes that identify many proteins and RNAs influencing the regulatory processes in the nucleus. Additionally these also provide dynamic measurements of expression levels for many genes. However, though such data are available extensively our understanding still remains relatively incomplete. This is mainly because most of the available data normally provide selective view of dynamics of the systems or averages over many cell-states. Therefore, using exclusively experimental approaches to obtain a complete operational picture is a challenge.

Mathematical modelling addresses this problem by presenting approaches that incorporate details of the dynamics of biological systems in model construction. This necessarily requires an assimilation of inputs from all levels and integration of experimental, computational, and theoretical approaches. Molecular interactions need to be precisely described in a series of mathematical formulations that take into account the necessary parameters, such as initial concentration of each component and kinetic constants, which are normally estimated from experimental data. Gene regulatory models mostly employ analytical or statistical approaches. Both of these approaches can effectively provide non-intuitive insights into gene regulatory systems. The statistical approach is appropriate for datasets representing the expression levels of thousands of genes. To represent regulatory interactions, probabilistic models that are graph-based such as neural, Boolean, and Bayesian networks are used [12, 13]. Varied

conditions bring in statistical correlations that highlight which groups of genes function together thereby indicating possible regulatory relations. This analysis identifies shared motifs that might be involved in transcriptional control. The final aim of this approach is to uncover regulatory networks that possibly underlie the given data. However, though such statistical approaches provide a big picture based on large proportions of genes in a system, they are not really able to explain complex relations between polymerase, transcription factors or the details of the cellular architecture. Development of gene array data has led to better-evolved statistical approaches and some have been discussed in recent reviews [13-17].

The analytical approach is represented by a variety of mathematical models focusing on gene-expression. These models incorporate terms that define binding of RNA-polymerase and transcription factors to the DNA, cooperative and/or inhibitory interactions between transcription factors and mRNA, protein degradation and mRNA translation rate. The analytical approach is based on extensive knowledge of components in the system and on hypotheses about the structure of the system. As covered in the review by Ay & Arnosti (2011) there are three major classes of mathematical models that are useful for us in the context of this thesis and characterize the analytical approach to systems biology: thermodynamic, Boolean, and differential equation-based models [18]. In existing literature we can find these models being used to summarize experimental data [19], to infer new relations from complex experimental data thereby presenting new testable hypotheses [20] and to find properties of the system that are challenging to measure directly but lead to accurate modelling of emergent behaviour [21]. Some general features characterize these models. In most cases the models are deterministic such that the independent variable has a reproducible and predictable impact on the dependent variable. However, they can also be constructed as stochastic models that can capture any erratic behaviour that is observable in many biological systems due to extrinsic or intrinsic noise [15, 22, 23].

At the next level, modelling approaches might be categorized as discrete or continuous. Boolean models are an example of the discrete form and represent time, state or space as a set of discrete values thereby simplifying calculations. On the other hand, differential equation models utilize continuous values to facilitate (mathematically) “smoother” representation of dynamical changes. In the following sections we provide an overview of three major classes of models. As in the case of our work on hepcidin modelling, the choice of model depends entirely on the biological system under consideration. A successful model should fit existing experimental data providing alongside new insights on the system under consideration and be predictive. We hope such a description would provide the reader with a baseline understanding of modelling approaches in systems biology and an introductory reasoning regarding the use of thermodynamic modelling in the framework of our project.

1.1.4 Thermodynamic Model

Approach

This modelling approach efficiently extracts gene regulation information from sequences of cis-regulatory regions (cis-element is a region of DNA/ RNA that regulates the expression of genes located on that same molecule of DNA) and known or hypothesized binding of elements to sequence-specific transcription factors. These models aim to predict how and whether a gene will be activated or repressed given that the relevant transcription factors and promoters are well explained for the system. Based on the underlying tenet that gene expression is (inversely) directly proportional to the level of (repressors) activators bound to the promoter, these models can also predict how varied spatial and temporal expressions can

result from various combinations of binding sites present on a regulatory region.

Thermodynamic (also known as fractional occupancy) models are based on simple biophysical descriptions of DNA–protein interactions and statistical physics (*Fig. 1A*) [18]. The first column shows a simplified enhancer region with two binding sites for a repressor (R) and an activator (A). The mathematical formulation represents binding efficiency for the repressor site. In the second column, all four possible states of this enhancer region are shown. The third column represents the probability of this state occurring, which is not simply one-fourth but rather a function of the protein concentration and quality of the binding site(s). The fourth column indicates the efficiencies with which a particular state drives gene expression. This may be a simple additive expression of activators minus repressors, or a more complex expression. The last column represents the total expressions coming from each state (the probability that a state will occur multiplied by the potential of this configuration of proteins) and their summation, which provides a measure of the total output of the cis-element. Generally implementations ignore additional processes like DNA methylation or chromatin structure and modification and do not independently account for the recruitment of cofactors or general transcription machinery, although some aspects may be incorporated in the later sequence of models. Such a simplification allows for more comprehensibility of the directed application of this modelling approach.

There are two main ways to implement these models. To start with, potential transcription factor-DNA interactions that illustrate all the variant states of the enhancer are listed and each state has a statistical weight is assigned to it (*Fig. 1A*). States where activators are bound to specific binding-sites are considered to be active. For example, a simple regulatory region with only one binding site will have two states (2^1 states: simply bound and unbound) and one with four sites would have 16 states (2^4). Calculating statistical weight takes into account concentration and binding affinity of transcription factors. Proteins binding with high affinity have greater statistical weights than where binding is weak. Each state has a probability of

occurrence that is given by dividing the statistical weight of the particular state by the sum of statistical weights of all possible states. Calculation of this probability incorporates various properties that are known to affect rate of transcription. As an example, inhibitory effects of repressors on activators or competitive or cooperative interactions between transcription factors may be included in an existing model by assigning higher or lower weights. At the next stage gene expression output is calculated from each of the states. High expression is likely to be induced from the states having high activator occupancy whereas low expression levels mostly characterize repressor occupancy. Various approaches have been used with the aim of converting occupancy to gene expression. As is the underlying tenet of thermodynamic modelling, gene expression output can be modeled as proportional to the weighted sum of the transcription factors or the probability of RNA-polymerase binding to the transcription start site [24, 25].

Applications

The theoretical foundations of thermodynamic modelling have been studied majorly using prokaryotic systems. Simple bacterial systems provide an extremely comprehensible setting for quantitative research since the regulatory regions contained by them is generally small and bind to very few transcription factors. The lac operon in *Escherichia coli* and the lysis/lysogeny switch of phage lambda are two examples that have been treated ([1, 26]. [27]). Additional promoters and configurations are considered in Bintu et al.[24]. Zhou and Su generalized the results of Bintu et al. to derive a single formula calculating transcriptional probability for all simple regulatory configurations [27]. In eukaryotic systems, thermodynamic modelling shows great potential to predict the function of different combinations of transcription factor binding sites since they are able to efficiently incorporate complex *cis*-regulatory regions [27].

The possible effectiveness and limitations of this approach are illustrated by the recent uses of thermodynamic modelling in yeast and *Drosophila* [28-30]. With a wide range of applications, thermodynamic modelling has also aided in discovery of functions of single, complex regulatory regions in great detail. Reinitz and colleagues modeled the activity of a 1.7-kb promoter proximal region of the *Drosophila melanogaster* even-skipped (*eve*) gene that expresses as seven stripes in the stage of embryonic development [31]. This particular study was of specific importance since it concluded that enhancer-like outputs maybe generated by binding sites that are widely dispersed, further suggesting that developmental regulatory modules need not always exist as compact modules.

1.1.5 Differential Equation Models

Approach

In the previous section we observe that thermodynamic models are especially efficient in accounting for detailed, quasi-equilibrium activity of transcriptional elements that are well defined. At the other end of the scale, however, there are many multi-component and temporally evolving biological systems that require a model that is able to represent these properties. For such cases differential equation models are best suited. Differential equations can efficiently define regulatory networks by incorporating explicit rules defined by rate equations that embody the interactions of molecules such as mRNAs and proteins. As the system evolves, the levels of the respective protein or mRNA as a function of the other components can be specified by these equations. These models generally include space or time-dependent variables like mRNA and protein concentrations and parameters such as production and degradation rates. In Figure 1B, regulatory relationship between two genes is

depicted. Synthesis of gene 1 (G1) involves expression of mRNA (M1) and translation of protein (P1), which regulates gene 2 (G2). Both mRNA and protein are subjects to turnover and protein is subject to diffusion. mRNA and protein synthesis, degradation, and diffusion events are shown at left. This process can be modeled with reaction diffusion equations as shown at right. Each molecular constituent is assigned such an equation. There are two main groups of differential equation models: one using ordinary differential equations (ODEs) where dependence is on a single variable like time; the other group consists of application of partial differential equations (PDEs) which involves multiple variables like time and space together. Although ODEs are a well-established area of mathematics, it is hard to solve them analytically and numerical methods are applied to arrive at their solutions with the help of approximations and software applications. Similarly, though PDEs are established analytically and numerically, they are computation intensive and their theory is much more complex. Therefore even in case of PDEs, it is difficult to find analytical solutions and one must rely on numerical analysis tools.

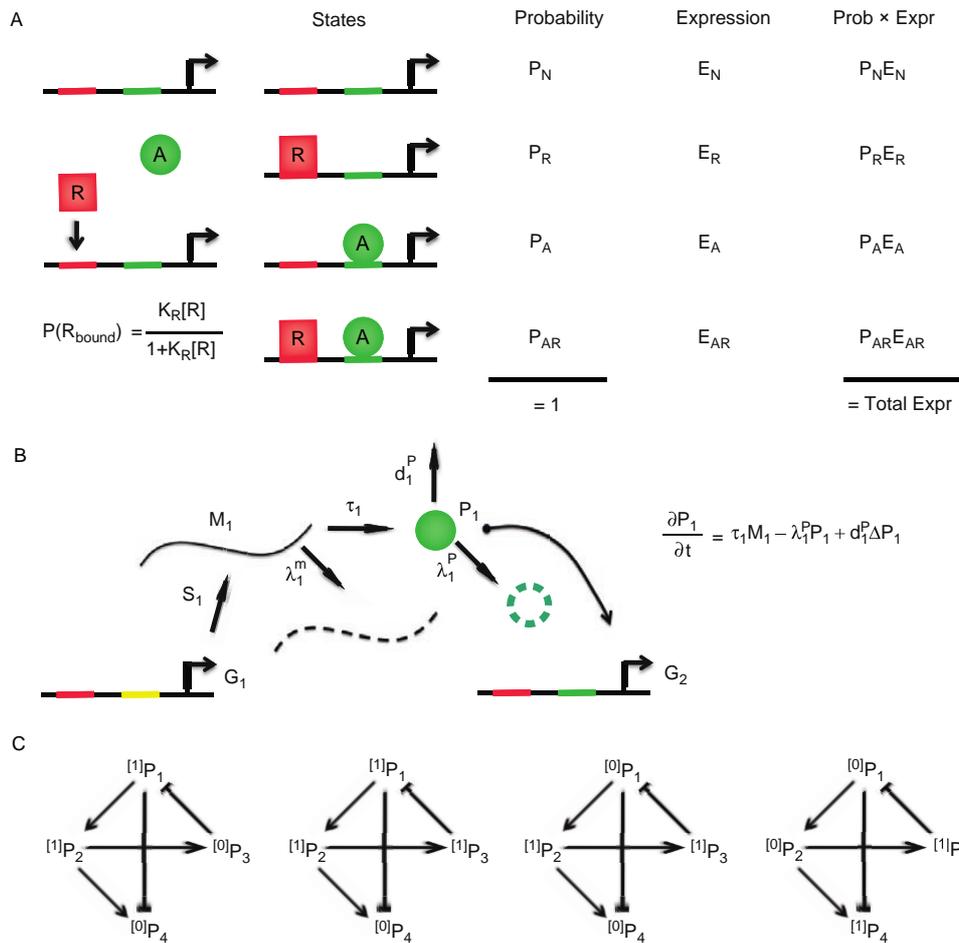


Figure 1: Analytical modelling approaches used in gene regulation studies. **(A)** Thermodynamic or fractional occupancy model of gene expression. **(B)** Differential equation model of gene expression. **(C)** Boolean model of gene expression. [18]

Application

ODEs were first applied to bacterial operons like tryptophan (*trp*) and *lac*. These operons each consist of structural genes and a small regulatory region in the DNA that controls gene expression [32-35]. Such operons have been subject to detailed experimental and quantitative studies. These models have been applied to systems with varying levels of complexity, from basic descriptions of diffusible morphogens to complex gene regulatory networks showcasing cell signalling. Features of differential equations are uniquely suited to reflect the dynamics of biological systems. These models, however, do have their limitations. They work best only in well-characterized systems with adequate quality and quantity of data. The important effects of addition of new proteins might even be missed due to overfitting of an incomplete model [36].

Significant improvement in results was observed by Reinitz and colleagues by simply improving data quality keeping the modelling and optimization techniques same as earlier efforts. This also showed much lower levels of error and provided more precision for parameter estimates [20, 31]. Larger regulatory networks are broken by parts into smaller modules for applying these models and this method keeps the analysis tractable. Extensive data often requires a large number of parameters and this might make modelling computationally challenging. Therefore it is often difficult to analyze a complex regulatory system that has numerous between-molecules interactions. Statistical methods along with advancement in computational techniques are expected to eventually ameliorate this problem [12, 13].

Another limiting factor is that differential equation models generally fail to consider translational regulation, sequence of transcriptional cis-regulatory elements or other such very fine-scale phenomena. Therefore these models are not able to provide much insight to the structure of the enhancer or that of its organization in a way provided by thermodynamic models. In effect ODE models are able to occupy the middle ground in the world of transcriptional modelling and provide an option of analysing moderate to high complexity biological systems by efficiently describing dynamical aspects that are lacking in other approaches like thermodynamic modelling.

1.1.6 Boolean Models

Approach

Some biological processes show on-off behaviour. Logic gates are building blocks of Boolean models and can represent such characteristics. In biological systems proteins or mRNAs normally have two states on or (1) and off or (0). In *Figure 1C*, a network describing the regulatory relationships among four proteins is shown; the directed arrows show activation, and the blunt arrows show repression. Starting from initial state, three temporal steps are demonstrated. In this model, protein turnover occurs in one time interval and repression is assumed to be dominant over activation. Here, superscripts [1] and [0] indicate active or inactive states, respectively. Interaction between these elements can be described by using logic gates like 'OR', 'AND' and 'NOT'. To illustrate the nature of these interactions, when there are two transcription factors, the 'OR' function indicates that transcription takes place if any one of them are bound (active), 'AND' implies that for transcription both need to be active and 'NOT' means there is no gene transcription if both transcription factors bind to

their respective sites. Boolean modelling can be efficiently used to combine qualitative experimental results logically in order to simulate the dynamic behaviour in biological frameworks with pre-defined theoretical components and interactions. The great advantage of Boolean models is that they can be easily analyzed through computational implementation and can be scaled from a smaller to a larger system. These models can efficiently create and analyze variants of the same network and might be used even in systems without detailed network description. They are simple yet efficient in illustrating basic characteristics underlying the biological system.

Application

There is extensive literature where Boolean measures have been applied to analyze gene regulation [19, 37]. Sánchez and Thieffry have studied the *Drosophila* gap gene network by applying Boolean models followed up later by Jaeger and colleagues who carried out the study using reaction diffusion models [20]. In this case, both differential equation-based models and Boolean models provide equable understanding of this study yet have their characteristic differences [20, 37]. Protein concentrations that are continuous in nature are discretized in the Boolean set-up and though this eases computation, such conversion of the data prevents detailed quantitative analysis of gene gap network features. Generally speaking, application of Boolean models sometimes requires a simplification of set up which in turn might heavily tax the accuracy of results. However these still provide a successful starting point for preliminary investigative analysis.

1.1.7 Summing up the reasoning

The project covered by this thesis might be seen as a case study for modelling gene-expression data based on the bottom-up approach. The experimental data we have is quantitative and provides details at the promoter level. We note that of all modelling approaches thermodynamic modelling fits the bill best since this modelling framework allows for a quantitative description of combinatorial promoter regulation. Comparatively, we see that applying ODEs would not work since this approach would neglect the details on promoters that are available in the data whereas Boolean modelling would be a failure as well due to its qualitative approach when the need is to quantify a rather complicated transcription procedure. Stochastic modelling is also not an option since we are not interested in single-cell variability and are dealing with population data. On this basis we aim to study how systemic iron homeostasis might be influenced by hepcidin expression brought about by promoter level interactions of transcription-factors and possible signalling level crosstalk of certain pathways (BMP & IL6) and protein molecules. By calculating how the model might behave *in silico* and then comparing to observations made at the system level, emergent systems properties are predicted. Even in case of model failure, we hoped that the lack of correspondence might bring more certainty to what are *not* the characteristics of the system and what *might* be hypothesized instead. As we see in the course of our investigation, correspondence with experimental data, mostly leads to the discovery of interactive or organizational properties that are important biological functions. The discovered properties are then inserted in a new generation of models, and eventually more detailed and accurate models can be obtained. As our work illustrates, the multi-level biological interactions eventually affect or effect processes that enable a living system to maintain its state or develop in time through processes that repair damage or compensate for dissipation. In some systems, for example, the molecular constituents are understood sufficiently enough to prompt the construction of detailed kinetic models of reaction networks ('silicon cells') [38-40]. As applications, these approaches might be used to design drugs *in silico* or to improve understanding of how

molecules might jointly bring about a particular cellular behaviour [41, 42]. It is also interesting to mention here that top-down and bottom-up systems biology approaches are not mutually exclusive. Ultimately these two approaches aim to be synergistic and mutually corroborative.

Given this background of systems biology and applications of various modelling approaches, the reader can better understand the way we have progressed to find the answer to the problem being addressed in this thesis. Further into the project now, we begin with an explanation of the biological background and the mechanisms explained in existing literature on hepcidin and its regulation. We then proceed with an overview of the experimental data from our collaborators followed by analysis of the data and modelling methods applied by us.

I.2 Hepcidin

1.2.1 Hepcidin and regulation of systemic iron homeostasis

Iron homeostasis is tightly regulated to provide this critical element for growth and survival while preventing the toxicity of iron excess. Intestinal absorption, reticulo- endothelial cell recycling and mobilization of hepatocyte(liver cells) stores maintain plasma iron levels. Circulating iron is secreted into serum transferrin and delivered primarily to the bone marrow for erythropoiesis (production of red-blood cells). Excess iron in the body can only be removed by sloughing of enterocytes and blood loss; the remaining iron is stored primarily in macrophages (white blood cells) and hepatocytes. There is no known regulated mechanism for iron excretion; therefore systemic iron homeostasis is maintained by tight regulation of intestinal iron absorption and macrophage and hepatocyte iron release [43]. The mechanism for iron regulation is yet to be fully understood but recent research suggests that hepcidin has a key role in regulating systemic iron homeostasis. Hepcidin is an iron-regulated hepatic peptide hormone that controls systemic iron homeostasis. Iron excess or inflammatory cytokines stimulate hepcidin expression, leading to reduction in plasma iron levels as the result of iron retention in macrophages and reduced intestinal iron absorption. Hypoxia (oxygen deficiency), high erythropoietic activity, and iron deficiency inhibit hepcidin expression by largely unknown mechanisms to mobilize iron stores and increase iron absorption. Hepcidin exerts its function by binding to the iron efflux channel ferroportin (*Fig. 2*), which is predominantly expressed on macrophages, intestinal enterocytes, and hepatocytes, causing ferroportin internalization and degradation [44]. Hepcidin thereby decreases both intestinal iron absorption and macrophage iron release. The mechanisms by which hepcidin expression is regulated remain poorly understood. Data suggest that hepcidin expression is enhanced by iron overload and inflammation, whereas it is inhibited by anaemia

and hypoxia. This is consistent with a compensatory role for hepcidin to limit intestinal absorption during iron overload and to increase iron availability when needed for erythropoiesis.

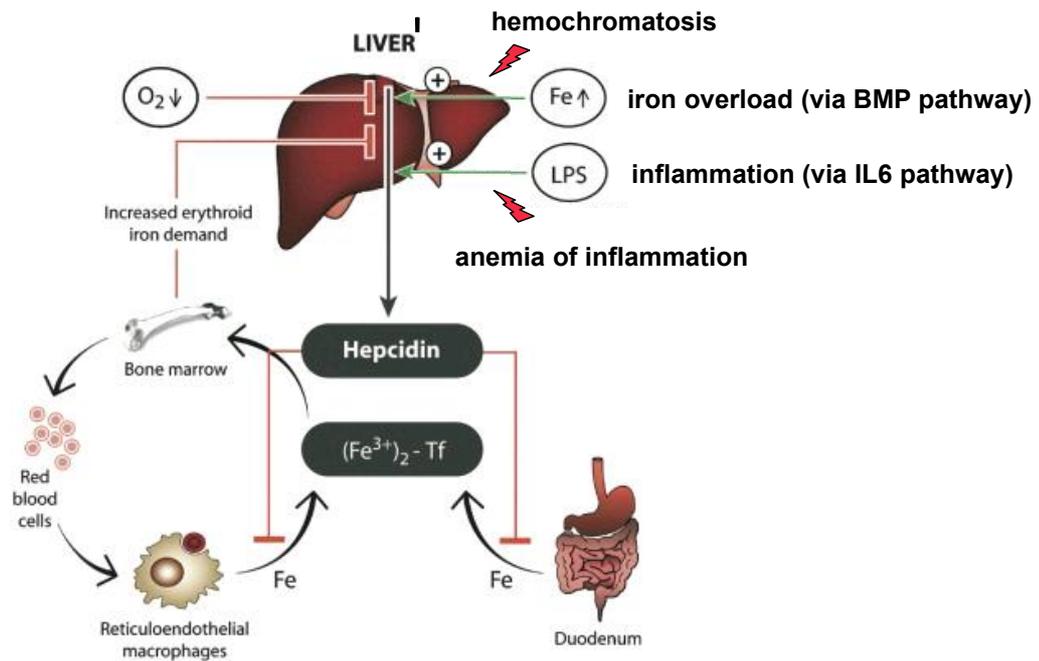


Figure 2: *Hepcidin is expressed by hepatocytes as a response to excess iron levels in blood. Human physiology does not allow for excretion of iron. Therefore strict balance is maintained by controlling the absorption of iron into blood plasma. Hepcidin is controlled by a negative feedback loop. Excess of iron stimulates hepcidin expression that degrades ferroportin thereby blocking iron-absorption from the duodenum resulting in prevention of iron overload in the liver.*

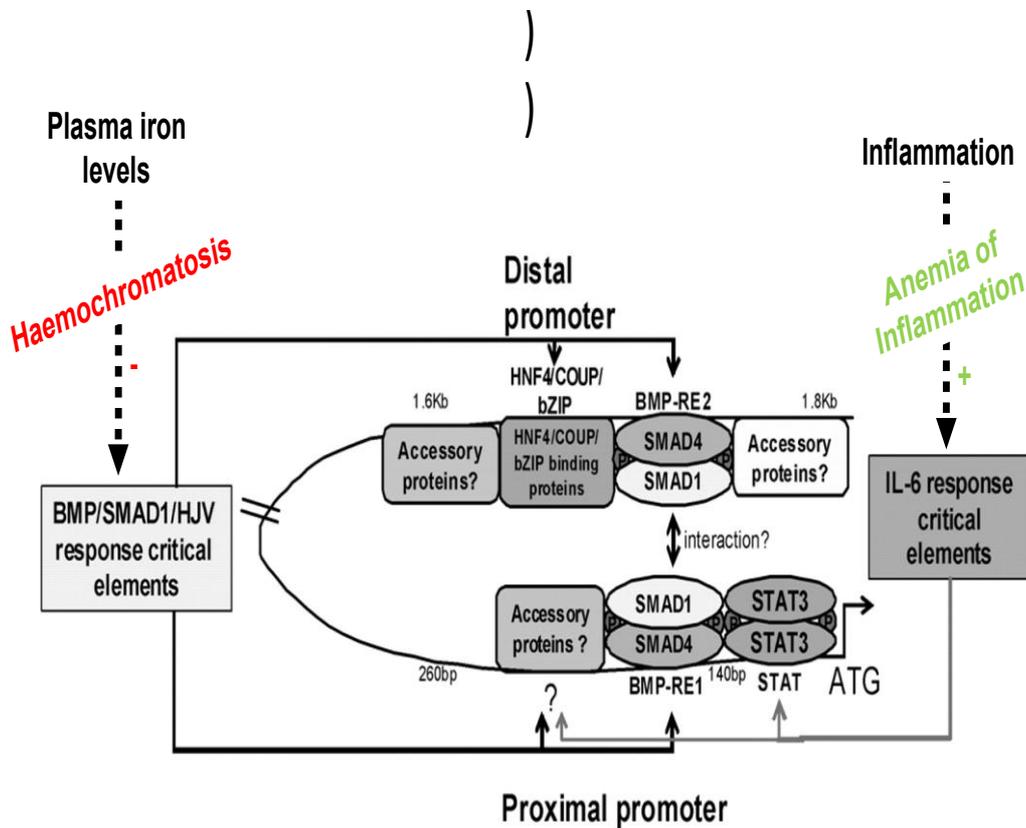


Figure 3: *Hepcidin* promoter with three major responsive sites. The proximal STAT-binding site is responsive to phosphorylated-STAT proteins (pSTATs) released on activation of the IL6/JAK/STAT pathway. There are two BMP responsive regions, one at the proximal and one at the distal end. Simultaneous presence of these two similarly responsive regions increases the BMP-responsiveness of the hepcidin promoter. How the binding sites influence each other is the question we attempt to answer with the help of modelling. [3]

1.2.2 Heparidin Control pathways

Heparidin expression is primarily controlled at the transcriptional level. Information about blood iron levels is transduced from the cell membrane to the nucleus by the bone-morphogenetic-protein (BMP) signalling pathway [43, 45]. HFE and TFR2, two proteins mutated in HH [46, 47], sense increasing blood iron concentrations. The signal is transmitted by the BMP co-receptor HJV and BMP receptor 1 to trigger the phosphorylation and nuclear translocation of SMAD1/5/8 transcription factors (referred to as SMADs hereafter). BMP6, which is regulated by hepatic iron levels, plays a critical role in this process [48, 49]. The heparidin promoter contains two BMP-responsive elements (BRE1 and BRE2) that are recognized by the SMADs (sometimes also abbreviated as BMP-RE1 and BMP-RE2) [50, 51].

1.2.3 Heparidin Regulation with focus on modelling

We see from existing literature [52, 53] that there are two major signalling pathways, which communicate to activate heparidin expression in hepatocytes, control heparidin expression:

- a) **Bone morphogenetic proteins (BMPs)**, a group of cytokines of the (TGF- β) family, induce heparidin expression by activating SMAD transcription factors that bind to **BMP-responsive elements (BREs)** in the heparidin promoter.
- b) **Inflammatory cytokines** (specifically **IL-6**) stimulate heparidin transcription via the JAK/STAT signalling pathway and a **STAT binding motif** proximal to the transcription start site.

BMP and IL6 signalling pathways play different roles in the regulation of hepcidin expression. BMP signalling is important for maintaining steady-state hepcidin expression. Moreover, BMP signalling is activated upon blood iron overload and thereby communicates systemic iron levels. IL-6 is not involved in sensing systemic iron levels, but induces hepcidin in case of inflammation to reduce intestinal iron absorption. Literature evidence suggests that both pathways not only contribute additively to hepcidin, but that there is complex interdependence between pathways: For example, SMAD4, a member of the BMP signalling cascade, contributes to the activation of hepcidin by inflammatory stimuli. Both BMP and IL-6 synergistically induce hepcidin, thus suggesting nonlinear amplification mechanisms, either at the signalling or promoter levels.

Our objective is to identify and quantitatively describe the crosstalk mechanisms between the IL6 and BMP signalling pathways in controlling hepcidin expression since these are incompletely understood (*Fig.3*).

Regulation of hepcidin by BMP

Hepcidin is critical for maintaining systemic iron homeostasis and hepcidin levels are inappropriately low in juvenile and adult forms of hereditary hemochromatosis (HH), a disease caused by mutations in HFE, transferrin receptor 2 (TFR2), hemojuvelin (HJV, HFE2), or hepcidin itself [51].

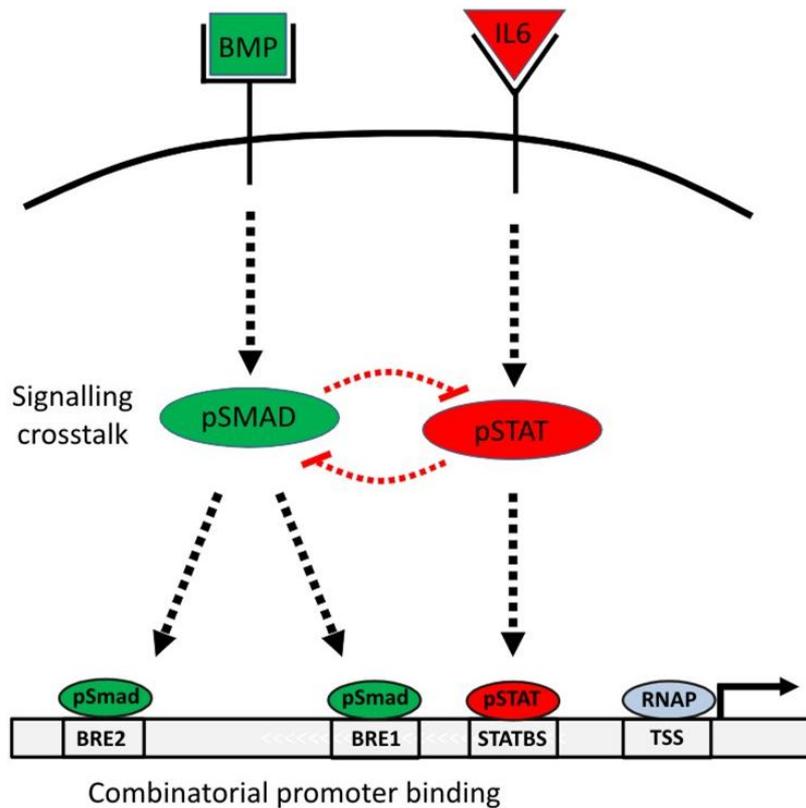


Figure 4: *Signal integration at the hepcidin promoter.* Schematic representation of two critical signalling pathways controlling hepcidin expression. SMAD and STAT transcription factors are phosphorylated upon BMP and IL6 stimulation, and bind BMP-responsive elements (BRE) and a STAT-binding site (STATBS) in the hepcidin promoter, respectively. The importance of signalling crosstalk is not clear.

These mutations thus lead to increased duodenal iron absorption and macrophage iron release resulting in elevated serum iron levels and liver iron overload. Recently, the pathogenesis of HH was linked to the bone morphogenetic proteins (BMPs), a group of cytokines that belongs to the TGF- β superfamily [54]. BMPs initiate intracellular signalling by binding to complexes of type I and type II serine/threonine kinase receptors, which when activated phosphorylate intracellular SMAD proteins. These translocate to the nucleus to control gene transcription. The HH (Hereditary hemochromatosis) protein HJV is a BMP co-receptor that enhances BMP signalling to regulate hepcidin expression and iron metabolism [55].

Regulation of hepcidin by IL6

Hepcidin is induced in anaemia of inflammation, which is an acquired condition that affects patients with a variety of disorders including infection, arthritis, cancer etc. It may range from being moderate to severe but is always characterized by loss of general health. Induction of hepcidin in this case can be majorly attributed to the inflammatory cytokine interleukin-6 (IL6). It has been established that IL6 acts directly to up-regulate hepcidin expression [56]. An IL6 responsive element was identified in the hepcidin promoter and it has been demonstrated that IL6 regulates hepcidin expression through direct binding of STAT3 to the promoter. It has also been shown by [56] that STAT3 is necessary and sufficient to confer IL6 responsiveness in a luciferase reporter assay. These observations have established that besides IL6 regulating hepcidin, even in the absence of elevated cytokine levels, aberrations in hepatic STAT3 regulation could lead to increased systemic hepcidin level and anaemia.

1.2.4 Heparin malfunction and related disease- Motivation for modelling

Heparin malfunction results in two main groups of disorders. The first one relates to dysregulation of iron. Most forms of genetic iron overload as in hemochromatosis 1,2 and 3 are characterized by inappropriately low heparin levels leading to parenchymal (hepatocytes, pancreatic cells and cardiac cells) iron overload. Mutations in the BRE1 promoter element and in the BMP signalling pathway are associated with human hemochromatosis [57]. On the other hand, heparin expression is additionally regulated by external cues other than iron blood levels, including inflammatory cytokines and hypoxia [58, 59]. Inflammatory cytokines like IL6 activate the STAT3 signalling pathway in hepatocytes. Phosphorylated STAT3 transcription factors are directly recruited to a STAT-binding site (STATBS) in the heparin promoter, thereby enhancing expression [56, 60]. Inflammatory cytokines thus diminish iron blood levels. In particular, chronic inflammation causes an iron-related disorder, known as anaemia of inflammation [61]. Patients with chronic inflammation show permanently elevated levels of IL6, increased heparin expression and plasma iron deficiency. The lack of iron availability blocks erythropoiesis (production of red blood cells) thereby causing anaemia of inflammation. This indicates that the integration of BMP and IL6 signals at the level of heparin expression plays a key role in systemic iron homeostasis. In anaemia of inflammation (which is an acquired condition in people suffering from chronic infections over a long period of time), heparin over-production is caused through the pathway involving inflammatory cytokine IL6 and its corresponding signal-transducer and transcription activator STAT3.

The second group of disorders is caused by implicit variations in heparin production. For example, alcoholism [62] and hepatitis C virus infection [63] results in suppression in heparin production that might contribute to the moderate iron-excess that sometimes characterizes these diseases. Therefore it is evident that finding the keys to heparin

regulation and understanding hepcidin expression would in the course of time aid to understand iron-metabolism better and help identifying a series of genetic and acquired iron-relation disorders to make way for novel therapeutic approaches.

II. Results: Towards modelling hepcidin expression

II.1 Conceptual model of systemic iron homeostasis

Existing literature shows that systemic iron homeostasis is maintained by an auto-regulatory negative feedback loop [44]. Iron overload triggers hepcidin expression, and hepcidin in turn lowers intestinal iron influx (*Fig. 2, Page 38*). Such feedback regulation is thought to compensate for fluctuations in dietary iron content [64]. We analyzed a minimal mathematical model of this circuitry to gain insights into iron homeostasis and the factors determining iron blood levels. As a measure of homeostasis, we investigated how well the negative feedback loop maintains constant iron levels in the face of imbalances in systemic iron availability and consumption. The ODE model for feedback (*Equation 1.2, Page 49*) also explicitly shows us what governs and ensures iron-homeostasis. The model suggests that the performance of the iron regulatory loop is optimal if hepcidin expression responds in a steep, nonlinear manner to alterations in iron blood levels. This finding supports previous studies showing that nonlinear negative feedback regulation is particularly efficient in promoting homeostasis [65]. We therefore reasoned that data of dose-dependent hepcidin promoter regulation, and its modulation by iron-independent stimuli, could provide valuable insights into the regulation of iron homeostasis.

The dynamics of the model species were described using the framework of ordinary differential equations.

The model equations read (1.1.1-1.1.2)

$$\frac{d[Fe_b]}{dt} = v_{influx} - v_{efflux} = \frac{k_{influx}}{1+k_{FB} \cdot [hepcidin]} \cdot [Fe_i] - k_{efflux} \cdot [Fe_b] \quad (\text{Equation 1.1.1})$$

$$\frac{d[hepcidin]}{dt} = v_{synthesis} - v_{degradation} = k_{induced} * [Fe_b]^n - k_{deg} \cdot [hepcidin] \quad (\text{Equation 1.1.2})$$

Iron blood levels were described by the species Fe_b , whose levels are controlled by influx and efflux reactions. The iron influx rate is proportional to the intestinal iron concentration (species Fe_i). The efflux rate is modeled as a first-order process, and represents iron excretion and incorporation into red blood cells. Iron blood levels are assumed to control hepcidin expression (parameter $k_{induced}$). The role of the BMP signalling in iron-induced hepcidin expression was not modeled explicitly. Induction was assumed to occur with an exponent n to take into account that the activity of the hepcidin promoter may depend in a nonlinear manner on the iron blood concentration. Hepcidin levels are additionally controlled by first-order degradation (k_{deg}) following its uptake into ferroportin expressing cells [66]. Negative feedback regulation was considered in the model by assuming that the iron influx is negatively influenced by hepcidin. The feedback strength is determined by the parameter k_{FB} .

For the analysis of long-term iron homeostasis, we are not interested in the temporal dynamics of the system but focus on the steady state behavior. The steady state can be

calculated by assuming that iron blood and hepcidin levels do not change over time, i.e., setting

$$\frac{d[Fe_b]}{dt} = 0$$

and

$$\frac{d[hhepcidin]}{dt} = 0$$

Thus, *Eq. 1.1.1* and *Eq. 1.1.2* simplifies to a set of algebraic equations that can be solved for the steady state concentration of iron in the blood. In the limit of strong feedback (large k_{FB}) we can approximate this solution as (1.2)

$$[Fe_b] = \left(\frac{k_{influx} \cdot [Fe_i] \cdot k_{deg}}{k_{efflux} \cdot k_{induced} \cdot k_{FB}} \right)^{1/(1+n)} \quad (\text{Equation 1.2})$$

This equation implies that the steady state iron blood level responds to a change in iron diet content $[Fe_i]$ in a less than linear manner, since the exponent $1/(n+1)$ ensures that $[Fe_i]$ enters as the $(n+1)$ -th root only. We see that this simple model reflects partial homeostasis. The term partial homeostasis implies that changes in the diet iron content cannot be fully compensated. Iron blood levels are less sensitive to changes in the diet content if feedback regulation is steep and nonlinear ($n > 1$). We conclude that the steepness of the hepcidin promoter response towards iron-BMP signalling determines how well the homeostasis loop compensates fluctuations in diet iron levels. Interestingly, tuning the steepness is the only way that homeostasis can be improved in this simple model. Other ways may exist in more

complex versions of the negative feedback model, but the steepness of the promoter will always be a very important factor controlling the performance of the homeostasis loop.

Another important aspect of iron homeostasis is the transcriptional regulation of hepcidin expression by pathways other than BMP signalling. Consider the following extended model which takes into account that hepcidin synthesis occurs by an iron-independent term (k_{basal}) (1.3)

$$\frac{d[Fe_b]}{dt} = v_{influx} - v_{efflux} = \frac{k_{influx}}{1+k_{FB} \cdot [hepcidin]} \cdot [Fe_i] - k_{efflux} \cdot [Fe_b] \quad (\text{Equation 1.3.1})$$

$$\frac{d[hepcidin]}{dt} = v_{synthesis} - v_{degradation} = k_{basal} + k_{induced} * [Fe_b] - k_{deg} \cdot [hepcidin] \quad (\text{Equation 1.3.2})$$

Here, we assumed that hepcidin expression by iron occurs with moderate steepness ($n = 1$), because the steady state of the iron blood level would be much more complicated for $n > 1$. However, the following general conclusions derived from the model continue to hold true for a steeper hepcidin promoter response as well.

The steady state iron blood concentration can be derived from *Eq. 1.3* for the limit of strong feedback (i.e., large k_{FB}) (1.4)

$$[Fe_b] = \frac{1}{2} \cdot \left(-\frac{k_{basal}}{k_{induced}} + \sqrt{\left(\frac{k_{basal}}{k_{induced}}\right)^2 + 4 \cdot \frac{k_{influx} \cdot [Fe_i] \cdot k_{deg}}{k_{efflux} \cdot k_{induced} \cdot k_{FB}}} \right) \quad (\text{Equation 1.4})$$

This equation reveals that the intestinal iron concentration ($[Fe_i]$) enters in a square-root manner for $k_{basal} = 0$, confirming that the system without basal expression shows partial homeostasis. For non-zero basal expression the iron blood concentration can depend on the intestinal iron concentration in a linear manner, i.e., homeostasis may be lost. This emergence of linear behavior is not immediately obvious from *Eq. 1.4*.

Linear behavior can be shown by calculating the gain of the iron blood concentration with respect to the intestinal iron level (1.5).

$$G = \frac{[Fe_i]}{[Fe_b]} \cdot \frac{d[Fe_b]}{d[Fe_i]} \quad (\text{Equation 1.5})$$

The gain is a normalized slope that describes how a percent change in the intestinal iron translates into a percent change in the iron blood level. A gain equal to one implies a linear

relationship between intestinal and blood iron concentrations, while $G = \frac{1}{2}$ and $G = 2$ imply sub-linear and quadratic behavior, respectively (i.e., dampening and amplification). $G < 1$, which means that external perturbations are dampened generally characterizes homeostatic systems. By calculating the gain of *Eq. 1.4*, and using l'Hopitals' rule it can be shown that

$$\lim_{k_{basal}/k_{induced} \rightarrow \infty} (G) = 1 \quad (\text{Equation 1.6})$$

Thus, a linear relationship between intestinal and blood iron levels, i.e., a loss of homeostasis, is observed for sufficiently high iron-independent, basal hepcidin expression. We conclude that the regulation of hepcidin by signalling pathways other than BMP may modulate how strongly iron diet content fluctuations can be compensated.

Whether iron-independent regulators of hepcidin expression compromise iron homeostasis, however, depend on the mode of signal integration at the hepcidin promoter: In *Eqs. 1.3.1-1.3.2 (Page 50)*, it was assumed that the iron-BMP-signalling axis and the iron-independent terms control hepcidin expression in an additive manner (logical OR gate). Alternatively, two pathways may control hepcidin expression multiplicatively (logical AND gate, *Fig. 5*). This scenario is covered by the differential equation system *Eqs. 1.1.1-1.1.2 (Page 48)*: In this model, the iron-independent pathway would change the value of the synthesis rate constant $k_{induced}$. From *Eq. 1.2*, we learned that the degree of homeostasis in this system is only determined by the steepness of promoter regulation by BMP, but not by other parameters such as $k_{induced}$. We therefore conclude that the performance of the iron homeostasis loop is determined by the steepness of the promoter response towards BMP stimulation. Whether or not this steepness (and thus iron homeostasis) is modulated by iron-independent regulators of hepcidin expression depends on the mode of signal integration at the hepcidin promoter.

Additionally we also conclude that iron-independent pathways would *not* modulate how strongly iron diet content fluctuations can be compensated in the multiplicative scenario. Instead, the iron-independent pathways would modulate the absolute iron blood levels, although to a limited extent only.

In the next sections, we therefore set out to systematically quantify hepcidin promoter regulation by BMP and IL6 using systematic promoter mutagenesis and co-stimulation experiments.

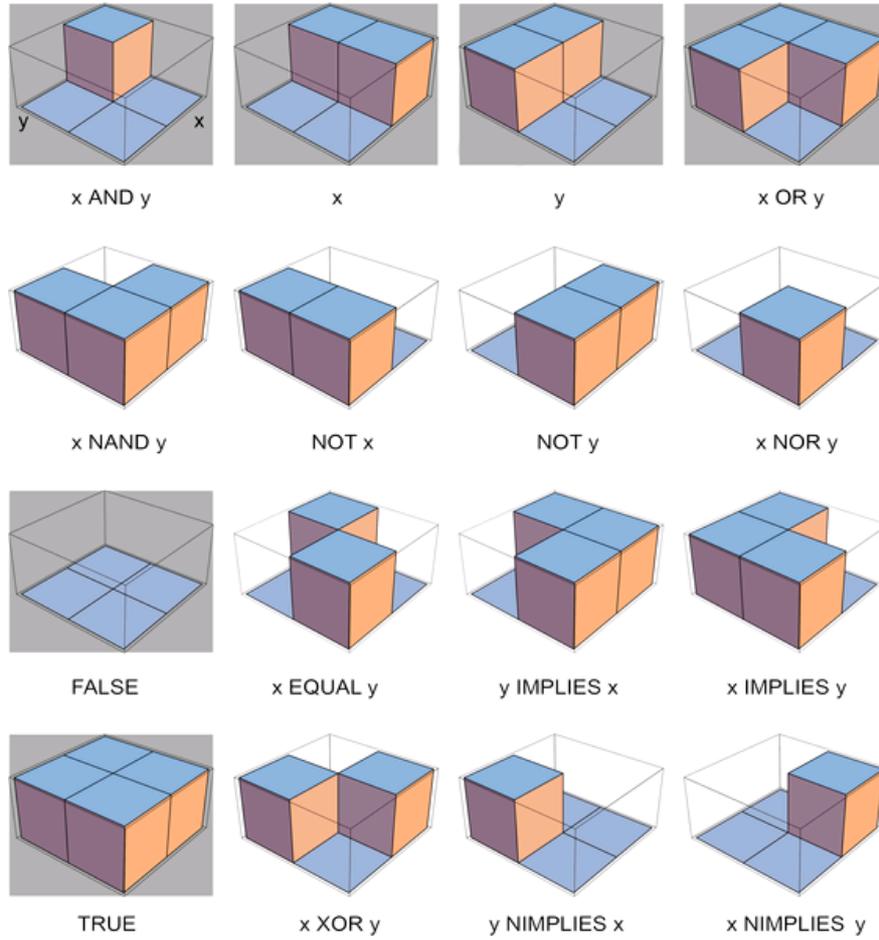


Figure 5: *Plasticity of combinatorial regulation.* Transcription may follow logic models such that presence, absence or combined effect of binding sites or signalling pathways directly influence expression.

II.2 Quantification of hepcidin cross-regulation by BMP and IL6

II.2.1 Experimental characterization of hepcidin promoter variants reveals impact of individual binding sites - performed by Muckenthaller group

Our model is based on in-vivo homeostasis where we use cell culture system to characterize promoter response. Genetic reporters were used by our experimental collaborators to study gene expression levels and the impact of transcription-factor binding-sites (TFBS) on hepcidin expression. The analysis focused on the influence of BMP-responsive elements (BRE1 and BRE2) and the STAT binding site (STAT-BS) on hepcidin expression. In particular, we hoped to understand the interdependency of TFBS by combinatorial TFBS mutations. Our collaborators cloned the hepcidin promoter variants upstream of the firefly luciferase reporter, and transfected the constructs into Human hepatoma (HuH7) cells. Luciferase is an enzyme that catalyzes the conversion of Luciferin into Oxyluciferin, a reaction that is accompanied by the emission of light ('bioluminescence'). The light signal is therefore a direct readout for promoter activity. To correct for different transfection efficiencies, the Firefly luciferase signal has to be normalized by the signal of a co-transfected Renilla luciferase reporter (that catalyzes a different light reaction). Luciferase reporter assays were allowed to measure the activity of WT (wild-type, *hepcidin promoter with its binding sites intact*) and mutant hepcidin promoter (*binding sites as seen in Fig. 3, Page 39, combinatorially mutated or deleted*) versions under various experimental conditions. There are eight different constructs where different transcription factor binding elements are mutated individually or in combination by direct mutagenesis. Some constructs (hepcidin DNA segments) were based on a short version of the hepcidin promoter (1 kb), which contains the STATBS and the BRE1. Additionally, there are characterized constructs that

were based on the native (as occurring in nature) full-length 2.7 kb promoter containing all three TFBS (STATBS, BRE1, and BRE2). For example, the construct BRE1_1kb refers to the 1kb promoter that lacks the BRE1. TFBS mutations were introduced in both constructs either singularly (BRE1_1kb, STATBS_1kb, BRE1_2.7kb, STATBS_2.7kb, BRE2_2.7kb), or in combination (rest). In total, 8 out of 12 possible constructs were characterized; the promoter variants lacking all TFBS (STATBS_BRE1_BRE2_2.7kb, STATBS_BRE1_1kb) were not investigated since luciferase expression was undetectable. Additionally, the following promoters were not available: STATBS_BRE2_2.7kb and STATBS_BRE1_2.7kb. Our collaborators carried out gene-expression analysis on the luciferase constructs but by combinatorially co-stimulating them with varying doses of BMP and IL6 giving rise to 4x4 matrices per construct which we analyse in three-dimensions by illustrating them as heatmaps (*Fig. 6*). Each replicate measurement series was normalized to correct for slight differences in absolute luciferase signals between replicate experiments: Each measured value was divided by the median over all 92 data points (7 constructs with 16 stimulation conditions each) of that replicate series.

QPCR measurements of endogenous hepcidin mRNA were performed under the same experimental conditions to confirm that the reporter measurements reflect endogenous hepcidin expression.

II.2.2 Theoretical analysis of promoter data

We estimated the experimental errors by calculating standard deviations over all replicates. In some cases, we calculated fold changes in luciferase expression by dividing the measured values by their respective basal level (i.e., expression in unstimulated cells) or by the expression of other promoter versions under the same experimental conditions. The errors of these fold-changes were calculated using a Monte-Carlo approach: In the simulations, random realizations were drawn from normal distributions with mean and standard deviation equal to those of the measured luciferase expression data. Fold-changes were calculated for 103 pairs of realizations, and the fold-change error was evaluated by calculating the standard deviation of the resulting probability distributions.

Subsequently, using qPCR (real-time polymerase chain reaction) data, we systematically compared qPCR and luciferase signals (fold-change over basal expression) over various stimulation and knockdown conditions [44] and found a high correlation coefficient of at least $R = 0.9835$ (replicate 2, *Fig. 7A, Page 62*) which confirms that luciferase measurements faithfully reflect endogenous regulation. Moreover, this eventually supports our model assumption that post-transcriptional regulation of hepcidin expression is negligible.

Mono-stimulation experiments reveal that the WT promoter primarily responds to BMP6 stimulation, since the highest BMP6 dose induced a 450-fold induction of luciferase activity, while IL6 only enhanced luciferase levels by 20-fold (*Fig. 6*). Notably, even higher IL6 levels could not raise luciferase activity any further, as can be seen from more detailed titration experiments with higher doses of IL6. We conclude that IL6 is an intrinsically weaker inducer of hepcidin expression when compared to BMP.

Co-stimulation of the WT construct with IL6 and BMP enhanced luciferase activity compared to mono-stimulation with IL6 or BMP (*Fig. 6*).

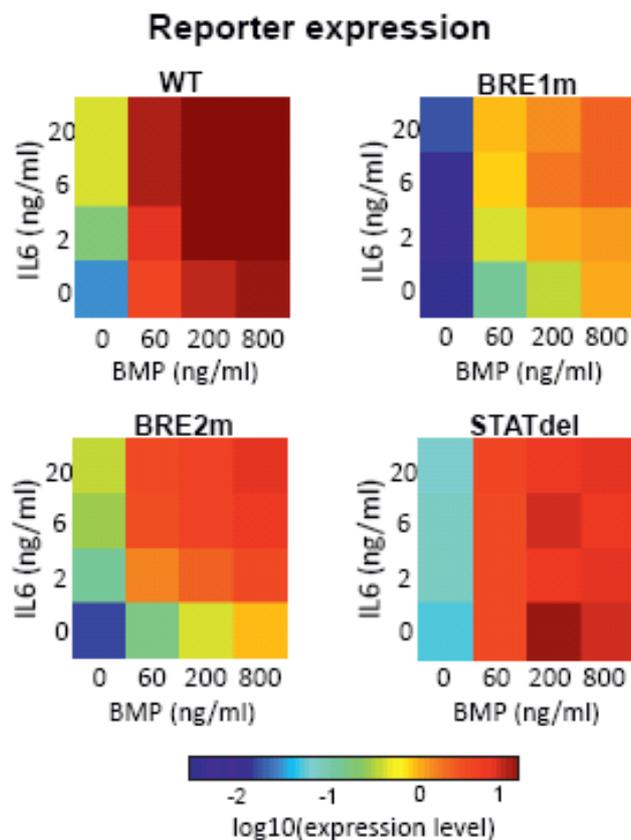


Figure 6: Analysis of transcription factor crosstalk at the promoter level by reporter gene assays. Luciferase expression is driven by the wildtype (WT) hepcidin promoter (3 kb upstream of TSS) or promoter mutants lacking one of the transcription factor binding sites (panel B; BRE1m = BRE1 mutated; STATdel = deleted for STATBS). Luciferase activity of each reporter construct (shown on a log₁₀-scale) was measured for increasing doses of IL6 and/or BMP (n = 6).

From a quantitative point of view, the co-stimulation response was more than additive, but less than multiplicative: the induction upon co-stimulation with the maximal doses of BMP and IL6 (600-fold) ranged between the sum and the product of the mono-stimulation effects (450- and 20-fold, respectively, *Fig. 6*). QPCR measurements quantitatively confirm this co-stimulation response at the level of endogenous hepcidin mRNA, thereby supporting the physiological relevance of the luciferase results (*Fig. 7*).

Whether transcriptional regulation follows an additive or multiplicative mode depends on the molecular events at the promoter level: Additive behavior would be expected if the two stimuli drive transcription in a mutually exclusive manner (e.g., if only one of the respective transcription factors establishes a loop with the TSS at a time). In operating multiplicatively, the transcription factors would simultaneously attract RNAP to the TSS: The simultaneous presence of multiple independent recruitment sites in the promoter could boost RNAP binding, thus establishing a more than additive response (avidity effect). Most thermodynamic models of gene expression implement this avidity concept and predict a multiplication of mono-stimulation effects in the absence of promoter saturation or cooperative transcription factor binding [67].

We quantified the co-stimulation response for all available promoter constructs and stimulus combinations to distinguish between additive and multiplicative modes in hepcidin promoter regulation. We find that the hepcidin promoter mutants show near-multiplicative behavior, while the WT promoter tends to respond sub-multiplicatively. Further, the sub-multiplicative behavior appears to be less pronounced the lower the luciferase signal, indicating a promoter saturation effect (*Fig. 8*). This suggests that the hepcidin promoter generally follows a multiplicative mode of regulation. Accordingly, after extensive application of a variety of modelling approaches, we find that standard thermodynamic models with multiplicative

behavior are able to describe the promoter data, while additive models fail to do so. The less-than multiplicative behavior that arises for certain constructs and stimulus conditions further suggests the existence of saturation effects or inhibitory crosstalk at the signalling and promoter levels, respectively.

We analyzed the co-stimulation response of mutant promoters to further understand the determinants of hepcidin cross-regulation. The BRE1m promoter primarily responds to BMP stimulation and exhibits a co-stimulation heatmap that is qualitatively similar to the WT promoter. However, the expression levels of BRE1m are generally ~10-fold lower than WT (*Fig. 6, Page 58, compare BRE1m and WT*). Thus, the BRE1 deletion ‘scales’ hepcidin expression and elicits a similar reduction in basal and induced expression. The impact of BRE1 on basal expression can be explained by partial phosphorylation of SMAD1/5/8 in unstimulated cells.

Analysis of the BRE1 mutant promoter (*in Fig. 6*) uncovers that BRE1 is required for full IL6 responsiveness of the hepcidin promoter: strong mono-stimulation with IL6 enhances luciferase activity 20-fold for the WT promoter, but only 5-fold for the BRE1m promoter. This observation supports previous reports concluding that BRE1 is required for IL6-induced hepcidin expression [50]. We additionally see that BRE1 is only partially required for the response towards high IL6 doses. We next investigated how a deletion of the STATBS in turn affects the BMP responsiveness. The STATdel promoter shows a slightly but significantly reduced BMP responsiveness when compared to the WT promoter (*Fig. 6; p < 0.001, paired t-test*). This indicates that the STATBS promotes full BMP responsiveness. We conclude that BRE1 and the STATBS synergistically control hepcidin expression: the presence of one element enhances the responsiveness towards the stimulus targeting the other element.

However, the synergism is not fully symmetric, as the BRE1 deletion has a stronger effect on IL6 responsiveness than a STAT deletion on BMP inducibility.

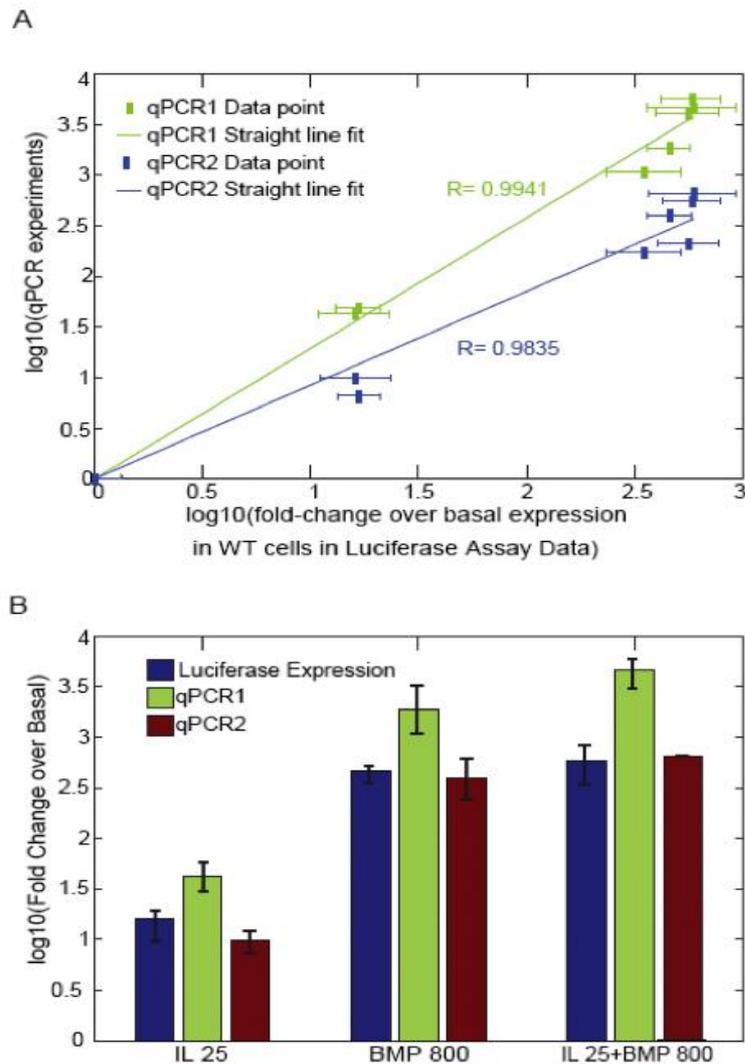


Figure 7: *Less than multiplicative co-stimulation response in the WT promoter preserved at the level of endogenous hepcidin mRNA expression.* The fold-expression-changes over basal were compared for qPCR and luciferase assays. (A) Scatter plots showing correlation between Luciferase Assay data (n=3) and qPCR data (green and blue are biological replicates, each with technical replicates (n=2)). The axes show log₁₀ of fold-changes over basal expression in unstimulated cells. The respective Pearson Correlation Coefficients(R) for qPCR1 vs. Luciferase Expression and qPCR2 vs. Luciferase Expression support a strong correlation. The blue and green solid lines show linear fits to the data. (B) Co-stimulation modestly increases expression when compared to BMP mono-stimulation. Bargraph showing log₁₀ of fold changes over basal (y- axis) for Luciferase Expression (Blue), and for the two biological replicate qPCR experiments (green and brown; same data as in panel A).

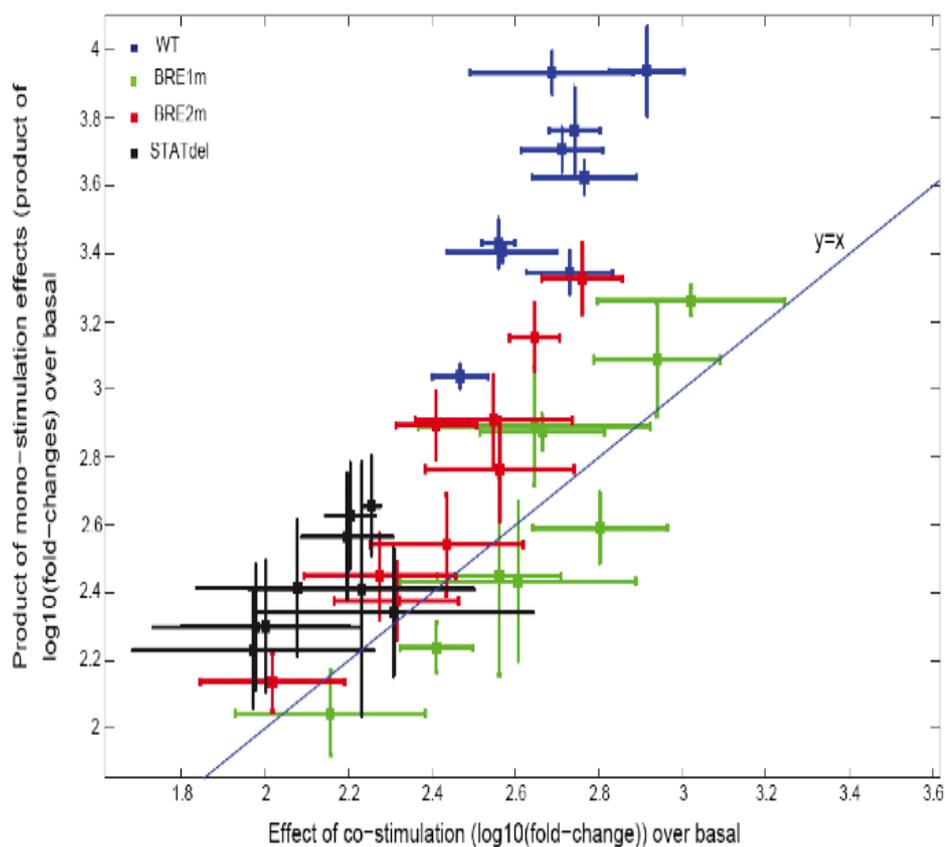


Figure 8: *Co-stimulation response of luciferase reporters reveals multiplicative/sub-multiplicative behavior for multiple stimulus concentrations and promoter mutants.* The x dimension shows the experimentally observed fold-change in expression upon co-stimulation with BMP and IL6. The y dimension shows the product mono-stimulation responses with the same doses of BMP and IL6, respectively. Each data point represents one co-stimulation condition (different concentrations of BMP and IL6 and/or different promoter constructs). The colors of the data points correspond to different promoter constructs (legend). The bisectrix (solid line) marks the expectation for a multiplicative system.

The co-stimulation response of the BRE2m construct is qualitatively different from that of the WT and BRE1m promoters (*Fig. 6*). The BRE2m promoter resembles a coincidence detector

(‘logical AND gate’) (*Fig. 5, Page 54*): Mono-stimulation with either BMP or IL6 raises luciferase activity to intermediate levels only, and co-stimulation with both ligands is required for maximal expression (*Fig. 6*). This suggests that BRE1 and BRE2 fulfill distinct functions in hepcidin regulation: The BRE2m promoter shows only a slight reduction in basal luciferase activity compared to the WT promoter. This reduction in basal activity did not reach statistical significance (paired t-test), and was much less pronounced than the effect of a BRE1 mutation. The promoter responsiveness to IL6 mono-stimulation was not significantly affected by the BRE2 mutation (paired t-test), indicating that BRE2 has lesser impact on IL6-induced hepcidin expression than BRE1. However, the responsiveness to BMP mono-stimulation is strongly reduced in the BRE2m promoter ($0.001 < p < 0.0025$, paired t-test): maximal BMP stimulation enhances luciferase activity by only 80-fold in BRE2m, when compared to 400-fold in WT and BRE1m. We conclude that the most prominent feature of the BRE2m promoter is the reduced ability to respond to BMP stimuli compared to the WT promoter. In contrast, the BRE1 mutation primarily affects basal activity and IL6 inducibility of the promoter. The loss of BRE1 hardly affects the promoter responsiveness to maximal BMP doses, although it has some impact at intermediate BMP doses. Taken together, these data raise the interesting question of how two transcription-factor binding-sites with very similar sequence can show qualitatively distinct behavior in hepcidin expression regulation.

This quantitative analysis shows that BMP and IL6 signalling cascades regulate hepcidin expression in an interdependent manner. Apparently, the BMP/IL6 co-stimulation response reflects a superposition of multiple regulation effects. Therefore for further analysis we adopted mathematical modelling to quantitatively understand the dynamics of hepcidin promoter regulation. In particular, we sought to understand why BRE1 and BRE2 play different roles in hepcidin expression with respect to basal expression, BMP inducibility and co-stimulation response.

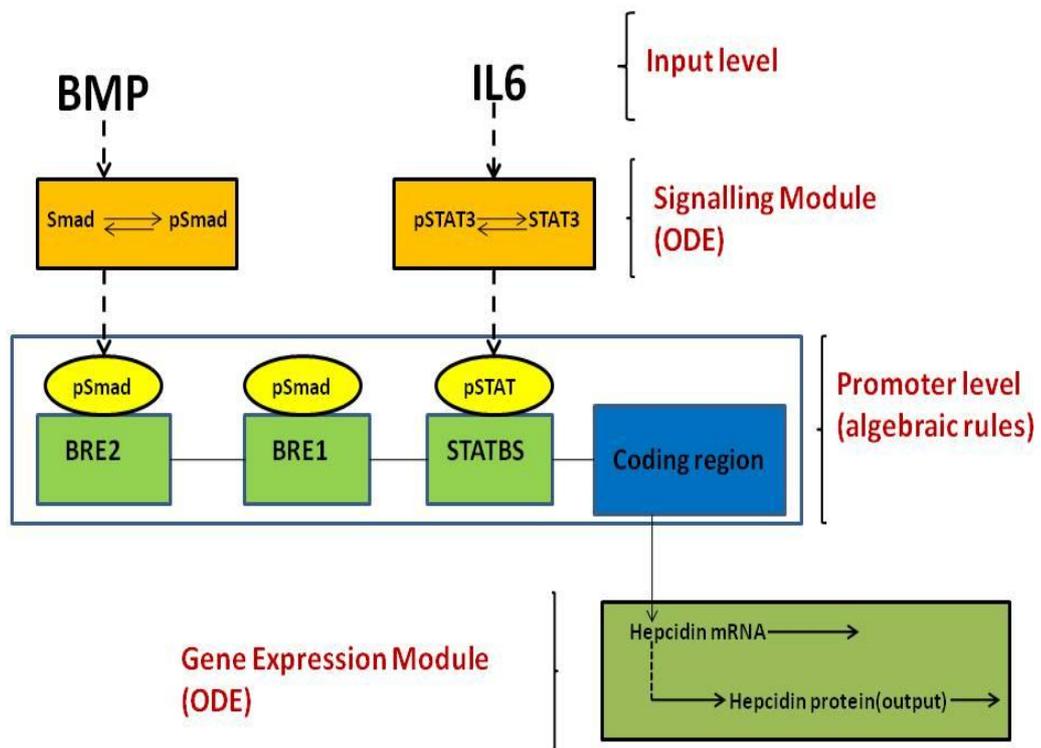


Figure 9: Modular modelling approach for arriving at gene expression data by integrating systemic behaviour at each level. This approach provides quantitative understanding of effects involved at the various levels leading to gene expression.

II.3 Modelling Regulation of Hecpudin

II.3.1 Definition of model assumptions

We derived a mathematical model of hepcidin regulation by IL6 and BMP that links the activity of signalling pathways to combinatorial promoter regulation and gene expression.

The molecular mechanisms underlying regulation of hepcidin expression by IL6 and BMP can be subdivided into three layers of regulation (*Fig. 9*): Each input rapidly activates an intracellular signalling pathway (*'signalling module'*). The active signalling molecules (i.e., *pSMAD* and *pSTAT*) in turn bind to transcription-factor binding-sites in hepcidin promoter (*'promoter level'*). This induces transcription of hepcidin mRNA, which is then either degraded or translated into protein (*'gene expression module'*). It is necessary for our model to represent this biological reality of a) *signalling*, b) *promoter-level interactions* and c) *protein expression* to quantitatively analyze the extent to which IL-6 and BMP regulate the expression of hepcidin.

The upstream signalling module of the model describes stimulus-induced transcription factor phosphorylation, while the downstream module characterizes combinatorial transcription factor binding to the promoter and gene expression. The model describes hepcidin regulation at steady state and neglects the temporal dynamics of gene expression. The kinetic parameters of the model were *a priori* unknown, and were estimated by fitting the model to luciferase activity data.

In this chapter, we would derive quantitative signalling module with simple Hill-equations, promoter model based on the foundation of thermodynamics and a gene expression module.

a) Modelling dose-response behaviour of signalling module- without assuming signalling crosstalk

For the signalling module, we chose a phenomenological modelling approach and did not focus on the molecular mechanisms underlying intracellular signal processing. Phenomenological modelling means that an arbitrary mathematical function is used to represent the dose-response curves of transcription factor phosphorylation. We chose the sigmoidal Hill equation, as BMP- and IL6-induced dose-response curves are typically of sigmoidal shape. We neglected putative signalling crosstalk between IL6 and BMP pathways in the initial versions of our model.

The dose-response of intracellular signalling pathways is typically sigmoidal in shape, and described by the Hill equation [46, 68] (2.0)

$$y = \frac{y_{basal} + y_{max} * S^n}{S^n + EC_{50}^n} \quad (\text{Equation 2.0})$$

This Hill equation expresses signalling (y) as a sigmoidal function of the stimulus (S), and takes into account basal signalling (y_{basal}), maximal pathway activation (y_{max}), the half-maximal-stimulus ($EC50$) and the Hill coefficient (n) as a measure of steepness of the dose-

response curve. In our initial model, we neglected inhibitory crosstalk between BMP and IL6 pathways, and described the steady state of each signalling cascade using a Hill equation. The concentrations of phospho-STAT (pST) and phospho-SMAD (pSM) are thus described by (2.1.1) and (2.1.2)

$$[pST] = y_{max,1} * f_{Hill,IL6} \quad (Equation 2.1.1)$$

$$[pSM] = y_{basal,2} + y_{max,2} * f_{Hill,BMP} \quad (Equation 2.1.2)$$

Where

$$f_{Hill,IL6} = \frac{[IL6]^{n_1}}{[IL6]^{n_1} + EC_{50,1}^{n_1}} \quad (Equation 2.1.3)$$

$$f_{Hill,BMP} = \frac{[BMP]^{n_2}}{[BMP]^{n_2} + EC_{50,2}^{n_2}} \quad (Equation 2.1.4)$$

The output concentrations of these Hill equations (pSM and pST) were in turn used as an input for the thermodynamic promoter model.

b) Promoter Module

The downstream promoter module describes steady state luciferase activity as a function of phosphorylated transcription factors. We neglected post-transcriptional steps in gene regulation, and could therefore assume that steady state gene expression is proportional to the transcription initiation rate [69]. The transcription initiation rate thus served as a model readout that was fitted to the luciferase expression data (using a scaling factor). We used the framework of thermodynamic modelling to mechanistically describe transcriptional initiation. In order to apply kinetic models to describe gene expression at the promoter level, we require kinetic laws for the transcription rates. Putting it simply, we assume in such cases that the transcription rate of a gene depends on regulator activities x_i and other influences on transcription are neglected. The rate y is defined by a gene regulation function

$$y(t) = f(x(t), \mathbf{p}) \quad (\text{Equation 2.1.5})$$

The mathematical function f and the parameter \mathbf{p} are specific to the gene in question. The vector x consists of information on activities of all the regulators in the system. We make two underlying assumptions for developing a quantitative kinetic model of gene regulation: (i) regarding the time-scale the different states are in thermodynamic equilibrium and the probability of each state depends on its binding energy and on the concentration of the regulator molecules present. (ii) Transcription initiation occurs randomly at a certain rate in each of the states.

The mathematical form of the gene regulation function can follow from these assumptions and the analysis of the possible state-variants.

Each of the states has a free energy associated with it, given by

$$F = E - T * S \quad (\text{Equation 2.1.6})$$

where T is the temperature, E stands for the energy and S denotes the entropy of the state. The free energy depends on the presence and sequence of binding sites on the promoter. The number of available regulator molecules influences the entropy term. The statistical weight (Z_i) of a promoter state i is determined by its free energy (F_i) and is defined as

$$Z_i = e^{\frac{F_i}{K_b * T}} \quad (\text{Equation 2.1.7})$$

where K_b is the Boltzmann constant.

The statistical modelling approach illustrated by Bintu et al. [24] establishes a way of linking the microscopic binding states and macroscopic gene regulation mechanism. The modelling framework focuses on gene regulation at the level of transcription initiation, and assumes that the activity of a gene can be understood based on the determinants controlling RNAP recruitment to the transcription start site (TSS). Thermodynamic modelling has been shown to accommodate various modes of signal integration on a promoter [1, 24]. Experimental and theoretical characterization of bacterial and yeast promoters revealed that even relatively simple promoter architectures could give rise to complex and non-intuitive modes of signal

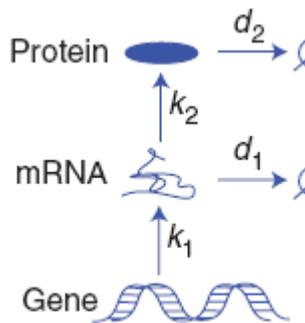
integration [70-72]. More recently, thermodynamic modelling was extended to aspects of eukaryotic gene regulation, including nucleosome-positioning effects [73].

In this project, we applied thermodynamic modelling to quantitatively analyze how the iron-sensing BMP and inflammatory IL6 pathways coordinately control hepcidin expression. We show that cross-regulation of hepcidin promoter activity by BMPs and IL-6 can be quantitatively understood as a superposition of two promoter regulatory events: weak cooperative binding of transcription factors and strong promoter saturation with RNA polymerase. We find that the presence of two BMP-responsive elements promotes high sensitivity towards BMP stimulation, which is a key feature promoting iron homeostasis *in vivo*. IL6 reduces the ability of the promoter to sense the BMP signal in co-stimulation experiments, suggesting that this mechanism leads to disturbance of iron homeostasis in the anaemia of inflammation. Taken together, we employed a combined experimental and theoretical approach to gain insights into the regulation of hepcidin and iron homeostasis in health and disease.

c) Gene-expression module

A simple promoter model predicts how the transcription rate depends on promoter architecture. However, the luciferase assays used as experimental readout for gene expression reflect the amount of luciferase protein produced. Thus, the experimental and theoretical readouts disagree. In this section, we eventually show that protein expression is proportional to the transcription rate in simple models of gene expression, thus justifying our assumption that luciferase is an appropriate readout for promoter activity and the quantity of protein

produced at the gene-expression level is proportional to the transcription initiation rate for the promoter module.



Inset 1: Expression of a gene into its protein as determined by the four stages: transcription, translation, mRNA degradation and protein degradation.

Consider a minimal model for gene expression (*Inset 1*), where mRNA is subject to synthesis and degradation. Moreover, the mRNA is translated into protein, which in turn might be degraded.

The system can be described by the following set of ODEs:

$$\frac{d(mRNA)}{dt} = k_1 - d_1 \cdot mRNA \quad (\text{Equation 2.2.1})$$

$$\frac{d(Protein)}{dt} = k_2 \cdot mRNA - d_2 \cdot Protein \quad (\text{Equation 2.2.2})$$

where k_1 , k_2 , d_1 and d_2 are the transcription rate, translation rate, mRNA degradation rate and protein degradation rate, respectively.

In order to understand the relationship between promoter activity (k_1) and protein expression, we integrate the differential equation system assuming that the promoter activity switches from low ($k_1=x$) to high ($k_1=y$) at $t=0$.

The time course of protein expression $P(t)$ is given by

$$P(t) = P_{ss} \left(1 - \frac{P_{ss}-P(0)}{P_{ss}} \cdot \frac{d_1 \cdot e^{-d_2 \cdot t} - d_2 \cdot e^{-d_1 \cdot t}}{d_1 - d_2} \right) \quad (\text{Equation 2.2.3})$$

Where P_{ss} is the steady state protein concentration and is given by:

$$P_{ss} = \left(\frac{k_1}{d_1} \right) \cdot \left(\frac{k_2}{d_2} \right) \quad (\text{Equation 2.2.4})$$

Experimental measurements revealed that the half-lives of luciferase mRNA and protein are 1.5h and 1h, respectively [74, 75].

Since the mRNA and protein degradation follow simple exponential decays, the half-life and the degradation rates d are related by

$$t_{\frac{1}{2}} = \ln(2)/d \quad (\text{Equation 2.2.5})$$

Thus, the mRNA and proteins degradation rates are approximately given by 0.5 h^{-1} and 0.7 h^{-1} . Given that our luciferase measurements of gene expression are performed at 24hours, the exponential part in *Eq. 2.2.3* is close to zero, implying that the system is close to steady state under our experimental conditions.

In steady state therefore, protein expression is proportional to promoter activity (k_I), thus justifying our assumption that luciferase signals are a good experimental readout for our promoter model

II.3.2 Thermodynamic Model for Hecpidin Promoter

a) Derivation of Thermodynamic model

The hepcidin promoter consists of three major regulatory elements, all of which are controlled by IL6 and/or BMP signalling [64]: (i) the STAT-binding site (STATBS); (ii) and (iii) the BMP-responsive elements 1 and 2 (BRE1 and 2). When occupied by cognate transcription factors, each of these sites may directly recruit RNA polymerase II and thereby initiate transcription. Additionally, the transcription factors may mutually enhance their impact on transcription, e.g., by cooperative promoter binding, DNA looping or opening of chromatin.

A thermodynamic model of transcription was derived in order to quantitatively describe signal integration by the hepcidin promoter (reviewed in [24] [76]). Thermodynamic modelling applies methods from statistical thermodynamics to describe combinatorial binding of transcription factors to promoters. The approach additionally takes into account protein-protein interactions on the promoter: (i) transcription factors may contact RNAP, and thereby promote RNAP recruitment and transcription. (ii) transcription factors may form pairwise complexes, thus cooperatively enhancing their promoter binding or transcriptional activation. In the following, we will derive a thermodynamic model comprising all possible pairwise protein-protein interactions on a promoter containing three transcription-factor binding-sites. We reduced the complexity of this model based on model fitting and model selection approaches.

A central concept in thermodynamic modelling is establishing the so-called promoter states that represent the transcription factor binding configurations of a promoter. Combinatorial binding of transcription factors to three specific binding sites in the promoter and polymerase

binding to the transcription start site (TSS) gives rise to $2^4 = 16$ promoter states (Fig. 10, Page 89).

Thermodynamic modelling assumes that the transcription initiation rate is proportional to the amount of polymerase bound to the promoter. We therefore derive an expression for the probability of polymerase binding based on the promoter states. The probability of polymerase binding is given by the sum over the statistical weights of the polymerase-bound promoter states (Z_{bound}) divided by the sum over all promoter weights (Z_{tot}) (2.3.1)

$$p_{bound} = \frac{\sum Z_{bound}}{Z_{tot}} \quad (\text{Equation 2.3.1})$$

We use the notation ‘P’ to describe empty promoter states where all P polymerase molecules are bound non-specifically to genomic DNA (or diffuse freely in the nucleoplasm). ‘P-1’ refers to the active promoter scenario where one polymerase molecule is bound specifically to the TSS, and the remaining P-1 molecules show background or no binding. Using a similar nomenclature for each of the activating transcription factor binding sites A_i , we can write the sum over the statistical weights of the polymerase-bound states as (2.3.2)

$$\begin{aligned} \sum Z_{bound} = & Z(P - 1, A_1, A_2, A_3) + Z(P - 1, A_1 - 1, A_2, A_3) + Z(P - 1, A_1, A_2 - 1, A_3) + \\ & Z(P - 1, A_1, A_2, A_3 - 1) + Z(P - 1, A_1 - 1, A_2 - 1, A_3) + Z(P - 1, A_1 - 1, A_2, A_3 - 1) + \\ & Z(P - 1, A_1, A_2 - 1, A_3 - 1) + Z(P - 1, A_1 - 1, A_2 - 1, A_3 - 1) \end{aligned}$$

$$(\text{Equation 2.3.2})$$

The total statistical weight additionally takes into account 8 polymerase-free promoter states (2.3.3)

$$\begin{aligned}
Z_{tot} &= \sum Z_{bound} + \sum Z_{free} \\
&= \sum Z_{bound} + Z(P, A_1, A_2, A_3) + Z(P, A_1 - 1, A_2, A_3) + Z(P, A_1, A_2 - 1, A_3) \\
&\quad + Z(P, A_1, A_2, A_3 - 1) + Z(P, A_1 - 1, A_2 - 1, A_3) + Z(P, A_1 - 1, A_2, A_3 - 1) \\
&\quad + Z(P, A_1, A_2 - 1, A_3 - 1) + Z(P, A_1 - 1, A_2 - 1, A_3 - 1)
\end{aligned}
\tag{Equation 2.3.3}$$

The weights of the 16 individual promoter states are given by (2.3.4-2.3.19)

$$Z(P, A_1, A_2, A_3) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^{NS}}{k_B * T}}}{P! * A_1! * A_2! * A_3! * (N_{NS} - P - A_1 - A_2 - A_3)!}
\tag{Equation 2.3.4}$$

$$Z(P, A_1 - 1, A_2, A_3) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^S + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^{NS}}{k_B * T}}}{P! * (A_1 - 1)! * A_2! * A_3! * (N_{NS} - P - (A_1 - 1) - A_2 - A_3)!}
\tag{Equation 2.3.5}$$

$$Z(P, A_1, A_2 - 1, A_3) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^S + \varepsilon_{A_3}^{NS}}{k_B * T}}}{P! * A_1! * (A_2 - 1)! * A_3! * (N_{NS} - P - A_1 - (A_2 - 1) - A_3)!} \quad (\text{Equation 2.3.6})$$

$$Z(P, A_1, A_2, A_3 - 1) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^S}{k_B * T}}}{P! * A_1! * A_2! * (A_3 - 1)! * (N_{NS} - P - A_1 - A_2 - (A_3 - 1))!} \quad (\text{Equation 2.3.7})$$

$$Z(P, A_1 - 1, A_2 - 1, A_3) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^S + \varepsilon_{A_2}^S + \varepsilon_{A_3}^{NS} + \varepsilon_{A_1 A_2}}{k_B * T}}}{P! * (A_1 - 1)! * (A_2 - 1)! * A_3! * (N_{NS} - P - (A_1 - 1) - (A_2 - 1) - A_3)!} \quad (\text{Equation 2.3.8})$$

$$Z(P, A_1 - 1, A_2, A_3 - 1) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^S + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^S + \varepsilon_{A_1 A_3}}{k_B * T}}}{P! * (A_1 - 1)! * A_2! * (A_3 - 1)! * (N_{NS} - P - (A_1 - 1) - A_2 - (A_3 - 1))!} \quad (\text{Equation 2.3.9})$$

$$Z(P, A_1, A_2 - 1, A_3 - 1) = \frac{N_{NS}! e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^S + \varepsilon_{A_3}^S + \varepsilon_{A_2 A_3}}{k_B T}}}{P! * A_1! * (A_2 - 1)! * (A_3 - 1)! * (N_{NS} - P - A_1 - (A_2 - 1) - (A_3 - 1))!}$$

(Equation 2.3.10)

$$Z(P, A_1 - 1, A_2 - 1, A_3 - 1) = \frac{N_{NS}! e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^S + \varepsilon_{A_2}^S + \varepsilon_{A_3}^S + \varepsilon_{A_1 A_2} + \varepsilon_{A_1 A_3} + \varepsilon_{A_2 A_3}}{k_B T}}}{P! * (A_1 - 1)! * (A_2 - 1)! * (A_3 - 1)! * (N_{NS} - P - (A_1 - 1) - (A_2 - 1) - (A_3 - 1))!}$$

(Equation 2.3.11)

$$Z(P - 1, A_1, A_2, A_3) = \frac{N_{NS}! e^{-\frac{\varepsilon_P^S + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^{NS}}{k_B T}}}{(P - 1)! * A_1! * A_2! * A_3! * (N_{NS} - (P - 1) - A_1 - A_2 - A_3)!}$$

(Equation 2.3.12)

$$Z(P - 1, A_1 - 1, A_2, A_3) = \frac{N_{NS}! e^{-\frac{\varepsilon_P^S + \varepsilon_{A_1}^S + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^{NS} + \varepsilon_{P A_1}}{k_B T}}}{(P - 1)! * (A_1 - 1)! * A_2! * A_3! * (N_{NS} - (P - 1) - (A_1 - 1) - A_2 - A_3)!}$$

(Equation 2.3.13)

$$Z(P - 1, A_1, A_2 - 1, A_3) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^S + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^S + \varepsilon_{A_3}^{NS} + \varepsilon_{PA_2}}{k_B * T}}}{(P-1)! * A_1! * (A_2-1)! * A_3! * (N_{NS} - (P-1) - A_1 - (A_2-1) - A_3)!}$$

(Equation 2.3.14)

$$Z(P - 1, A_1, A_2, A_3 - 1) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^S + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^S + \varepsilon_{PA_3}}{k_B * T}}}{(P-1)! * A_1! * A_2! * (A_3-1)! * (N_{NS} - (P-1) - A_1 - A_2 - (A_3-1))!}$$

(Equation 2.3.15)

$$Z(P - 1, A_1 - 1, A_2 - 1, A_3) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^S + \varepsilon_{A_1}^S + \varepsilon_{A_2}^S + \varepsilon_{A_3}^{NS} + \varepsilon_{PA_1} + \varepsilon_{PA_2} + \varepsilon_{A_1A_2}}{k_B * T}}}{(P-1)! * (A_1-1)! * (A_2-1)! * A_3! * (N_{NS} - (P-1) - (A_1-1) - (A_2-1) - A_3)!}$$

(Equation 2.3.16)

$$Z(P - 1, A_1 - 1, A_2, A_3 - 1) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^S + \varepsilon_{A_1}^S + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^S + \varepsilon_{PA_1} + \varepsilon_{PA_3} + \varepsilon_{A_1A_3}}{k_B * T}}}{(P-1)! * (A_1-1)! * A_2! * (A_3-1)! * (N_{NS} - (P-1) - (A_1-1) - A_2 - (A_3-1))!}$$

(Equation 2.3.17)

$$Z(P-1, A_1, A_2-1, A_3-1) = \frac{N_{NS}! e^{\frac{\varepsilon_P^S + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^S + \varepsilon_{A_3}^S + \varepsilon_{PA_2} + \varepsilon_{PA_3} + \varepsilon_{A_2A_3}}{k_B T}}}{(P-1)! A_1! (A_2-1)! (A_3-1)! (N_{NS} - (P-1) - A_1 - (A_2-1) - (A_3-1))!}$$

(Equation 2.3.18)

$$Z(P-1, A_1-1, A_2-1, A_3-1) = \frac{N_{NS}! e^{\frac{\varepsilon_P^S + \varepsilon_{A_1}^S + \varepsilon_{A_2}^S + \varepsilon_{A_3}^S + \varepsilon_{PA_1} + \varepsilon_{PA_2} + \varepsilon_{PA_3} + \varepsilon_{A_1A_2} + \varepsilon_{A_1A_3} + \varepsilon_{A_2A_3}}{k_B T}}}{(P-1)! (A_1-1)! (A_2-1)! (A_3-1)! (N_{NS} - (P-1) - (A_1-1) - (A_2-1) - (A_3-1))!}$$

(Equation 2.3.19)

Here, N_{NS} is the number of non-specific binding sites in the genome. The Boltzmann weights ε_i^S characterize specific binding of protein i to the promoter, while ε_i^{NS} describes non-specific binding to the genomic background. Transcription factor complexes with RNAP and other transcription factors are described by the weights ε_{PA_i} and $\varepsilon_{A_iA_j}$, respectively.

Using Eqs. 2.1.1 – 2.3.19, the probability of polymerase binding (Eq. 2.3.1) can be rewritten as (2.4)

$$p_{bound} = \frac{\sum Z_{bound}}{Z_{tot}} = \frac{1}{1 + \frac{\sum Z_{free}}{\sum Z_{bound}}} = \frac{1}{1 + \frac{1}{F_{Reg}} * \frac{Z(P, A_1, A_2, A_3)}{Z(P-1, A_1, A_2, A_3)}} \approx \frac{1}{1 + \frac{N_{NS}}{P * F_{Reg}} * e^{\frac{\Delta \varepsilon_P}{k_B T}}} \quad (\text{Equation 2.4})$$

In the last step, we calculated the ratio of $Z(P, A_1, A_2, A_3)$ and $Z(P-1, A_1, A_2, A_3)$, and additionally assumed that the number of polymerase and transcription factor molecules is much smaller than the total number of non-specific binding sites in the genome ($N_{NS} \gg P$, $N_{NS} \gg A_1$, $N_{NS} \gg A_2$, $N_{NS} \gg A_3$). This leads to the following approximation (2.5)

$$\frac{Z(P, A_1, A_2, A_3)}{Z(P-1, A_1, A_2, A_3)} = \frac{N_{NS} - (P-1) - A_1 - A_2 - A_3}{P} * e^{\frac{\Delta \epsilon_P}{k_B T}} \approx \frac{N_{NS}}{P} * e^{\frac{\Delta \epsilon_P}{k_B T}} \quad (\text{Equation 2.5})$$

Additionally, we used the notation

$$\Delta \epsilon_P = \epsilon_P^S - \epsilon_P^{NS} \quad (\text{Equation 2.6})$$

and will use similar definitions to describe specific vs. non-specific binding of other proteins below.

The regulation factor may be seen as describing an effective increase (for $F_{reg} > 1$), or decrease (for $F_{reg} < 1$), of the number of RNA polymerase molecules that are available to bind the promoter. The regulation factor in Eq. 2.4 is given by (2.7)

$$F_{Reg} = \frac{\sum Z_{bound}/Z(P-1, A_1, A_2, A_3)}{\sum Z_{free}/Z(P, A_1, A_2, A_3)} \quad (\text{Equation 2.7})$$

Using Eq. 2.3.4-2.3.19, we obtain (2.8)

$$F_{Reg} = \frac{1+c_1+c_2+c_3+c_4+c_5+c_6+c_7}{1+c_8+c_9+c_{10}+c_{11}+c_{12}+c_{12}+c_{14}} \quad (\text{Equation 2.8})$$

Where (2.9.1-2.9.14)

$$c_1 = \frac{Z(P-1, A_1-1, A_2, A_3)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta\varepsilon_{A_1}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_1}}{k_B * T}} \quad (\text{Equation 2.9.1})$$

$$c_2 = \frac{Z(P-1, A_1, A_2-1, A_3)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_2}{N_{NS}} * e^{-\frac{\Delta\varepsilon_{A_2}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_2}}{k_B * T}} \quad (\text{Equation 2.9.2})$$

$$c_3 = \frac{Z(P-1, A_1, A_2, A_3-1)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_3}{N_{NS}} * e^{-\frac{\Delta\varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_3}}{k_B * T}} \quad (\text{Equation 2.9.3})$$

$$c_4 = \frac{Z(P-1, A_1-1, A_2-1, A_3)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta\varepsilon_{A_1}}{k_B * T}} * \frac{A_2}{N_{NS}} * e^{-\frac{\Delta\varepsilon_{A_2}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_1} + \varepsilon_{PA_2} + \varepsilon_{A_1 A_2}}{k_B * T}} \quad (\text{Equation 2.9.4})$$

$$c_5 = \frac{Z(P-1, A_1-1, A_2, A_3-1)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta\varepsilon_{A_1}}{k_B * T}} * \frac{A_3}{N_{NS}} * e^{-\frac{\Delta\varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_1} + \varepsilon_{PA_3} + \varepsilon_{A_1 A_3}}{k_B * T}} \quad (\text{Equation 2.9.5})$$

$$C_6 = \frac{Z(P-1, A_1, A_2-1, A_3-1)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}} * \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_2} + \varepsilon_{PA_3} + \varepsilon_{A_2 A_3}}{k_B * T}}$$

(Equation 2.9.6)

$$C_7 = \frac{Z(P-1, A_1-1, A_2-1, A_3-1)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}} * \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}} * \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_1} + \varepsilon_{PA_2} + \varepsilon_{PA_3} + \varepsilon_{A_1 A_2} + \varepsilon_{A_1 A_3} + \varepsilon_{A_2 A_3}}{k_B * T}}$$

(Equation 2.9.7)

$$C_8 = \frac{Z(P, A_1-1, A_2, A_3)}{Z(P, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}}$$

(Equation 2.9.8)

$$C_9 = \frac{Z(P, A_1, A_2-1, A_3)}{Z(P, A_1, A_2, A_3)} = \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}}$$

(Equation 2.9.9)

$$c_{10} = \frac{Z(P, A_1, A_2, A_3 - 1)}{Z(P, A_1, A_2, A_3)} = \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}}$$

(Equation 2.9.10)

$$c_{11} = \frac{Z(P, A_1 - 1, A_2 - 1, A_3)}{Z(P, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}} * \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}} * e^{-\frac{\varepsilon_{A_1 A_2}}{k_B * T}}$$

(Equation 2.9.11)

$$c_{12} = \frac{Z(P, A_1 - 1, A_2, A_3 - 1)}{Z(P, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}} * \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\Delta \varepsilon_{A_1 A_3}}{k_B * T}}$$

(Equation 2.9.12)

$$c_{13} = \frac{Z(P, A_1, A_2 - 1, A_3 - 1)}{Z(P, A_1, A_2, A_3)} = \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}} * \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\varepsilon_{A_2 A_3}}{k_B * T}}$$

(Equation 2.9.13)

$$c_{14} = \frac{Z(P, A_1 - 1, A_2 - 1, A_3 - 1)}{Z(P, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}} * \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}} * \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\varepsilon_{A_1 A_2} + \varepsilon_{A_1 A_3} + \varepsilon_{A_2 A_3}}{k_B * T}}$$

(Equation 2.9.14)

By lumping constant terms together, we can rewrite Eqs. 2.8 and 2.4 as (2.10) and (2.11) respectively.

$$F_{Reg} =$$

$$\frac{1 + \frac{[A_1]}{K_{A_1}}f_1 + \frac{[A_2]}{K_{A_2}}f_2 + \frac{[A_3]}{K_{A_3}}f_3 + \frac{[A_1][A_2]}{K_{A_1}K_{A_2}}f_1f_2\omega_{12} + \frac{[A_1][A_3]}{K_{A_1}K_{A_3}}f_1f_3\omega_{13} + \frac{[A_2][A_3]}{K_{A_2}K_{A_3}}f_2f_3\omega_{23} + \frac{[A_1][A_2][A_3]}{K_{A_1}K_{A_2}K_{A_3}}f_1f_2f_3\omega_{12}\omega_{13}\omega_{23}}{1 + \frac{[A_1]}{K_{A_1}} + \frac{[A_2]}{K_{A_2}} + \frac{[A_3]}{K_{A_3}} + \frac{[A_1][A_2]}{K_{A_1}K_{A_2}}\omega_{12} + \frac{[A_1][A_3]}{K_{A_1}K_{A_3}}\omega_{13} + \frac{[A_2][A_3]}{K_{A_2}K_{A_3}}\omega_{23} + \frac{[A_1][A_2][A_3]}{K_{A_1}K_{A_2}K_{A_3}}\omega_{12}\omega_{13}\omega_{23}}$$

(Equation 2.10)

$$p_{bound} = \frac{F_{Reg}}{F_{Reg} + K_P} \quad (\text{Equation 2.11})$$

Here, the parameters K_{A_i} and K_P characterize specific vs. non-specific transcription factor or polymerase binding to DNA (2.12.1-2.12.2).

$$K_{A_i} = N_{NS} * e^{\frac{\Delta \epsilon_{A_i}}{k_B * T}} \quad (\text{Equation 2.12.1})$$

$$K_P = \frac{N_{NS}}{P} * e^{\frac{\Delta \epsilon_P}{k_B * T}} \quad (\text{Equation 2.12.2})$$

Protein-protein complexes are described by ω_{ij} and f_i that determine interactions between transcription factors, and transcription factors and RNAP, respectively (2.12).

$$\omega_{ij} = e^{-\frac{\varepsilon_{A_i A_j}}{k_B * T}} \quad (\text{Equation 2.13.1})$$

$$f_i = e^{-\frac{\varepsilon_{P A_i}}{k_B * T}} \quad (\text{Equation 2.13.2})$$

Luciferase expression was modeled using Eq. 2.10, assuming that steady state expression is proportional to the transcription initiation rate. The transcription factor concentrations $[A_i]$ represent the phospho-SMAD and phospho-STAT input into the promoter model, and were modeled as described in the previous section on signalling module (*Eq. 1.1.1-1.1.2, Page 48*).

b) Promoter Variants

During our modelling analyses, we fitted multiple variants of the model to the experimental data. In the following, we will describe the implementation of different model variants before discussing the model fitting strategy and the model selection procedure.

We analyzed eight variants of the promoter model (*Fig. 11*). All model variants contained well-known aspects of promoter regulation such as pSTAT/pSMAD binding to the promoter and RNAP activation by transcription factors (grey arrows in *Fig. 11*). In the full model (topology 8 in *Fig. 11*), we additionally allowed all possible cooperative protein-protein

interactions between transcription factors (indicated by red arrows in *Fig. 11*). The thermodynamic promoter model derived in Equation 2.10 describes this full model.

Let us define two terms as:

$$\begin{aligned} Numerator_{reg} = & 1 + \frac{[pSM]^{nSM}}{K_{B1}} f_{B1} + \frac{[pSM]^{nSM}}{K_{B2}} f_{B2} + \frac{[pST]^{nST}}{K_{ST}} f_{ST} + \frac{[pSM]^{2nSM}}{K_{B1}K_{B2}} f_{B1}f_{B2}\omega_{B1B2} + \\ & \frac{[pSM]^{nSM}[pST]^{nST}}{K_{B1}K_{ST}} \end{aligned}$$

(Equation 2.14.1)

$$Denominator_{reg} = 1 + \frac{[pSM]^{nSM}}{K_{B1}} + \frac{[pSM]^{nSM}}{K_{B2}} + \frac{[pST]^{nST}}{K_{ST}} + \frac{[pSM]^{2nSM}}{K_{B1}K_{B2}} \omega_{B1B2} + \frac{[pSM]^{nSM}[pST]^{nST}}{K_{B1}K_{ST}}$$

(Equation 2.14.2)

Then, in the context of hepcidin promoter regulation, *Eq. 2.10* may be rewritten as (3.1)

$$p_{bound} = \frac{F_{Reg}}{F_{Reg} + K_P} \quad (Equation 3.1)$$

Where,

$$F_{reg} = \frac{Numerator_{reg}}{Denominator_{reg}} \quad (\text{Equation 3.2})$$

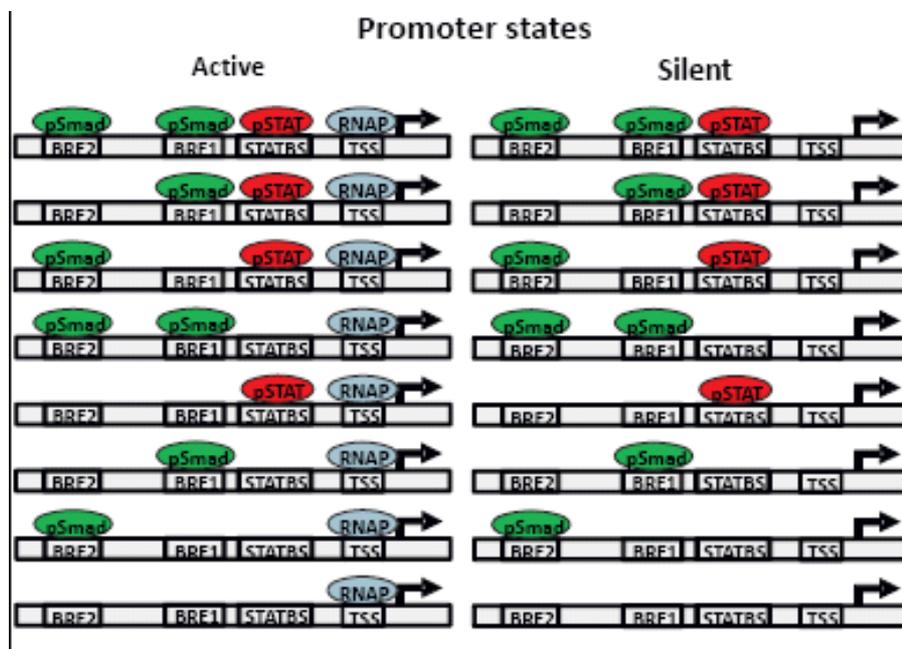


Figure 10: Possible promoter states for applying thermodynamic model. Basis of mathematical modelling of signalling and promoter crosstalk

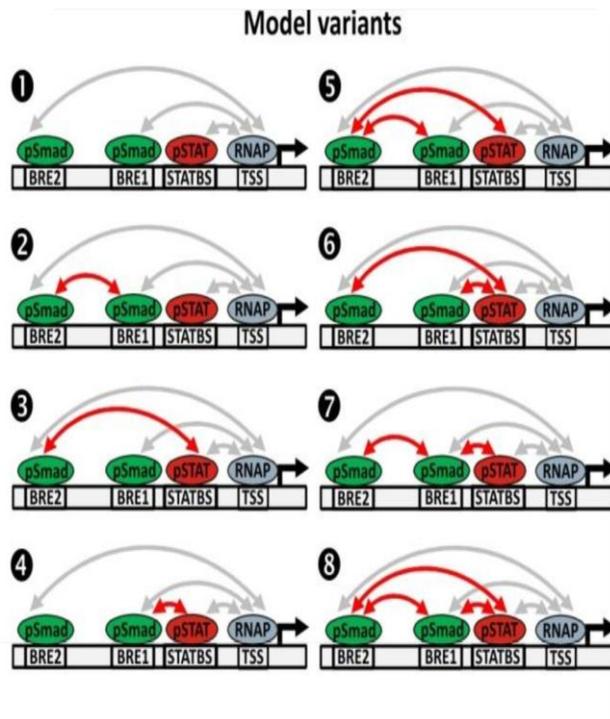


Figure 11: Model selection approach allows for the identification of protein-protein interactions on the promoter. Various model variants were tested for their ability to fit the data in Fig. 6 (Page 58). The minimal model (model 1) assumes that each transcription factor independently activates RNAP (grey arrows), while more complex variants additionally take into account cooperativity among transcription factors (red arrows).

Here, we have replaced the generic transcription factor concentrations A_i by the levels of phosphorylated SMAD and STAT transcription factors, designated as [pSM] and [pST], respectively. Transcription factors of the SMAD and STAT families homo- and hetero-oligomerize upon phosphorylation, and oligomerization is thought to be important for gene expression regulation [77, 78]. Oligomerization would result in a nonlinear relationship between the level of transcription factors and gene expression. To accommodate possible oligomerization reactions we allowed that pSM and pST control hepcidin expression with the exponents n_{SM} and n_{ST} , respectively. The values of n_{SM} and n_{ST} were allowed to vary between 1 and 3 during fitting to reflect dimerization and trimerization reactions. An exponent larger than one for SMAD signalling improved the fitting result, while $n_{ST} > 1$ was not beneficial (see *Inset 2, Page 98*, for best-fit parameters).

The affinity of the transcription factors for the BRE1, the BRE2 and the STATBS is described by the corresponding dissociation constants (K_{B1} , K_{B2} , K_{ST}) (*Eqs. [2.14.1-2.14.2]-3.2, Page 88*). Each transcription factor bound to DNA contacts polymerase, and enhances RNAP recruitment to the TSS with the binding energies f_{B1} , f_{B2} and f_{ST} (denoting the effects of transcription factors bound to BRE1, BRE2 and STATBS, respectively). The half-maximal saturation parameter K_P takes into account that the TSS may be fully occupied by RNAP, implying that a further increase in gene expression is not possible if $F_{Reg} \gg K_P$. Protein-protein interactions among transcription factors on DNA may mutually enhance promoter binding and/or RNAP activation; in the model, such complex formation is described by the binding energies ω_{B1B2} (bridging between transcription factors bound to BRE1 and BRE2), ω_{B1ST} (bridging between BRE1 and STATBS) and ω_{B2ST} (bridging between BRE2 and STATBS).

We also analyzed the ability of simpler model variants (topology 1-7 in *Fig. 11*) to fit the experimental data that allowed us to derive a minimal essential model of hepcidin promoter regulation. These simpler variants contained only a subset of all possible cooperative protein-protein interactions between transcription factors (red arrows in *Fig. 11*). Cooperative interactions were eliminated by fixing the parameters ω_{ij} of the forbidden cooperative protein-protein interaction to unity (*Eq. 3.1, Page 88*). For example, the minimal model variant 1 did not contain any of the hypothetical protein-protein-interactions among transcription factors (i.e., the corresponding interaction parameters in *Eqs. 2.14.1 & 2.14.2(Page 88)* were set to $\omega_{B1B2} = \omega_{B1ST} = \omega_{B2ST} = 1$).

Model fitting to the luciferase data required the simulation of promoter mutants. *Equations 2.14.1 & 2.14.2(Page 88)* describes the wildtype construct, but was adjusted to mutant promoters by eliminating transcription factor binding terms (i.e., by setting $[pSM]/K_{B1}$, $[pSM]/K_{B2}$ and/or $[pST]/K_{ST}$ to zero).

The simple equations (*Eq. 1.1-2.3.19; Eq. 3.1*) are used to define the promoter model in an efficient way. The promoter model (*Eq. 3.1, Page 88; 10 parameters*) was either combined with a signalling model that does not take into account pathway crosstalk (*Eq. 1.1.1-1.1.2, Page 48; 7 parameters; best-fit in Fig. 12*) or with a signalling model considering signalling crosstalk (See *Section 3.1; 9 parameters; best-fit in Figs. 14-15*).

II.3.3 Model Calibration

a) Fitting strategy: All model variants were fitted to the data by minimizing the χ^2 metric ($\chi^2 = (M_i - s_j \cdot D_i)^2 / \sigma_i$, as given by Eq. 3.4 in the next section; where M_i , D_i and σ_i are the simulated value, the measured value and the experimental error, respectively). The model species [pSM] and [pST] were fitted to the transcription factor phosphorylation data, while p_{bound} was fitted to the luciferase measurements. A scaling factor s_j was used to adjust the simulated transcription rate (p_{bound} ; Eq. 3.1, Page 88) to the luciferase data. Transcription factor phosphorylation was formulated in arbitrary units (Eqs. 2.1.1-2.3.3, Page 68; Inset 2, Page 98), and could thus be fitted without the use of a scaling factor (by adjusting $y_{\text{max},1}$ and $y_{\text{max},2}$). All parameters of the model were allowed to vary within a physiologically reasonable range; the ranges as well as the best-fit values are given in Inset 2. The parameters 1-9 belong to the signalling module of the model, while the remaining ones describe promoter regulation. Each parameter was constrained to a physiologically feasible range during fitting to ensure reasonable modelling results. The Hill coefficients of the signalling module (parameters 3 and 7) were restricted to values typical for biochemical response curves. The other parameters of the signalling module (y_{max} , EC_{50}) represent a combination of multiple signalling reaction constants, and were thus constrained such that they match the experimental measurements of transcription factor phosphorylation (Fig. 6, Page 58). Most parameters of the promoter module were taken from literature. Some were allowed to vary over a broad range to accommodate different kinds of qualitative behavior. For example, the wide range of half-maximal promoter saturation constants (K_P ; parameter 10) allows for promoter saturation to occur upon stimulation. Likewise, the K_D values of transcription factor binding to cognate promoter sites (parameters 11-13) were allowed to vary over a broad range to accommodate weak and strong binding. The parameter ranges for the constants describing protein-protein interactions on the promoter (parameters 14-19) represent the typical interaction energies of 1-5 kcal/mol reported in the literature [1]. The exponents n_{SM} and n_{ST} reflect transcription factor dimerization and trimerization, implying that values of up to 3 can be expected.

Parameter optimization was done using a deterministic trust region optimizer in Matlab. In order to circumvent local minima, we repeatedly fitted the model starting from 80.000 quasi randomly distributed positions in the space of allowed parameter ranges. The results of this repeated fitting strategy apparently converged to a global optimum, since ~50% of the fitting runs yielded χ values close to the minimum all 80.000 runs.

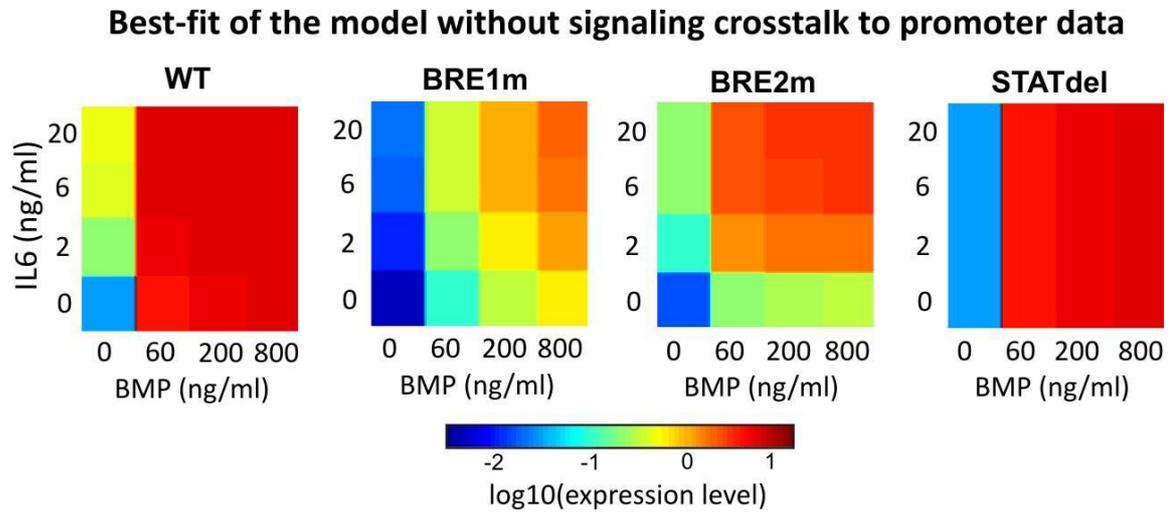


Figure 12: Best-fit of a hepcidin expression model without crosstalk at the level of BMP and IL6 signalling pathways. Luciferase expression was simulated using Eqs. 2.1 and 3.1, and the transcription rate in the model (pbound) were fitted to the data in (using a scaling factor). The best-fit parameter values of this model are given in Inset 2(Page 98).

b) Model selection: To derive a minimal essential model of hepcidin regulation, we employed a model selection approach considering model topologies of different complexity, and systematically compared their ability to fit the experimental observations. Models of different complexities (*Fig. 11*) were compared using the Akaike Information criterion and the Likelihood ratio test. This statistical goodness of fit criteria indicates whether a more complex model with more parameters fits the data significantly better than simpler variants.

Briefly, we used the following formula for the Akaike information criterion (3.3)

$$AIC = \chi^2 + 2 \cdot k \quad (\text{Equation 3.3})$$

Here, k is the number of parameters of that model and χ^2 is given by:

$$\chi^2 = \sum_{i=1}^n \frac{(\text{model value}_i - \text{data value}_i)^2}{\text{data value}_i} \quad (\text{Equation 3.4})$$

Where n is the number of data points. The model with the least AIC was taken as the variant that is most suited to describe the data. The likelihood ratio statistic (3.5)

$$D = \chi_2^2 - \chi_1^2 \quad (\text{Equation 3.5})$$

equals the difference in the χ^2 values of two nested models. Nested means that the larger model (characterized by χ_2^2) must contain the smaller model (characterized by χ_1^2), and can

thus be interconverted into the smaller variant by appropriate parameter choice. D was compared to a tabulated value of the χ^2 statistic using the difference in the number of model parameters as the degree of freedom. If the tabulated values were less than D we rejected H_0 , null hypothesis that the extra parameters are necessary for a better model.

Both statistical measures revealed that a relatively simple model containing a single protein-protein interaction among SMAD and STAT transcription factors bound to BRE1 and the nearby STATBS was sufficient to explain the data (model 4 in *Fig. 11, Page 90*). The best fitting result of the BRE1-STATBS interaction model without signalling crosstalk is shown in *Fig. 12 (Page 95)*. Statistical measures indicate that the model matched the data close to experimental measurement noise ($\chi = 124$, $N = 80$).

c) *Parameter identifiability analysis:* Parameter identifiability was analysed using the strategy proposed by Hengl et al. [79]: Briefly, the parameter vectors of the top 45% fitting results had a similar goodness of fit ($\chi < 135$), and were analysed with respect to parameter ranges and parameter correlations. The robustness of model predictions was estimated by repeatedly simulating predictions for the top 45% of the model solutions. The upper and lower bounds of the computed solution are given as a prediction range in *Fig. 21(Page 119)*.

#	Parameter	Description	Fitting range	Value with signalling crosstalk (Fig. 2C and D)	Value without signalling (Fig. S3)	without crosstalk
1	$y_{\max,1}$	Maximal STAT activation	0.5 – 7 (a.u.)	2.8520	0.2598	
2	$EC_{50,1}$	Half-maximal IL6 concentration for STAT activation	0.5 – 50 (ng/ml)	7.7388	2.5048	
3	n_1	Hill coefficient STAT activation	0.5 - 5	1.0242	2.3717	
4	$Y_{\text{basal},2}$	Basal SMAD activation	0.03 – 0.11 (a.u.)	0.0583	0.0458	
5	$y_{\max,2}$	Maximal SMAD activation	1.5 – 3 (a.u.)	1.9490	33.3224	
6	$EC_{50,2}$	Half-maximal BMP concentration for SMAD activation	20 – 5000 (ng/ml)	140.2440	464.7168	
7	n_2	Hill coefficient SMAD activation	1 - 5	1.4481	1.2969	
8	$k_{C,1}$	Cross-inhibition of STAT signalling by pSMAD	0.01 - 5	0.4135	-	
9	$k_{C,2}$	Cross-inhibition of SMAD signalling by pSTAT	0.01 - 5	0.1285	-	
10	K_P	Half-maximal promoter saturation	1 - 10000	6804.7	5413.9	
11	K_{B1}	K_D of pSMAD binding to BRE1	0.001 - 5000	0.4391	1.6370	
12	K_{B2}	K_D of pSMAD binding to BRE2	0.001 - 5000	16.8738	94.8517	
13	K_{ST}	K_D of pSTAT binding to STATBS	0.001 - 5000	206.3988	339.9842	
14	f_{B1}	Binding of BRE1-bound pSMAD to RNAP	1 - 5000	537.6490	257.3671	
15	f_{B2}	Binding of BRE2-bound pSMAD to	1 - 5000	4972.6	2777.9	

16	f_{ST}	RNAP Binding of STATBS-bound pSTAT to RNAP	1 - 5000	584.75	4691.8
17	ω_{B1B2}	cooperativity between BRE1 and BRE2	1 - 300	1 (fixed)	1 (fixed)
18	ω_{B1ST}	cooperativity between BRE1 and STATBS	1 - 300	5.3869	6.1287
19	ω_{B2ST}	cooperativity between BRE2 and STATBS	1 - 300	1 (fixed)	1 (fixed)
20	n_{SM}	Exponent reflecting pSMAD oligomerization	1 - 3	1.7807	1 (fixed)
21	n_{ST}	Exponent reflecting pSMAD oligomerization	1 - 3	1	1 (fixed)

Inset 2: Best-fit parameters for the models with and without signalling crosstalk (*Figs. 15(Page 107) and 11(Page 95), respectively*).

III. Discussion and analysis of model and approach

III.1 Mathematical modelling of the IL6 and BMP crosstalk at the signalling and promoter levels

Model fitting and statistical evaluation revealed that the data were best explained by a model containing a single cooperative interaction among SMAD and STAT transcription factors (model 4 in *Fig. 11, Page 90*). This minimal essential model corroborated our initial hypothesis concerning a synergism between BRE1 and STATBS promoter elements. Moreover, the analysis supports that hepcidin expression can be described by a standard thermodynamic model that is based on the assumption of a multiplicative mode of transcriptional regulation.

Interestingly, the fitting result (*Fig. 12, Page 95*) suggested that the luciferase data could be fully explained without assuming pathway crosstalk at the level of signal transduction. Moreover, inclusion of signalling crosstalk in the model did not improve the fit to the data when compared to the crosstalk-less formulation (*compare Figs. 12 and 15, Page 109*). Thus, the model predicted that crosstalk at the level of transcription factor phosphorylation is weak or absent.

We confirmed this model prediction by directly measuring signalling crosstalk. Phosphorylation of STAT3 and SMAD1/5/8 was monitored after stimulation with BMP and/or IL6 for 12 h using quantitative immunoblotting. The phosphorylation of STAT and

SMAD transcription factors upon co-stimulation with BMP and IL6 was monitored using immunoblotting with phospho-specific antibodies. Our collaborators from the Muckenthaller group performed two replicate measurements, and the corresponding gels are shown in *Fig. 13*. Bands were quantified by densitometry, and duplicate measurements were merged by multiplying one of the duplicates with a fitted scaling factor. This procedure adjusts the duplicate dose-response curves according to their shape, thereby correcting for differences in arbitrary units between gels. A single scaling factor was used to merge all STAT measurements, and another one for the SMAD measurements. Some experimental errors estimated from scaling were unreasonably small; therefore we assumed a minimal experimental error, based on typical variability in Western Blot measurements (relative error of 5% plus an absolute error value).

On our part, we analysed the immunoblotting data and *Figure 13E* and *12F* show mean and standard deviation of the duplicate experiments. As shown previously in *Section II.III (Eq. 2, Page 67)*, here as well the Hill equation (4)

$$y = \frac{y_{basal} + y_{max} * S^n}{S^n + EC50^n} \quad (\text{Equation 4})$$

expresses phosphorylation (y) as a sigmoidal function of the stimulus (S), and considers basal signalling (y_{basal}), maximal pathway activation (y_{max}), the half-maximal-stimulus ($EC50$) and the Hill coefficient (n) as a measure of steepness of the sigmoidal curve. The co-stimulation data (green) can be fitted well by the Hill equation if it is assumed that the non-canonical stimuli (i.e., BMP in panel E and IL6 in panel F) affect only the maximal activation level (y_{max}) of the dose-response curve.

As expected, analysis of promoter data shows stimulation with BMP or IL6 alone resulted in dose-dependent increases of SMAD1/5/8 and STAT3 phosphorylation, respectively (see data points in *Fig. 13*). Co-stimulation with a saturating dose of IL6 appeared to slightly reduce BMP-induced SMAD1/5/8 phosphorylation (*Fig. 13F, Page 104*), but the effect is not statistically significant (paired t-test). High doses of BMP had significant and stronger crosstalk effects on IL6-mediated STAT3 phosphorylation ($p < 0.001$, paired t-test), but again the effects were moderate and never exceeded a two-fold change (*Fig. 13E*). We conclude that crosstalk at the level of transcription factor phosphorylation is relatively weak. This supports the model prediction that crosstalk at the promoter level dominates over signalling effects.

Modelling including signalling crosstalk

To allow for a quantitative description of signalling and promoter events, we refined our model by simultaneous fitting to the transcription factor phosphorylation and luciferase data. During our experimental analyses, we observed that the IL6 and BMP mutually inhibit each other at the level of transcription factor phosphorylation, although to a minor extent (*Fig. 13*). We therefore extended our initial signalling model, and took inhibitory crosstalk into account.

The main objective of the signalling crosstalk model was to represent the existing data (*Figs. 6 and 12*), and to extrapolate the concentrations of phosphorylated transcription factors for conditions where experimental measurements were not available: STAT phosphorylation was only assessed for increasing doses of IL6 in the presence or absence of saturating amounts of BMP (800 ng/ml). Likewise, SMAD phosphorylation was monitored for increasing doses of BMP, alone or in combination with 25 ng/ml IL6 (*Figs. 13*). Since luciferase expression was additionally assessed for intermediate IL6 and BMP doses, less input than output conditions

are known in the promoter model. To overcome this problem, we estimated transcription factor activity using a model-based extrapolation strategy.

Transcription factor phosphorylation upon co-stimulation

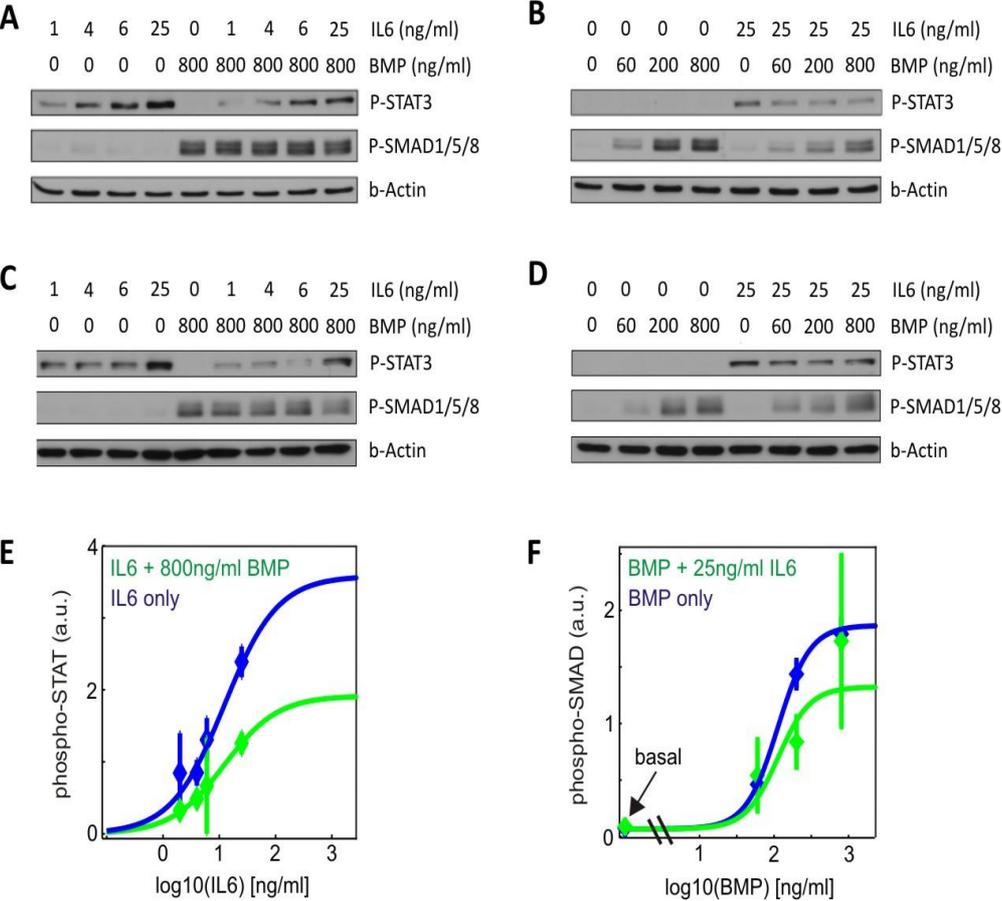


Figure 13: *Moderate inhibitory crosstalk at the signalling level.* (A) - (D) Analysis of crosstalk at the signalling level by immunoblotting against phosphorylated SMAD and STAT. HuH7 cells were stimulated with increasing doses of IL6 in the presence or absence of BMP (A, C) or vice versa (B, D). Actin levels serve as loading controls. Two biological replicates were performed (Replicate 1: panels A and B; Replicate 2: panels C and D). (E) and (F) Quantification of signalling crosstalk. Data points represent mean and standard deviation of densitometric analyses of Western Blots (N =2). Lines are fits of the Hill equation to the data.

A model describing signalling crosstalk should be able to simultaneously describe dose-response curves of pSMAD and pSTAT in the presence or absence of the non-canonical inhibitory stimulus (IL6 and BMP, respectively). To investigate the mode of crosstalk regulation, we fitted the Hill equation to the dose-response data (*Figs. 13E and F*). Different scenarios of signalling crosstalk were analyzed: the non-canonical stimulus (i.e., BMP for STAT and IL6 for SMAD) was assumed to affect dose-response of transcription factor activation at the level of one or more parameters of the Hill equation (y_{basal} , y_{max} , EC50 and n).

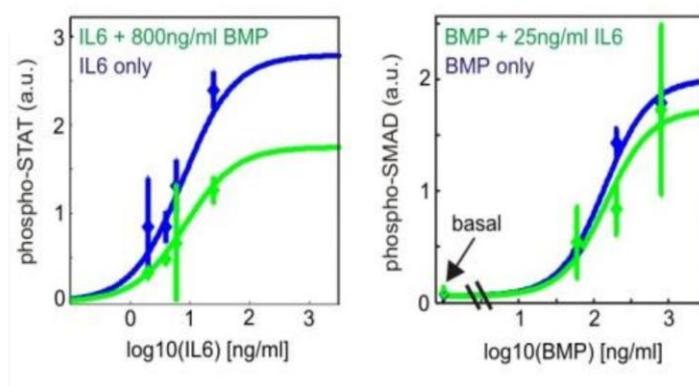


Figure 14: Integrative crosstalk model fits the dose-response curves of transcription factor phosphorylation. Solid lines represent model trajectories in comparison to experimentally measured data points (shown as mean \pm std).

The fits in *Fig. 13E and F* reveal that crosstalk can be described quantitatively if it is assumed that non-canonical stimulation modulates that maximal activation level in the Hill equation (y_{max}). The assumption of selective crosstalk modulation at the level of y_{max} was used to

derive a simple mathematical model for signalling crosstalk as described in the following part of this section (*Fig. 14*).

In a simple crosstalk formulation the concentration of phosphorylated transcription factors can be written as (5.1)

$$[pST] = \frac{y_{max,1}}{1+k_{C,1}*[pSM]} * f_{Hil,IL6} \quad (\text{Equation 5.1})$$

$$[pSM] = y_{basal,2} + \frac{y_{max,2}*f_{Hil,BMP}}{1+k_{C,2}*[pST]} \quad (\text{Equation 5.2})$$

Where

$$f_{Hil,IL6} = \frac{[IL6]^{n_1}}{[IL6]^{n_1} + EC_{50,1}^{n_1}} \quad (\text{Equation 5.31})$$

And

$$f_{Hil,BMP} = \frac{[BMP]^{n_2}}{[BMP]^{n_2} + EC_{50,2}^{n_2}} \quad (\text{Equation 5.3.2})$$

For each transcription factor, the maximal activation level is reduced by the presence of the opposite factor (with crosstalk strength constants $k_{C,1}$ and $k_{C,2}$). We have taken into account that basal STAT phosphorylation is negligible in our experimental setup ($y_{basal,1} = 0$).

We can solve for pSTAT by plugging in the pSMAD concentration, and obtain (5.4)

$$\begin{aligned}
 [pST] = & \\
 & \frac{1}{2} * \frac{1}{k_{C,2}} * \\
 & \left(\frac{k_{C,2} * y_{max,1} * f_{Hill,IL6} - (1 + k_{C,1} * (y_{basal,2} + y_{max,2} * f_{Hill,BMP}))}{1 + k_{C,1} * y_{basal,2}} + \right. \\
 & \left. \sqrt{\left(\frac{k_{C,2} * y_{max,1} * f_{Hill,IL6} - (1 + k_{C,1} * (y_{basal,2} + y_{max,2} * f_{Hill,BMP}))}{1 + k_{C,1} * y_{basal,2}} \right)^2 + 4 * k_{C,2} * \frac{y_{max,1} * f_{Hill,IL6}}{1 + k_{C,1} * y_{basal,2}}} \right)
 \end{aligned}$$

(Equation 5.4)

As expected, this expression simplifies to the normal Hill equation

$$[pST] = y_{max,1} * f_{hill,IL6}$$

in the absence of SMAD signalling (i.e., if $y_{basal,2} = 0$ and $y_{max,2} = 0$).

This minimal crosstalk model was sufficient to accurately and simultaneously describe the phospho-STAT and phospho-SMAD dose-response curves (*Figure 14* shows the fit of *Eqs. 5.1* and *5.4* to the Western Blot data). The crosstalk model could extrapolate transcription factor phosphorylation for intermediate levels of BMP and IL6 that were not measured experimentally. Thus, the signalling crosstalk provided a complete input map for the promoter module. The promoter level was thus described using the model topology containing a single cooperative interaction between BRE1 and STATBS (variant 4 in *Fig. 11*). The best fitting result model is shown in *Fig. 15*. The model described the data with accuracy close to experimental measurement noise ($\chi = 124$, $N = 80$).

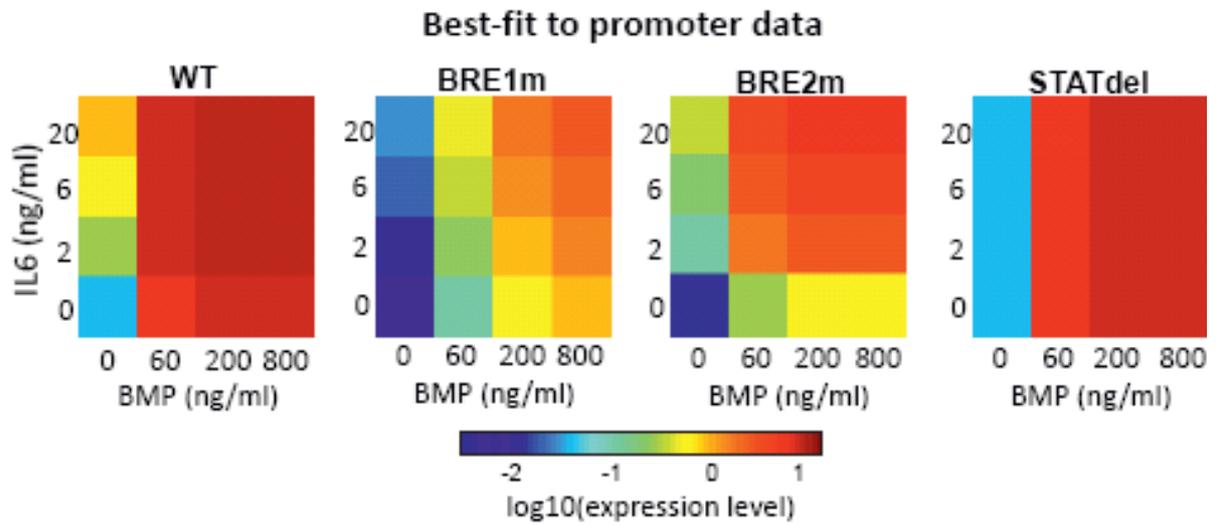


Figure 15: Integrative crosstalk model fits promoter data from luciferase assay. The thermodynamic promoter model (topology 4 in Fig. 10, Page 89) was coupled to a simple signalling model describing inhibitory crosstalk between phospho-SMAD and phospho-STAT transcription factors (Fig. 14). The simulated luciferase activities in the heatmaps agree well with the corresponding data in Fig. 6.

III.2 Experimental validation of model-predicted mechanisms of hepcidin promoter regulation

As a next step, we sought to verify our model by an independent set of experiments not used for model calibration. To this end, we first analyzed whether the parameters of our model were sufficiently constrained by the experimental data used for model fitting ('parameter identifiability analysis'). In case of insufficiently constrained parameters, the model may not be able to generate reliable predictions for previously untested conditions. The analysis revealed that most (but not all) predictions could be made with reasonable accuracy. Based on our parameter identifiability analysis, we formulated all model predictions to be discussed in the following as a range of expected behaviors, not as a single predicted value (*Figs. 16 and 17*). The formulation and verification of our model predictions was focused on double mutant promoters (*Fig. 16*), since these are experimentally accessible and allow for confirmation of predicted promoter mechanisms.

One central promoter mechanism predicted by the model is the cooperative interaction between pSMAD and pSTAT transcription factors bound to BRE1 and STATBS, respectively. The double mutant promoter lacking functional BRE1 and STAT elements (BRE1mSTATdel promoter, *Fig. 16*) was employed to confirm the cooperativity effect. The model predicted that the combined mutation of BRE1 and STATBS should reduce expression less than the sum of the individual mutation effects: Each individual mutation can be considered to independently eliminate the cooperative enhancement of transcription by the two sites, while the combined mutation does so only once.

Co-stimulation of double mutants

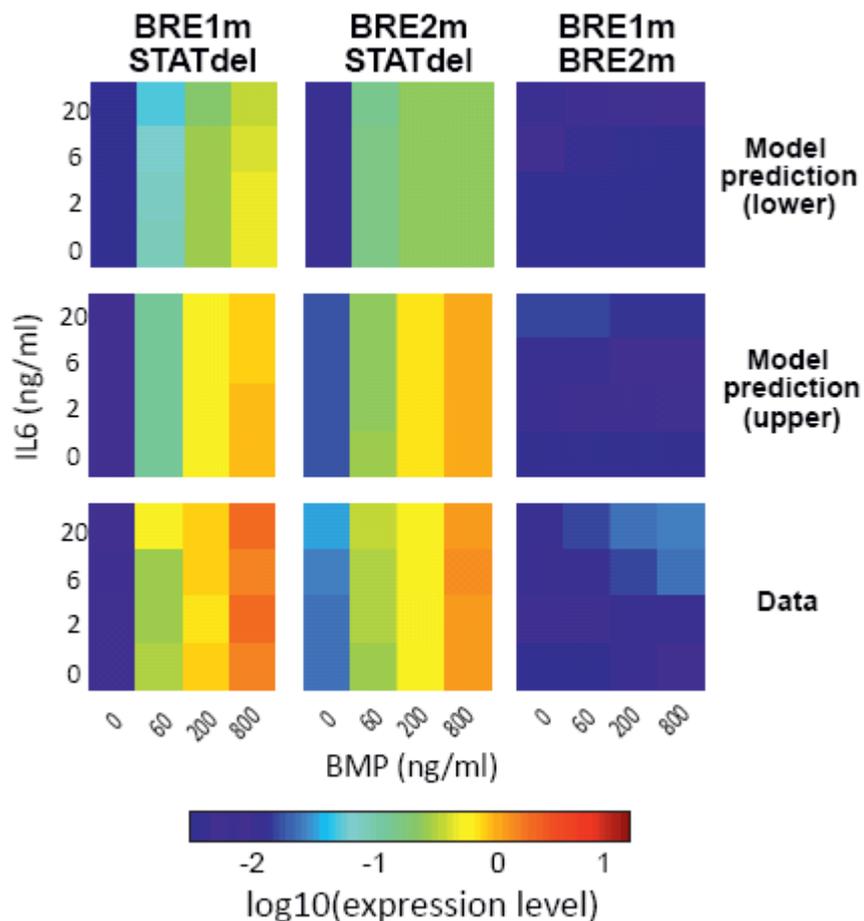


Figure 16: *Co-stimulation response of double mutant promoters.* Confirms that the isolated BRE1 and BRE2 behave similarly. This figure shows heatmaps of luciferase activity under co-stimulation conditions along with the range of model predictions from ‘lower’ to ‘upper’. Data points show the mean values and the two rows of model predictions indicate the range of variation.

BRE dose-response

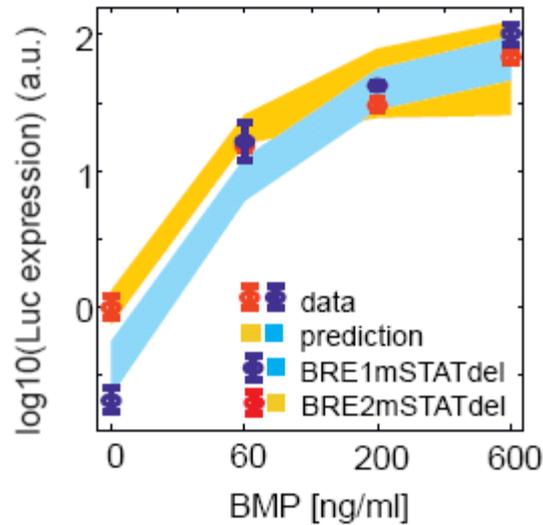


Figure 17: *Model prediction range.* The shaded corridor in the figure represent model predictions and show measurement-compliant parameter sets with highest and lowest predicted effects. Data and model were normalized to basal luciferase expression in the BRE2mSTATdel construct.

As predicted by the model, we found that the reduction of expression in BRE1mSTATdel relative to WT is less than the sum of the individual BRE1 and STATBS mutation effects (*Fig. 18*). This finding supports a cooperative interaction of binding sites. Interestingly, we find that the effect of a BRE1 single mutation is similar to the combined BRE1-STATBS mutation at low IL6 doses (compare green and red bars, *Fig. 18*). STATBS thus acts as a ‘helper site’ that hardly affects expression in a BRE1 mutant, yet it affects expression if the cooperating BRE1 site is present (blue bars, *Fig. 18*). At higher IL6 doses, STATBS tends to affect the expression even in the absence of the cooperating site, because the single BRE1 and double BRE1-STATBS mutation effects are significantly different. The role of STATBS thus

shifts from a pure helper site to an independent site that enhances transcription on its own. The model explains this shift by increased STAT3 binding to the STATBS at higher IL6 doses: low-level STAT3 binding is sufficient to establish the high-affinity cooperation with BRE1, while stronger STAT3 binding is required to directly recruit RNAP with low affinity. These data explain our previous observation that the BRE1m promoter fails to respond to IL6 at low but not high doses (*Fig. 15, Page 109*). BRE1 and BRE2 play different roles in hepcidin expression with respect to basal expression, BMP inducibility and co-stimulation response. One difference between the two sites is the above-mentioned cooperative interaction between BRE1 and STATBS. How do the BMP-responsive elements differ beyond this interaction? The model predicted that BRE1 has higher affinity for phosphorylated SMAD than BRE2, explaining why BRE1 plays a predominant role under basal conditions. Upon sufficiently strong BMP stimulation both sites are predicted to activate RNAP with comparable efficiency. In conclusion, the model suggested that BRE1 and BRE2 should behave similarly in the absence of cooperative promoter interactions. This prediction can be tested by co-stimulation of BRE2mSTATdel and BRE1mSTATdel promoters that solely contain BRE1 and BRE2, respectively (*Fig. 16, Page 111*). The experimental data was in good qualitative agreement with model predictions: Both mutants showed very similar co-stimulation heatmaps and primarily responded to BMP stimulation (*Fig. 16, bottom row*). Maximal luciferase activity at high BMP levels was comparable for both constructs, indicating that BRE1 and BRE2 indeed drive RNAP activation with similar efficiency (*Figs. 16 and 17*). Basal activity was approximately 10-fold higher in the BRE2STATdel promoter, suggesting that the isolated BRE1 has indeed a higher pSMAD affinity than BRE2 (*Fig. 17*).

Quantitative model predictions for the BRE2mSTATdel and BRE1mSTATdel heatmaps were only possible up to a certain range of absolute luciferase activities owing to non-identifiability of model parameters (*Fig. 23, Page 123, top and middle row*). The experimentally observed luciferase activities were within the predicted range, which further supports the validity of our

model (Fig. 15, Page 109). In the co-stimulation heatmap of the WT promoter, we observed that BMP and IL6 regulate hepcidin expression in a less than multiplicative manner (Fig. 6, Page 58; Fig. 7, Page 62). The model suggested that sub-multiplicative regulation arises from saturating RNAP binding to the promoter: BMP alone elicits near-complete occupancy of the TSS with RNAP, implying that IL6 co-stimulation cannot enhance expression much further. Accordingly, the transcription factor binding sites mediating BMP and IL6 responsiveness are predicted to recruit RNAP in a highly redundant manner upon co-stimulation.

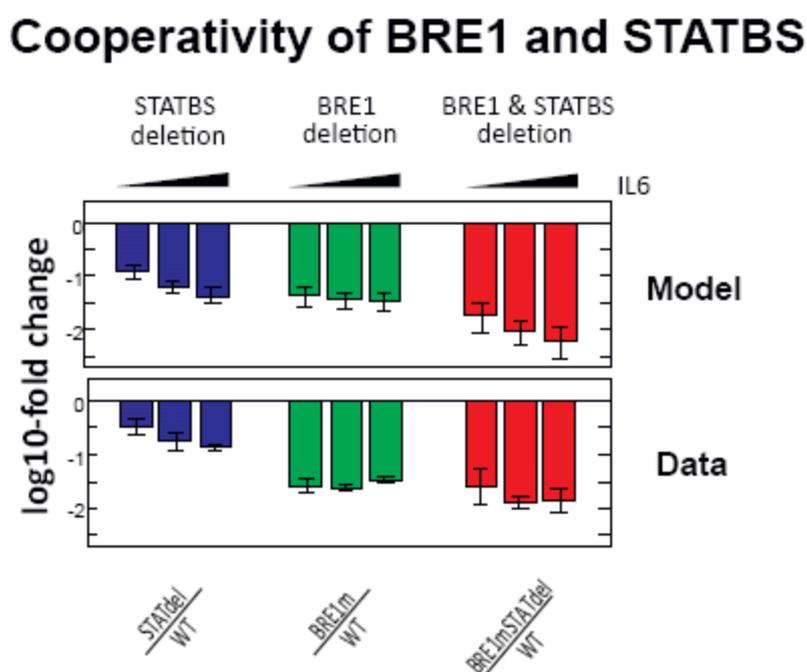


Figure 18: *BRE1 and STATBS cooperativity effect.* Systematic analysis of transcription factor binding site deletion effects confirms cooperativity of BRE1 and STATBS. The impact of binding site deletions was calculated by taking the luciferase activity ratios of different promoters (indicated on the bottom) and expressed as a log₁₀-fold change (y axis). Data points are mean and standard deviation, and model predictions represent the range of measurement-compliant parameter sets.

The model further predicted that similar redundancy also occurs between BRE1 and BRE2 elements, provided that BMP stimulation is sufficiently strong. In the following, we will present independent experimental evidence for the functional redundancy of BRE1 and BRE2, as the model predicted the strongest effects between these two sites. A similar analysis for BRE2 and STATBS can be found in the following figure (*Fig. 19*) where impact of binding site deletions was calculated by taking the expression ratios of different promoter constructs (indicated on the bottom) and expressed as a log₁₀-fold change (y axis). The analysis was restricted to co-stimulation conditions, where BRE2 and STATBS are both occupied. The combined deletion of BRE2 and STATBS (red) has stronger impact than the sum of individual deletion effects (green and blue, left). This supports that BRE2 and STATBS are functionally redundant and serve as buffers. The buffering effect is, however, less pronounced in the data than predicted by the model. The degree of buffering is directly visible if one compares the STATBS deletion effect in the WT construct (left blue bars) with the STATBS deletion effect in a mutant promoter that lacks the buffering BRE2 site (right blue bars). Similarly, the buffering in opposite direction can be assessed by comparing the BRE2 deletion effect in the WT construct (left green bars) with the BRE2 deletion effect in a mutant promoter that lacks the buffering STATBS (right green bars).

Redundancy is expected to reduce the impact of single transcription factor binding site deletions owing to buffering effects of backup sites. We used BRE1mBRE2m double mutant promoter to confirm redundancy, since buffering effects should no longer compensate for the combined deletion of multiple sites. *Figure 20* compares the expression of BRE1m, BRE2m and BRE1mBRE2m promoters relative to WT (*Fig. 20*). In line with model predictions, we find that the combined deletion of BRE1 and BRE2 affects expression much more than the sum of the corresponding single deletions (*Fig. 20, red bars*). This confirms the functional redundancy of the binding sites.

Functional redundancy of BRE2 and STATBS

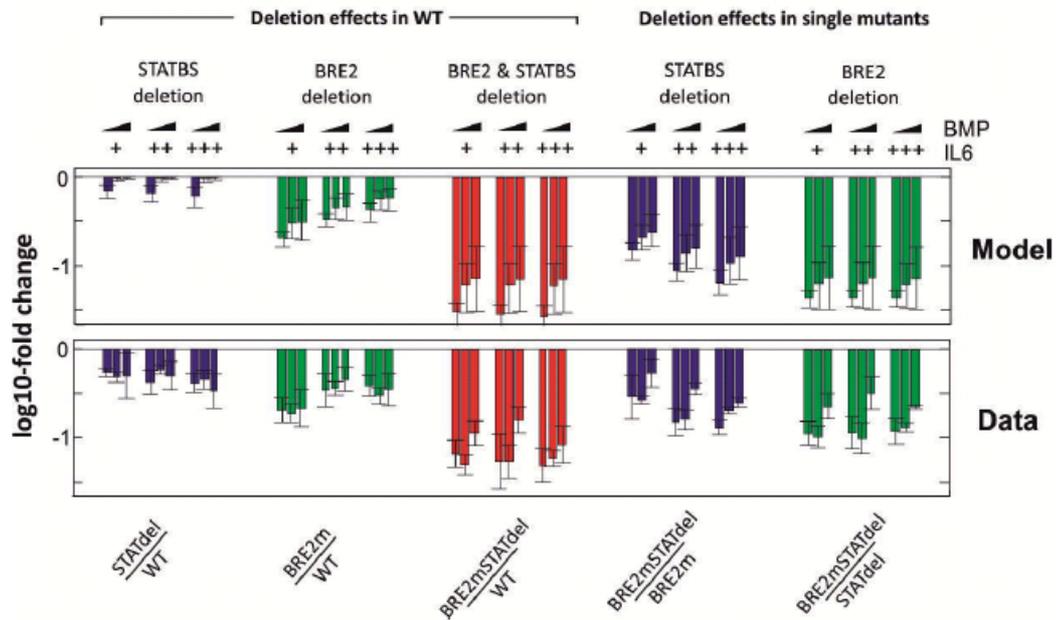


Figure 19: *Redundancy of BRE2 and STATBS.* Systematic analysis of transcription factor binding site deletion effects supports promoter saturation and redundancy of BRE2 and STATBS. The impact of binding site deletions was calculated by taking the expression ratios of different promoter constructs (as indicated along the y-axis) and expressed as log10 fold-change. Data points are mean and standard deviation, and model predictions represent the range of measurement-compliant parameter sets. The combined deletion of BRE2 and STATBS (red bars) are functionally redundant and act as buffers.

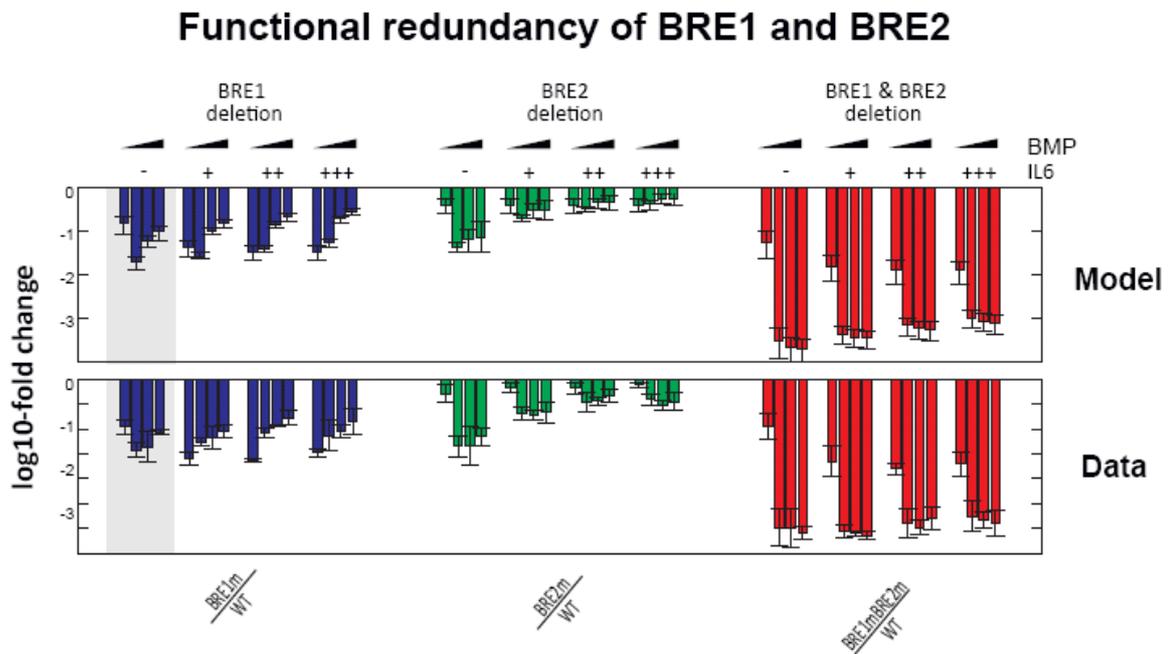


Figure 20: *Redundancy of BRE1 and BRE2.* Systematic analysis of transcription factor binding site deletion effects confirms promoter saturation and redundancy of BRE1 and BRE2. Concepts similar to *Figure 17.* A range of BMP stimulation conditions were considered to ensure visible contribution of both BRE1 and BRE2.

The validity of our model is further supported by the quantitative agreement of the model's predictions with the data: the buffering tends to be more pronounced the stronger the (co-)stimulation strength (*Fig. 20*). This is consistent with promoter saturation being an important aspect of buffering. Non-specific saturation in RNAP binding to the promoter would also suggest that buffering effects occur for binding site combinations other than BRE1 and BRE2. Accordingly, we find similar, albeit less pronounced, buffering effects between STATBS and BRE2 (*Fig. 19*). Taken together, the promoter response to binding site deletions is consistent with binding site redundancy arising from saturating RNAP binding to the TSS. Promoter saturation effects and the differential affinity of BRE1 and BRE2 for phospho-SMAD explain why BRE1 mutations affect expression most strongly at intermediate BMP concentrations (*Fig. 19*, blue bars): BRE1 is incompletely occupied by phosphorylated SMAD transcription factors under basal conditions. BMP stimulation promotes complete phospho-SMAD binding to BRE1, explaining why the impact of BRE1 mutations initially increases for increasing BMP doses. Stronger BMP stimulation results in phospho-SMAD binding to the low-affinity BRE2 site and promotes redundancy, implying a reduced impact of BRE1 single mutations. These data explain why BRE1 primarily affects the promoter inducibility by intermediate BMP stimuli, while it is dispensable for the responsiveness to high BMP doses.

We conclude that a superposition of relatively simple crosstalk effects fully explains interdependent regulation of hepcidin expression by BMP and IL6. The two BMP-responsive elements are functionally redundant and show very similar behavior in isolation, but are differentially modulated by a cooperative interaction with the STATBS.

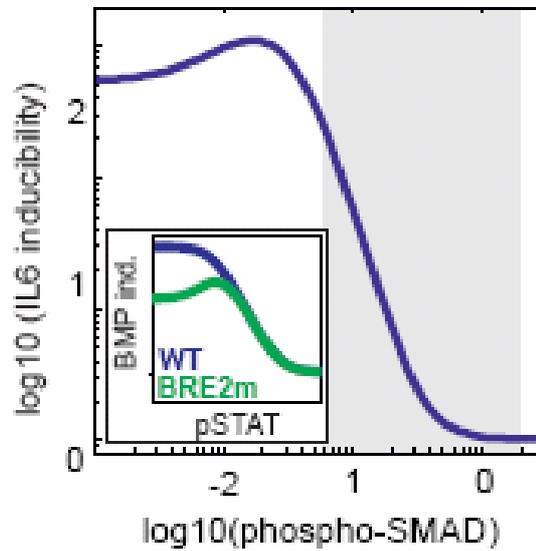


Figure 21: *Systems properties of hepcidin expression:* BMP signalling pathway activity is required for optimal IL6 responsiveness of the hepcidin promoter. The IL6 inducibility, defined as the maximal fold expression change by IL6 (over basal), is shown as a function of phospho-SMAD levels (best-fit WT model). The grey corridor indicates range of phospho-SMAD levels in HuH7 cells. The inset shows how the BMP inducibility of WT and BRE2m promoters is in turn affected by the STAT signalling pathway activity.

III.3 Analysis of signal integration in the promoter model

Thermodynamic modelling showed that hepcidin expression could be quantitatively understood as a superposition of weakly cooperative transcription factor binding and strong promoter saturation with RNA polymerase. The net result of this superposition under our experimental conditions is less-than multiplicative regulation of hepcidin expression by BMP and IL6 (*Fig. 7, Page 62*). Interestingly, the best-fit model suggested that the system can also attain a regime where SMAD and STAT signalling pathways control hepcidin synergistically (i.e., more than multiplicatively): *Figure 21* shows that phospho-SMAD signalling may enhance the IL6 inducibility of the promoter in the regime of very weak BMP signalling. Thus, a synergism between signalling pathways can be observed as long as promoter saturation is negligible. Synergism in the model required the phospho-SMAD concentration to be below the basal level in HuH7 cells (grey shading in *Fig. 21*). Our collaborators performed siRNA-mediated knockdown of the co-factor SMAD4 to lower the basal activity of the BMP pathway. We find that SMAD4 siRNA indeed lowers the IL6 inducibility of the WT hepcidin promoter (Mleczko-Sanecka et al., unpublished observation). This qualitatively supports the model prediction that low-level BMP signalling enhances the IL6 responsiveness of the promoter

As a next step, we addressed in silico whether IL6 signalling in turn affects the BMP inducibility of the promoter (*Fig. 21, inset*). We find that the BMP inducibility is never enhanced by the IL6 pathway (inset, blue line). Pathway synergy is thus an asymmetric phenomenon, and the model explains this observation as follows: BMP stimulation enhances the IL6 inducibility of the promoter owing to the cooperative BRE1-STATBS interaction. Conversely, the impact of IL6 signalling on the BMP inducibility is less pronounced, since the redundant BRE2 functions independently of STATBS. Accordingly, we find in silico that

the BMP inducibility of a BRE2-less promoter partially depends on IL6 signalling (*Fig. 21*, inset, green line). Thus, the presence of two BMP-responsive elements and only one STAT-binding site may establish a stimulus hierarchy, where the iron-sensing BMP pathway dominantly regulates expression. However, in the model this hierarchy was not very pronounced and could only be observed in a narrow range of phospho-SMAD levels (*Fig. 21*). This suggested that the promoter design with two redundant BMP-responsive elements might be more relevant for aspects other than the integration of BMP and IL6 signals.

We investigated *in silico* how the presence of two BREs affects the BMP mono-stimulation response. To this end, we compared BMP dose-response curves of WT, BRE1m and BRE2m promoters *in silico* (*Fig. 22*). Given the functional redundancy of BRE1 and BRE2, it was not surprising that either site is sufficient to confer (partial) BMP responsiveness to the promoter. However, the dose-response curve of the WT promoter was much steeper than that of BRE-mutated constructs: Given a certain increase in phospho-SMAD levels, the WT promoter responds with a larger fold-change in expression when compared to BRE1m and BRE2m promoters (*Fig. 22*). This suggests that the presence of two BMP-responsive elements establishes a highly sensitive response towards BMP stimulation (*Fig. 21*). Such a high BMP sensitivity may allow the iron-BMP signalling axis to sense minor changes in iron blood levels, and to maintain systemic iron homeostasis.

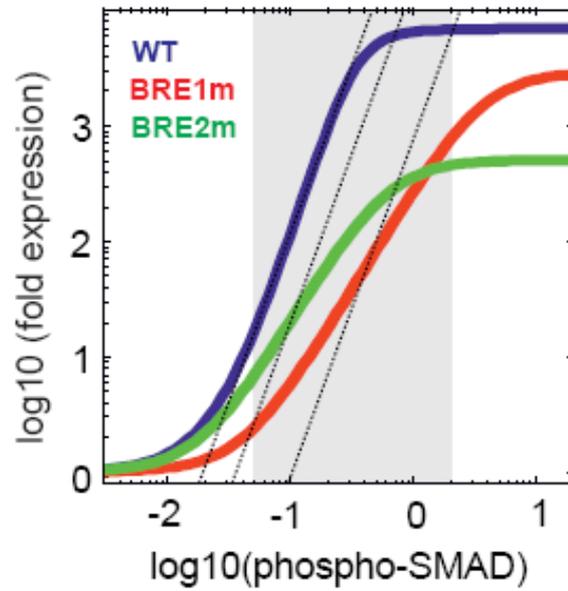


Figure 22: *Systems properties of hepcidin expression:* Hepcidin expression (fold over basal) is shown as a function phospho-SMAD level for the WT, BRE1m, and BRE2m promoter (phospho-STAT was assumed zero). The dashed lines indicate the maximal steepness of the WT dose-response. Grey corridor same as in *Fig. 21*.

The hepcidin promoter contains a single STAT-binding site as opposed to two BREs. Accordingly, the model predicts that the dose-response curve of IL6 is much shallower than that of BMP (*Fig. 23*). We confirmed this model prediction by performing detailed dose-response measurements with multiple doses of IL6 and BMP, respectively (*Fig. 24*). These analyses qualitatively support that the hepcidin promoter is much more sensitive towards BMP stimulation when compared to IL6. In quantitative terms, the steepness of the experimentally observed BMP dose-response curve is, however, less pronounced than predicted by the model.

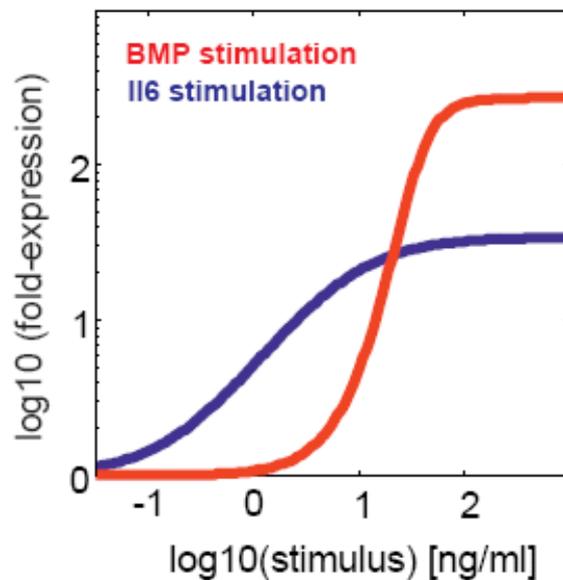


Figure 23: *Systems properties of hepcidin expression:* The luciferase activity (fold over basal) is plotted as function of the IL6 (blue) or BMP (red concentration) concentration. The figure shows simulations of the best-fit model (*Fig. 14, Page 105*)

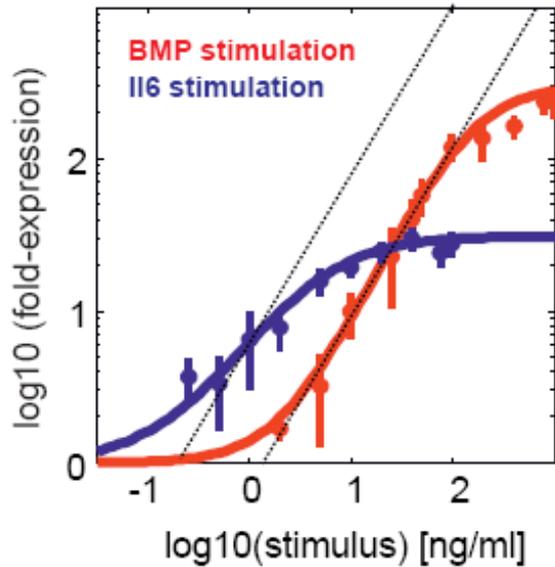


Figure 24: *Systems properties of hepcidin expression:* The luciferase activity (fold over basal) is plotted as function of the IL6 (blue) or BMP (red concentration) concentration. The figure contains experimental data ($n = 3-6$) and fits of the Hill equation (solid lines). Dashed lines indicate the maximal steepness of the BMP response.

One possible explanation is that cellular heterogeneity reduces the steepness of the dose-response curve in population-based reporter gene assays [80]. We conclude that the promoter design with two BMP-responsive elements as opposed to one STATBS specifically confers high sensitivity towards BMP stimulation. In the next step, hepcidin expression regulation by BMP and IL6 in the model described using the best-fit parameters (*Fig. 12, Page 95*), leads us to quantitatively analyze how the hepcidin promoter architecture affects systemic iron homeostasis.

III.4 Description of detailed homeostasis model

Systemic iron homeostasis is maintained by an auto-regulatory negative feedback loop. Iron overload triggers hepcidin expression, and hepcidin in turn lowers intestinal iron influx (*Fig. 25*). Previous theoretical studies showed that negative feedback is particularly efficient in mediating homeostasis if the sensitivity of the feedback is sufficiently high (reviewed in [65]). We therefore reasoned that the high BMP sensitivity of the WT promoter might be essential for maintaining systemic iron homeostasis. To confirm this hypothesis, iron homeostasis in the living animal was simulated using an extended model with feedback (*Fig. 25*). Iron blood levels were described by the species Fe_b , whose levels are controlled by influx and efflux reactions. The iron influx rate is proportional to the intestinal iron concentration (species Fe_i). Iron blood levels control the activity of the BMP signalling pathway, and thus hepcidin expression. Negative feedback regulation was considered in the model by assuming that the iron influx is negatively influenced by hepcidin.

We sought to investigate how the experimentally verified model of signal integration at the hepcidin promoter affects systemic iron homeostasis. We therefore extended the homeostasis model (*Eq. 1.1.1-1.1.2*, Page 48) to (6.1-6.2)

$$\frac{d[Fe_b]}{dt} = \frac{k_{influx}}{1+k_{FB} \cdot [hepcidin]} \cdot [Fe_i] - k_{efflux} \cdot [Fe_b] \quad (\text{Equation 6.1})$$

$$\frac{d[hepcidin]}{dt} = k_{induced} * p_{bound} - k_{deg} \cdot [hepcidin] \quad (\text{Equation 6.2})$$

Here, p_{bound} equals the probability being bound to the transcription start site according to the best-fit promoter and signalling crosstalk model (Eqs. 3.1, 3.2 (Page 88) and 5.1 (Page 106)).

The best-fit promoter and signalling crosstalk model describes p_{bound} as a function of BMP and IL6 concentrations. In mammals, the iron blood levels control the activity of the BMP signalling pathway. For simplicity, we assumed that the effective BMP concentration in the body is proportional to the iron blood level. Thus, the BMP concentration entering p_{bound} was replaced by the species $[\text{Fe}_b]$ in the detailed model of systemic iron homeostasis. For the simulations in Fig. 26, we chose the following kinetic parameters ($k_{\text{influx}} = 1$; $k_{\text{FB}} = 100$; $k_{\text{efflux}} = 1$; $k_{\text{induced}} = 1$; $k_{\text{deg}} = 1$). The IL6 concentration was set to zero in most simulations, and assumed to be saturating in the dashed blue line in Fig. 26.

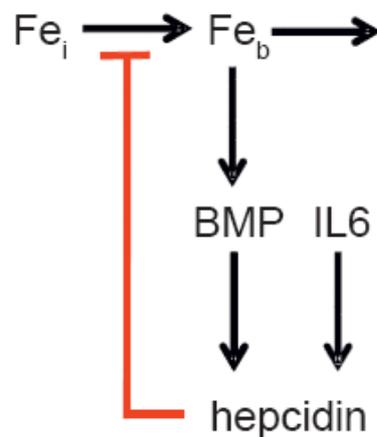


Figure 25: *Conceptual model of iron homeostasis:* Extended mathematical model describing negative feedback regulation of iron blood levels by hepcidin in vivo. Iron blood levels (Fe_b) are controlled by influx and efflux reactions, and the iron influx rate is proportional to the intestinal iron concentration (species Fe_i).

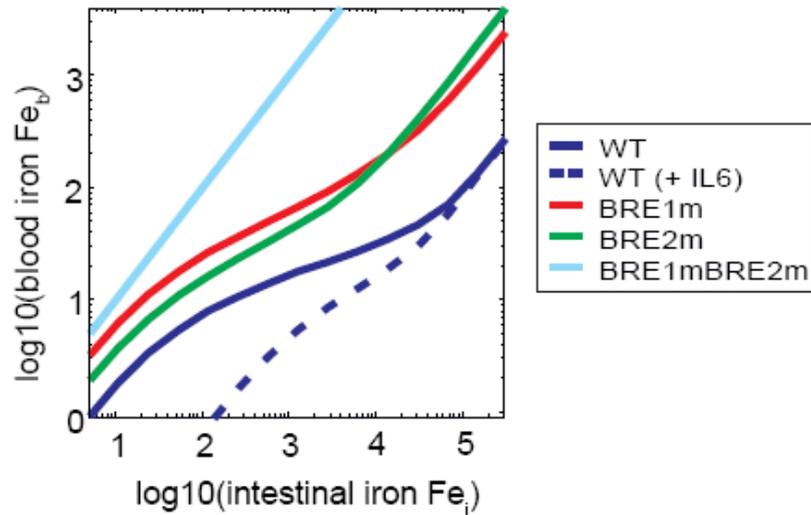


Figure 26: *Systems properties of hepcidin expression:* Iron homeostasis requires two BMP-responsive elements and is abolished by inflammatory stimulation. The extended model (*Fig. 25*) was used to simulate how iron blood levels respond to changes in the intestinal iron concentration.

The detailed model of systemic iron homeostasis thus describes how promoter mutations and IL6 co-stimulation affect the performance of the iron homeostasis loop. Fig. 26 shows corresponding simulations for one particular set of kinetic parameters. It should be noted that similar conclusions concerning the modulation of iron homeostasis continue to hold for other parameter values of k_{influx} , k_{FB} , k_{efflux} , k_{induced} and k_{deg} . This is because these parameter values affect the absolute iron blood levels in the model, but not the qualitative features of homeostasis.

Qualitatively, homeostasis was analyzed by assessing how the iron blood levels in the model (Fe_b) respond to changing intestinal iron concentrations (Fe_i). This scenario mimics a change in the diet iron content. The simulations in *Fig. 26* show that a model with the WT hepcidin promoter efficiently maintains systemic iron homeostasis, as the iron blood levels remain essentially constant over a broad range of intestinal iron concentrations. Models with BRE1m and BRE2m promoters perform less well, as the perturbation-response curves are steeper and homeostasis is restricted to a narrower range of influx rates (*Fig. 26*, green and red curves). This suggests that the simultaneous presence of two BMP-responsive elements in the promoter indeed optimizes the performance of the systemic iron homeostasis loop.

One important question is why IL6 stimulation reduces iron blood levels and induces anaemia of inflammation even though the homeostasis feedback works efficiently. At a first glance, one might expect that the auto-regulatory feedback loop effectively buffers IL6-induced perturbations in hepcidin expression. Nevertheless, simulations of the extended feedback model show strongly diminished iron levels and a loss of homeostasis if high IL6 levels are assumed (*Fig. 26*, blue dashed line). We analyzed the dose-response behavior of the best-fit promoter model (*Fig. 12, Page 95*) to understand this effect. The simulations reveal that increasing IL6 levels not only increase hepcidin expression, but also reduce the sensitivity of the BMP dose-response (*Fig. 27*). Moreover, significant changes in hepcidin expression are restricted to a narrower range of phospho-SMAD levels. Thus, IL6 stimulation impairs the performance of the auto-regulatory homeostasis loop, and thereby efficiently lowers iron blood levels. We conclude that the promoter saturation by BMP and IL6 observed in cultured cells may help to explain the loss of iron homeostasis in anaemia of inflammation.

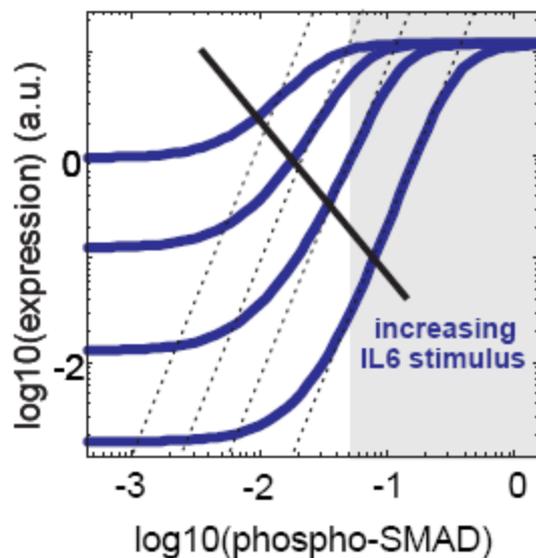


Figure 27: *Systems properties of hepcidin expression:* The best-fit model (Fig. 14, Page 105) was employed to simulate how increasing IL6 stimulation affects the BMP dose-response curve of the promoter. Dashed lines indicate the maximal slope in the absence of IL6. Grey corridor same as in Fig. 21.

On the other hand, iron blood levels are chronically elevated in hereditary hemochromatosis, in most cases due to inactivating mutations in the iron-sensing BMP signalling axis. This gives rise to another unexplored question which is why HH is commonly associated with inactivating mutations in the SMAD signalling pathway, while mutations in the BRE1 promoter element are rare and BRE2 mutations are not associated with disease. We analyzed how iron blood levels (Fe_b) change upon mutational inactivation of BREs to understand iron overload associated with HH (Fig. 26). Our iron homeostasis model predicts that a deletion of BRE1 affects the iron blood levels more strongly than the BRE2 deletion, at least in the range where homeostasis occurs in WT (compare green and red lines in Fig. 26, respectively). The

model explains the more critical role of BRE1 at low iron levels with its higher affinity for the phosphorylated SMAD transcription factors when compared to BRE2. The differential phospho-SMAD affinity may thus explain why only BRE1 mutations have been associated with HH. The model further predicts strong redundancy of BRE1 and BRE2: Single mutations in either site have much weaker effects than a complete feedback ablation by a BRE1mBRE2m double mutation (light blue line in *Fig. 26*). The more-than-additive effect of a BRE double deletion in cultured cells (*Fig. 20, Page 117*) is thus predicted to be preserved in vivo. These simulations may explain why BMP signalling pathway mutations that simultaneously inactivate expression regulation via BRE1 and BRE2 are by far the most common cause of HH.

Taken together, it seems that the steepness of the hepcidin promoter response is a key parameter controlling how well the systemic iron homeostasis loop compensates for fluctuations in iron diet content. Our results indicate that the hepcidin promoter is no longer able to efficiently sense changes in iron blood levels in clinically relevant iron disorders. This leads to a breakdown of the homeostasis loop, which is accompanied by a strong drop in the absolute iron blood levels.

IV. Outlook and scope of work done

IV.1 Relation and context to previous studies on hepcidin promoter regulation

Hepcidin, a circulating peptide hormone, is the master regulator of systemic iron homeostasis, coordinating the use and storage of iron with iron acquisition [53]. This hormone is primarily produced by hepatocytes and is a negative regulator of iron entry into plasma. De Domenico et al. have discussed how regulation of hepcidin expression appears to occur at the level of transcription [52]. Inflammatory cytokines, predominately IL-6, induce transcription of *HAMP* in hepatocytes. This induction involves the activation of Stat3 and binding of Stat3 to a regulatory element in the hepcidin promoter [56, 60]. A second mode of hepcidin regulation depends upon signalling through the bone morphogenetic protein/SMAD (BMP/SMAD) pathway. In particular, BMPs play essential roles in cardiac, neural, and cartilaginous differentiation. Wang et al. demonstrated that deletion of SMAD4 results in embryonic lethality but that liver-specific inactivation of SMAD4 results in loss of hepcidin synthesis and an iron overload phenotype similar to the phenotype seen in hepcidin-knockout mice [81]. Babitt et al. also showed that mice with a deletion in the SMAD4 gene were unable to synthesize hepcidin in response to inflammatory stimuli or to iron load [82]. This result was the first to our knowledge to show that the BMP/SMAD4 pathway is critical to hepcidin expression. Babitt and colleagues previously demonstrated that HJV acts as a BMP co-receptor in vitro, which facilitates the activation of the BMP-type I/type II receptor complex [43]. Mutations in *HFE2*, the gene that encodes HJV, lead to early-onset iron overload disease [83]. This form of juvenile hemochromatosis is typified by the absence of hepcidin and leads to heart disease, liver iron overload, and diabetes and is indistinguishable from the effects of *HAMP* mutations in patients. It has also been shown by Wang et al. that neither iron nor IL-6 can induce hepcidin expression in mice with a liver-specific *SMAD4* gene deletion [81].

Alongside Babitt et al. and Wang et al. provide compelling data that the BMP signalling pathway is critical for IL-6 induction [81, 82]. The architecture of the hepcidin promoter was characterized in detail in previous studies, and BRE1, BRE2 as well as STATBS were identified as central cis-regulatory elements mediating the responsiveness to BMP and IL6 [50]. Our results confirm the central role of these elements, as STATBS deleted promoters and BRE1mBRE2m double mutants showed strongly reduced IL6 and BMP inducibility, respectively. However, we also observed that stimulation with high doses of BMP enhances expression from the BRE1mBRE2m promoter three-fold (*Fig. 23, Page 123*). This suggests that weak BMP-responsive elements other than BRE1 and BRE2 exist. We hypothesize that BMP receptors phosphorylate SMAD2/3 transcription factors to a minor extent, which in turn allows for activation of the previously described TGF β -responsive elements in the hepcidin promoter [43].

IV.2 Signal integration and implication of the hepcidin promoter

description

The hepcidin promoter showed remarkable plasticity in the co-stimulation response, since IL6 and BMP appeared to control expression more-than multiplicatively at low doses, while sub-multiplicative behavior was observed at high doses (*Fig. 21, Page 119*). Synergistic control at low doses arises from a cooperative interaction between transcription factors bound to BRE1 and STATBS. Previous studies reported that BRE1 is required for full IL6 responsiveness of the hepcidin promoter [44, 50], and thus already indicated cooperation between pathways. However, it remained unclear whether synergistic regulation occurs at the level of signal transduction as well. Our co-stimulation data excludes synergistic regulation at the level of signalling pathways, and rather suggests moderate pathway cross-inhibition (compare *Figs. 11, Page 90 and 15, Page 109*). Our mathematical model could not identify the molecular nature of the cooperative interaction between BRE1 and STATBS promoter elements. Cooperativity may arise from a protein-protein complex between STAT and SMAD transcription factors that stabilizes promoter binding. Alternatively, the two transcription factors may establish complementary activating marks on chromatin that would imply an epigenetic mechanism of cooperation.

High-level stimulation with BMP and IL6 saturates hepcidin expression in HuH7 cells, most likely due to saturating binding of RNAP to the TSS. Saturation already starts at much lower stimulus doses, explaining why IL6 and BMP generally control expression in a less-than multiplicative, antagonistic manner (*Fig. 7, Page 62*). Similar sub-multiplicative behavior was reported in previous studies analyzing a hepcidin reporter construct or the expression of hepcidin peptides under co-stimulation conditions [84, 85]. Promoter saturation promotes redundancy of BRE1 and BRE2, and may help explaining why single mutations in BRE1 and

BRE2 generally have moderate effects on systemic iron homeostasis (*Fig. 20, Page 117*). Moreover, promoter saturation effects explain our experimental observation that BRE1 and BRE2 single mutations have particularly strong impact on expression at intermediate BMP concentrations, while the effects are less pronounced at lower and higher BMP concentrations (*Fig. 25, Page 119*): At low BMP concentrations, the BREs are incompletely occupied by phosphorylated SMAD transcription factors. Stronger stimulation promotes phospho-SMAD binding, explaining why the impact of binding site mutations initially increases for increasing BMP doses. Strong BMP stimulation saturates the hepcidin promoter and promotes redundancy, implying that a further increase in the BMP signal reduces the impact of BRE1 and BRE2 single binding site deletions.

The presence of two BMP-responsive elements as opposed to a single STAT binding site on the hepcidin promoter raises the question of why such a promoter design may be advantageous for the regulation of systemic iron homeostasis. Our results suggest that the presence of two BREs has little impact on the qualitative features of IL6 and BMP signal integration at the promoter level (*Fig. 21, Page 119*). Rather we propose that the presence of two BREs renders the expression of hepcidin more sensitive towards changes in the iron-sensing BMP pathway (*Fig. 22 - D*). This makes the negative auto-regulation loop more nonlinear, thereby promoting systemic iron homeostasis (*Fig. 26, Page 127*).

IV.3 Limitations of current modelling approach

The fine-tuned expression of hepcidin plays a central role in systemic iron homeostasis, and is deregulated in two major clinical settings, hemochromatosis and anaemia of inflammation. In this project, we comprehensively and mathematically characterized the gene regulatory function of the hepcidin promoter using data from systematic promoter mutagenesis and co-stimulation experiments with BMP and IL-6. We presented an integrative mathematical model simultaneously describing crosstalk events at the signalling and promoter levels. Our modelling approach thus captures multiple scales, ranging from fast post-translational modifications to slow gene expression responses. We showed that crosstalk mainly arises at the level of hepcidin promoter regulation, while crosstalk between IL6 and BMP at the level of signalling plays only a minor role.

Gene expression may be a gradual or a binary event at the single cell level [86]. The present study in this thesis was based on cell population measurements of transcription factor phosphorylation and reporter gene expression. The model assumes that population-based measurements reflect the behavior at the single-cell level, and thus presumes a gradual mode of hepcidin expression. Single-cell measurements indicate that BMP-induced target gene expression is indeed a gradual event at the single-cell level [87], thus supporting this basic assumption of our model. In any case, our population-based mathematical model reflects upon the physiologically relevant aspects of hepcidin expression, since systemic hepcidin levels *in vivo* are governed by expression in an ensemble of hepatocytes.

Thermodynamic modelling assumes that transcription factor binding to the promoter has reached an equilibrium state. This presumption holds for our experimental setup, since we

focused on signalling and gene expression responses at steady state. A plausible promoter model could be derived despite uncertainties about cooperativity between transcription factors by systematically comparing the ability of different model topologies to fit the experimental data (*Fig. 11, Page 90*). In future studies, model selection approaches may be combined with transcription factor binding site information and mRNA half-life data to model genome-wide gene expression responses to co-stimulation. Our simple model of negative autoregulation by hepcidin (*Fig. 25, Page 126*) constitutes a first step towards a quantitative description of systemic iron homeostasis. Future studies are required to investigate whether this simple model can predict the dynamics of iron metabolism in vivo.

IV.4 Extension of the proposed modelling approach

We employed the framework of thermodynamic modelling to describe the combinatorial transcription factor binding to the promoter. Thermodynamic models also are fractional occupancy models and are based on formulations originating from statistical physics. In modelling transcriptional analysis we see that these models take into consideration all possible states of the promoter where each ‘state’ implies a specific configuration of regulatory proteins/transcription factors bound to the DNA. As described in the previous sections on modelling, each of these states is awarded a weight that depends on properties like concentration or binding affinity of the site-specific proteins. The mathematical formulation we used for calculating the expression level (F_{reg}) is a rational function in which the estimable parameters are present in the numerator as well as the denominator.

Thermodynamic modelling was previously applied mainly to bacterial gene regulation [70-72] or allowed for the coarse-grained analysis of eukaryotic gene expression [25, 88]. The present work shows that the method can successfully predict the detailed behavior of a complex mammalian promoter as well: calibration of the homeostasis model has been observed in mouse data [89]. Time-resolved hepcidin expression data might in future provide us with the necessary foundation for developing ODE models for describing the dose-dependent dynamics for the hepcidin promoter. However such a modelling approach assumes that the dynamics at the promoter level are exclusively driven by internal deterministic mechanisms. Therefore there is an important scope of our present study that we extend our current approach to include more variations in the dynamics. One way of modelling the system accounting for more variations would be by using stochastic influences. If data on single-observation variability is available then one could extend the existing results to making a model based on stochastic differential equations where relevant parameters are modeled as

suitable stochastic processes or stochastic processes are added to the set of equations defining the system under consideration. Here however, It would be necessary to understand the weightage of noise in the dynamics of the hepcidin promoter since all biological dynamical systems evolve under stochastic processes where stochasticity maybe defined as the part the model cannot predict or is not explicitly included in the modelling. In a way of being computationally and experimentally intensive, it is still realistic that to draw a robust conclusion from a biological system we ought to take into account the random influences since they are an integral part and cannot be entirely isolated from the external and inherent effects that the system in consideration is exposed to.

IV.5 Implications for disease and therapy

Iron is the main stimulus for hepcidin production and needs to be tightly regulated to avoid a variety of diseased conditions. Iron is sensed through the hepatocytes' plasma membrane and involves the bone-morphogenic proteins (BMPs) and their receptors along with some ancillary proteins as the informers for iron-concentrations in the blood. Low hepcidin levels lead to high accumulation of iron in organs whereas consistently high hepcidin level leads to anaemia. For the whole range of iron-related disorders, there is now evidence for involvement of hepcidin in a number of them, either hereditary (hemochromatosis, ferroportin disease etc.) or acquired (anaemia of inflammation, inefficient erythropoiesis in thalassemia intermedia) [90]. Experimental evidence suggests that alcohol consumption [91] and chronic viral hepatitis [92] effect hepcidin transcription and cause hepatic iron overload. Since hemochromatosis is caused by defective synthesis of hormone hepcidin it might benefit from hormonal (hepcidin) therapy. However patients with latent anaemia may not be able to tolerate this treatment since an excess of hepcidin inflow might induce the anaemia to get worse. All genetic or acquired factors that cause increased hepcidin synthesis activity necessarily lead to decreased iron transfer into plasma and hypoferremia. If hepcidin stimulation persists, iron-restricted erythropoiesis and anaemia might follow as an effect. It is our hope that our efforts to mathematically analyse the behaviour of hepcidin promoter and interdependence of relevant pathways would spark off a steady lease of further theoretical and quantitative studies that, along with relevant experimental validation, would help develop well-designed clinical studies leading to long-term efficacy with minimal damaging side-effects and reduce the risks and improve benefits of hepcidin-targeted treatments.

APPENDIX

I. Numbered list of equations for easy reference

$$\frac{d[Fe_b]}{dt} = v_{influx} - v_{efflux} = \frac{k_{influx}}{1+k_{FB} \cdot [hepcidin]} \cdot [Fe_i] - k_{efflux} \cdot [Fe_b] \quad (\text{Equation 1.1.1})$$

$$\frac{d[hepcidin]}{dt} = v_{synthesis} - v_{degradation} = k_{induced} * [Fe_b]^n - k_{deg} \cdot [hepcidin] \quad (\text{Equation 1.1.2})$$

$$[Fe_b] = \left(\frac{k_{influx} \cdot [Fe_i] \cdot k_{deg}}{k_{efflux} \cdot k_{induced} \cdot k_{FB}} \right)^{1/(1+n)} \quad (\text{Equation 1.2})$$

$$\frac{d[Fe_b]}{dt} = v_{influx} - v_{efflux} = \frac{k_{influx}}{1+k_{FB} \cdot [hepcidin]} \cdot [Fe_i] - k_{efflux} \cdot [Fe_b] \quad (\text{Equation 1.3.1})$$

$$\frac{d[hepcidin]}{dt} = v_{synthesis} - v_{degradation} = k_{basal} + k_{induced} * [Fe_b] - k_{deg} \cdot [hepcidin] \quad (\text{Equation 1.3.2})$$

$$[Fe_b] = \frac{1}{2} \cdot \left(-\frac{k_{basal}}{k_{induced}} + \sqrt{\left(\frac{k_{basal}}{k_{induced}}\right)^2 + 4 \cdot \frac{k_{influx} \cdot [Fe_i] \cdot k_{deg}}{k_{efflux} \cdot k_{induced} \cdot k_{FB}}} \right)$$

(Equation 1.4)

$$G = \frac{[Fe_i]}{[Fe_b]} \cdot \frac{d[Fe_b]}{d[Fe_i]}$$

(Equation 1.5)

$$\lim_{k_{basal}/k_{induced} \rightarrow \infty} (G) = 1$$

(Equation 1.6)

$$y = \frac{y_{basal} + y_{max} \cdot S^n}{S^n + EC_{50}^n}$$

(Equation 2.0)

$$[pST] = y_{max,1} * f_{Hill,IL6}$$

(Equation 2.1.1)

$$[pSM] = y_{basal,2} + y_{max,2} * f_{Hill,BMP}$$

(Equation 2.1.2)

$$f_{Hill,IL6} = \frac{[IL6]^{n_1}}{[IL6]^{n_1} + EC_{50,1}^{n_1}}$$

(Equation 2.1.3)

$$f_{Hil,BMP} = \frac{[BMP]^{n_2}}{[BMP]^{n_2} + EC_{50,2}^{n_2}}$$

(Equation 2.1.4)

$$y(t) = f(x(t), \mathbf{p})$$

(Equation 2.1.5)

$$F = E - T * S$$

(Equation 2.1.6)

$$Z_i = e^{\frac{F_i}{K_b * T}}$$

(Equation 2.1.7)

$$\frac{d(mRNA)}{dt} = k_1 - d_1 \cdot mRNA$$

(Equation 2.2.1)

$$\frac{d(Protein)}{dt} = k_2 \cdot mRNA - d_2 \cdot Protein$$

(Equation 2.2.2)

$$P(t) = P_{ss} \left(1 - \frac{P_{ss} - P(0)}{P_{ss}} \cdot \frac{d_1 \cdot e^{-d_2 \cdot t} - d_2 \cdot e^{-d_1 \cdot t}}{d_1 - d_2} \right)$$

(Equation 2.2.3)

$$P_{ss} = \left(\frac{k_1}{d_1} \right) \cdot \left(\frac{k_2}{d_2} \right)$$

(Equation 2.2.4)

$$t_{\frac{1}{2}} = \ln(2)/d$$

(Equation 2.2.5)

$$p_{bound} = \frac{\sum Z_{bound}}{Z_{tot}}$$

(Equation 2.3.1)

$$\begin{aligned} \sum Z_{bound} = & Z(P - 1, A_1, A_2, A_3) + Z(P - 1, A_1 - 1, A_2, A_3) + Z(P - 1, A_1, A_2 - 1, A_3) + \\ & Z(P - 1, A_1, A_2, A_3 - 1) + Z(P - 1, A_1 - 1, A_2 - 1, A_3) + Z(P - 1, A_1 - 1, A_2, A_3 - 1) + \\ & Z(P - 1, A_1, A_2 - 1, A_3 - 1) + Z(P - 1, A_1 - 1, A_2 - 1, A_3 - 1) \end{aligned}$$

(Equation 2.3.2)

$$\begin{aligned}
Z_{tot} &= \sum Z_{bound} + \sum Z_{free} \\
&= \sum Z_{bound} + Z(P, A_1, A_2, A_3) + Z(P, A_1 - 1, A_2, A_3) + Z(P, A_1, A_2 - 1, A_3) \\
&\quad + Z(P, A_1, A_2, A_3 - 1) + Z(P, A_1 - 1, A_2 - 1, A_3) + Z(P, A_1 - 1, A_2, A_3 - 1) \\
&\quad + Z(P, A_1, A_2 - 1, A_3 - 1) + Z(P, A_1 - 1, A_2 - 1, A_3 - 1)
\end{aligned}
\tag{Equation 2.3.3}$$

$$Z(P, A_1, A_2, A_3) = \frac{N_{NS}! * e^{-\frac{\epsilon_P^{NS} + \epsilon_{A_1}^{NS} + \epsilon_{A_2}^{NS} + \epsilon_{A_3}^{NS}}{k_B * T}}}{P! * A_1! * A_2! * A_3! * (N_{NS} - P - A_1 - A_2 - A_3)!}
\tag{Equation 2.3.4}$$

$$Z(P, A_1 - 1, A_2, A_3) = \frac{N_{NS}! * e^{-\frac{\epsilon_P^{NS} + \epsilon_{A_1}^S + \epsilon_{A_2}^{NS} + \epsilon_{A_3}^{NS}}{k_B * T}}}{P! * (A_1 - 1)! * A_2! * A_3! * (N_{NS} - P - (A_1 - 1) - A_2 - A_3)!}
\tag{Equation 2.3.5}$$

$$Z(P, A_1, A_2 - 1, A_3) = \frac{N_{NS}! * e^{-\frac{\epsilon_P^{NS} + \epsilon_{A_1}^{NS} + \epsilon_{A_2}^S + \epsilon_{A_3}^{NS}}{k_B * T}}}{P! * A_1! * (A_2 - 1)! * A_3! * (N_{NS} - P - A_1 - (A_2 - 1) - A_3)!}
\tag{Equation 2.3.6}$$

$$Z(P, A_1, A_2, A_3 - 1) = \frac{N_{NS}! e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^S}{k_B T}}}{P! A_1! A_2! (A_3 - 1)! (N_{NS} - P - A_1 - A_2 - (A_3 - 1))!}$$

(Equation 2.3.7)

$$Z(P, A_1 - 1, A_2 - 1, A_3) = \frac{N_{NS}! e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^S + \varepsilon_{A_2}^S + \varepsilon_{A_3}^{NS} + \varepsilon_{A_1 A_2}}{k_B T}}}{P! (A_1 - 1)! (A_2 - 1)! A_3! (N_{NS} - P - (A_1 - 1) - (A_2 - 1) - A_3)!}$$

(Equation 2.3.8)

$$Z(P, A_1 - 1, A_2, A_3 - 1) = \frac{N_{NS}! e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^S + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^S + \varepsilon_{A_1 A_3}}{k_B T}}}{P! (A_1 - 1)! A_2! (A_3 - 1)! (N_{NS} - P - (A_1 - 1) - A_2 - (A_3 - 1))!}$$

(Equation 2.3.9)

$$Z(P, A_1, A_2 - 1, A_3 - 1) = \frac{N_{NS}! e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^S + \varepsilon_{A_3}^S + \varepsilon_{A_2 A_3}}{k_B T}}}{P! A_1! (A_2 - 1)! (A_3 - 1)! (N_{NS} - P - A_1 - (A_2 - 1) - (A_3 - 1))!}$$

(Equation 2.3.10)

$$Z(P, A_1 - 1, A_2 - 1, A_3 - 1) = \frac{N_{NS}! e^{-\frac{\varepsilon_P^{NS} + \varepsilon_{A_1}^S + \varepsilon_{A_2}^S + \varepsilon_{A_3}^S + \varepsilon_{A_1 A_2} + \varepsilon_{A_1 A_3} + \varepsilon_{A_2 A_3}}{k_B T}}}{P! (A_1 - 1)! (A_2 - 1)! (A_3 - 1)! (N_{NS} - P - (A_1 - 1) - (A_2 - 1) - (A_3 - 1))!}$$

(Equation 2.3.11)

$$Z(P - 1, A_1, A_2, A_3) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^S + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^{NS}}{k_B * T}}}{(P-1)! * A_1! * A_2! * A_3! * (N_{NS} - (P-1) - A_1 - A_2 - A_3)!}$$

(Equation 2.3.12)

$$Z(P - 1, A_1 - 1, A_2, A_3) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^S + \varepsilon_{A_1}^S + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^{NS} + \varepsilon_{PA_1}}{k_B * T}}}{(P-1)! * (A_1-1)! * A_2! * A_3! * (N_{NS} - (P-1) - (A_1-1) - A_2 - A_3)!}$$

(Equation 2.3.13)

$$Z(P - 1, A_1, A_2 - 1, A_3) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^S + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^S + \varepsilon_{A_3}^{NS} + \varepsilon_{PA_2}}{k_B * T}}}{(P-1)! * A_1! * (A_2-1)! * A_3! * (N_{NS} - (P-1) - A_1 - (A_2-1) - A_3)!}$$

(Equation 2.3.14)

$$Z(P - 1, A_1, A_2, A_3 - 1) = \frac{N_{NS}! * e^{-\frac{\varepsilon_P^S + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^S + \varepsilon_{PA_3}}{k_B * T}}}{(P-1)! * A_1! * A_2! * (A_3-1)! * (N_{NS} - (P-1) - A_1 - A_2 - (A_3-1))!}$$

(Equation 2.3.15)

$$Z(P - 1, A_1 - 1, A_2 - 1, A_3) = \frac{N_{NS}! * e^{\frac{\varepsilon_P^S + \varepsilon_{A_1}^S + \varepsilon_{A_2}^S + \varepsilon_{A_3}^{NS} + \varepsilon_{PA_1} + \varepsilon_{PA_2} + \varepsilon_{A_1 A_2}}{k_B * T}}}{(P-1)! * (A_1-1)! * (A_2-1)! * A_3! * (N_{NS} - (P-1) - (A_1-1) - (A_2-1) - A_3)!}$$

(Equation 2.3.16)

$$Z(P - 1, A_1 - 1, A_2, A_3 - 1) = \frac{N_{NS}! * e^{\frac{\varepsilon_P^S + \varepsilon_{A_1}^S + \varepsilon_{A_2}^{NS} + \varepsilon_{A_3}^S + \varepsilon_{PA_1} + \varepsilon_{PA_3} + \varepsilon_{A_1 A_3}}{k_B * T}}}{(P-1)! * (A_1-1)! * A_2! * (A_3-1)! * (N_{NS} - (P-1) - (A_1-1) - A_2 - (A_3-1))!}$$

(Equation 2.3.17)

$$Z(P - 1, A_1, A_2 - 1, A_3 - 1) = \frac{N_{NS}! * e^{\frac{\varepsilon_P^S + \varepsilon_{A_1}^{NS} + \varepsilon_{A_2}^S + \varepsilon_{A_3}^S + \varepsilon_{PA_2} + \varepsilon_{PA_3} + \varepsilon_{A_2 A_3}}{k_B * T}}}{(P-1)! * A_1! * (A_2-1)! * (A_3-1)! * (N_{NS} - (P-1) - A_1 - (A_2-1) - (A_3-1))!}$$

(Equation 2.3.18)

$$Z(P - 1, A_1 - 1, A_2 - 1, A_3 - 1) =$$

$$\frac{N_{NS}! * e^{\frac{\varepsilon_P^S + \varepsilon_{A_1}^S + \varepsilon_{A_2}^S + \varepsilon_{A_3}^S + \varepsilon_{PA_1} + \varepsilon_{PA_2} + \varepsilon_{PA_3} + \varepsilon_{A_1 A_2} + \varepsilon_{A_1 A_3} + \varepsilon_{A_2 A_3}}{k_B * T}}}{(P-1)! * (A_1-1)! * (A_2-1)! * (A_3-1)! * (N_{NS} - (P-1) - (A_1-1) - (A_2-1) - (A_3-1))!}$$

(Equation 2.3.19)

$$p_{bound} = \frac{\sum Z_{bound}}{Z_{tot}} = \frac{1}{1 + \frac{\sum Z_{free}}{\sum Z_{bound}}} = \frac{1}{1 + \frac{1}{F_{Reg}} * \frac{Z(P, A_1, A_2, A_3)}{Z(P-1, A_1, A_2, A_3)}} \approx \frac{1}{1 + \frac{N_{NS}}{P * F_{Reg}} * e^{\frac{\Delta \epsilon_P}{k_B * T}}}$$

(Equation 2.4)

$$\frac{Z(P, A_1, A_2, A_3)}{Z(P-1, A_1, A_2, A_3)} = \frac{N_{NS} - (P-1) - A_1 - A_2 - A_3}{P} * e^{\frac{\Delta \epsilon_P}{k_B * T}} \approx \frac{N_{NS}}{P} * e^{\frac{\Delta \epsilon_P}{k_B * T}}$$

(Equation 2.5)

$$\Delta \epsilon_P = \epsilon_P^S - \epsilon_P^{NS}$$

(Equation 2.6)

$$F_{Reg} = \frac{\sum Z_{bound} / Z(P-1, A_1, A_2, A_3)}{\sum Z_{free} / Z(P, A_1, A_2, A_3)}$$

(Equation 2.7)

$$F_{Reg} = \frac{1 + c_1 + c_2 + c_3 + c_4 + c_5 + c_6 + c_7}{1 + c_8 + c_9 + c_{10} + c_{11} + c_{12} + c_{13} + c_{14}}$$

(Equation 2.8)

$$c_1 = \frac{Z(P-1, A_1-1, A_2, A_3)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta \epsilon_{A_1}}{k_B * T}} * e^{-\frac{\epsilon_{PA_1}}{k_B * T}}$$

(Equation 2.9.1)

$$C_2 = \frac{Z(P-1, A_1, A_2-1, A_3)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_2}}{k_B * T}}$$

(Equation 2.9.2)

$$C_3 = \frac{Z(P-1, A_1, A_2, A_3-1)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_3}}{k_B * T}}$$

(Equation 2.9.3)

$$C_4 = \frac{Z(P-1, A_1-1, A_2-1, A_3)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}} * \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_1} + \varepsilon_{PA_2} + \varepsilon_{A_1 A_2}}{k_B * T}}$$

(Equation 2.9.4)

$$C_5 = \frac{Z(P-1, A_1-1, A_2, A_3-1)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}} * \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_1} + \varepsilon_{PA_3} + \varepsilon_{A_1 A_3}}{k_B * T}}$$

(Equation 2.9.5)

$$C_6 = \frac{Z(P-1, A_1, A_2-1, A_3-1)}{Z(P-1, A_1, A_2, A_3)} = \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}} * \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_2} + \varepsilon_{PA_3} + \varepsilon_{A_2 A_3}}{k_B * T}}$$

(Equation 2.9.6)

$$C_7 = \frac{Z(P-1, A_1-1, A_2-1, A_3-1)}{Z(P-1, A_1, A_2, A_3)} =$$

$$\frac{A_1}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}} * \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}} * \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\varepsilon_{PA_1} + \varepsilon_{PA_2} + \varepsilon_{PA_3} + \varepsilon_{A_1 A_2} + \varepsilon_{A_1 A_3} + \varepsilon_{A_2 A_3}}{k_B * T}}$$

(Equation 2.9.7)

$$c_8 = \frac{Z(P, A_1-1, A_2, A_3)}{Z(P, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}}$$

(Equation 2.9.8)

$$c_9 = \frac{Z(P, A_1, A_2-1, A_3)}{Z(P, A_1, A_2, A_3)} = \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}}$$

(Equation 2.9.9)

$$c_{10} = \frac{Z(P, A_1, A_2, A_3-1)}{Z(P, A_1, A_2, A_3)} = \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}}$$

(Equation 2.9.10)

$$c_{11} = \frac{Z(P, A_1-1, A_2-1, A_3)}{Z(P, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}} * \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}} * e^{-\frac{\varepsilon_{A_1 A_2}}{k_B * T}}$$

(Equation 2.9.11)

$$c_{12} = \frac{Z(P, A_1-1, A_2, A_3-1)}{Z(P, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}} * \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\Delta \varepsilon_{A_1 A_3}}{k_B * T}}$$

(Equation 2.9.12)

$$c_{13} = \frac{Z(P, A_1, A_2-1, A_3-1)}{Z(P, A_1, A_2, A_3)} = \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}} * \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\varepsilon_{A_2 A_3}}{k_B * T}}$$

(Equation 2.9.13)

$$C_{14} = \frac{Z(P, A_1-1, A_2-1, A_3-1)}{Z(P, A_1, A_2, A_3)} = \frac{A_1}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_1}}{k_B * T}} * \frac{A_2}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_2}}{k_B * T}} * \frac{A_3}{N_{NS}} * e^{-\frac{\Delta \varepsilon_{A_3}}{k_B * T}} * e^{-\frac{\varepsilon_{A_1 A_2} + \varepsilon_{A_1 A_3} + \varepsilon_{A_2 A_3}}{k_B * T}}$$

(Equation 2.9.14)

$$F_{Reg} =$$

$$\frac{1 + \frac{[A_1]}{K_{A_1}} f_1 + \frac{[A_2]}{K_{A_2}} f_2 + \frac{[A_3]}{K_{A_3}} f_3 + \frac{[A_1][A_2]}{K_{A_1} K_{A_2}} f_1 f_2 \omega_{12} + \frac{[A_1][A_3]}{K_{A_1} K_{A_3}} f_1 f_3 \omega_{13} + \frac{[A_2][A_3]}{K_{A_2} K_{A_3}} f_2 f_3 \omega_{23} + \frac{[A_1][A_2][A_3]}{K_{A_1} K_{A_2} K_{A_3}} f_1 f_2 f_3 \omega_{12} \omega_{13} \omega_{23}}{1 + \frac{[A_1]}{K_{A_1}} + \frac{[A_2]}{K_{A_2}} + \frac{[A_3]}{K_{A_3}} + \frac{[A_1][A_2]}{K_{A_1} K_{A_2}} \omega_{12} + \frac{[A_1][A_3]}{K_{A_1} K_{A_3}} \omega_{13} + \frac{[A_2][A_3]}{K_{A_2} K_{A_3}} \omega_{23} + \frac{[A_1][A_2][A_3]}{K_{A_1} K_{A_2} K_{A_3}} \omega_{12} \omega_{13} \omega_{23}}$$

(Equation 2.10)

$$p_{bound} = \frac{F_{Reg}}{F_{Reg} + K_P}$$

(Equation 2.11)

$$K_{A_i} = N_{NS} * e^{-\frac{\Delta \varepsilon_{A_i}}{k_B * T}}$$

(Equation 2.12.1)

$$K_P = \frac{N_{NS}}{P} * e^{-\frac{\Delta \varepsilon_P}{k_B * T}}$$

(Equation 2.12.2)

$$\omega_{ij} = e^{-\frac{\varepsilon_{A_i A_j}}{k_B * T}}$$

(Equation 2.13.1)

$$f_i = e^{-\frac{\varepsilon_{P A_i}}{k_B * T}}$$

(Equation 2.13.2)

$$\text{Numerator}_{reg} = 1 + \frac{[pSM]^{nSM}}{K_{B1}} f_{B1} + \frac{[pSM]^{nSM}}{K_{B2}} f_{B2} + \frac{[pST]^{nST}}{K_{ST}} f_{ST} + \frac{[pSM]^{2nSM}}{K_{B1}K_{B2}} f_{B1}f_{B2}\omega_{B1B2} + \frac{[pSM]^{nSM}[pST]^{nST}}{K_{B1}K_{ST}}$$

(Equation 2.14.1)

$$\text{Denominator}_{reg} = 1 + \frac{[pSM]^{nSM}}{K_{B1}} + \frac{[pSM]^{nSM}}{K_{B2}} + \frac{[pST]^{nST}}{K_{ST}} + \frac{[pSM]^{2nSM}}{K_{B1}K_{B2}} \omega_{B1B2} + \frac{[pSM]^{nSM}[pST]^{nST}}{K_{B1}K_{ST}}$$

(Equation 2.14.2)

$$p_{bound} = \frac{F_{Reg}}{F_{Reg} + K_P}$$

(Equation 3.1)

$$F_{reg} = \frac{\text{Numerator}_{reg}}{\text{Denominator}_{reg}}$$

(Equation 3.2)

$$AIC = \chi^2 + 2 \cdot k$$

(Equation 3.3)

$$\chi^2 = \sum_{i=1}^n \frac{(\text{model value}_i - \text{data value}_i)^2}{\text{data value}_i}$$

(Equation 3.4)

$$D = \chi_2^2 - \chi_1^2$$

(Equation 3.5)

$$y = \frac{y_{basal} + y_{max} * S^n}{S^n + EC50^n}$$

(Equation 4)

$$[pST] = \frac{y_{max,1}}{1 + k_{C,1} * [pSM]} * f_{Hill,IL6}$$

(Equation 5.1)

$$[pSM] = y_{basal,2} + \frac{y_{max,2} * f_{Hill,BMP}}{1 + k_{C,2} * [pST]}$$

(Equation 5.2)

$$f_{Hill,IL6} = \frac{[IL6]^{n_1}}{[IL6]^{n_1} + EC_{50,1}^{n_1}}$$

(Equation 5.3.1)

$$f_{Hill,BMP} = \frac{[BMP]^{n_2}}{[BMP]^{n_2} + EC_{50,2}^{n_2}}$$

(Equation 5.3.2)

$$[pST] = \frac{\frac{1}{2} * \frac{1}{k_{C,2}} * \left(\frac{k_{C,2} * y_{max,1} * f_{Hill,IL6} - (1 + k_{C,1} * (y_{basal,2} + y_{max,2} * f_{Hill,BMP}))}{1 + k_{C,1} * y_{basal,2}} + \sqrt{\left(\frac{k_{C,2} * y_{max,1} * f_{Hill,IL6} - (1 + k_{C,1} * (y_{basal,2} + y_{max,2} * f_{Hill,BMP}))}{1 + k_{C,1} * y_{basal,2}} \right)^2 + 4 * k_{C,2} * \frac{y_{max,1} * f_{Hill,IL6}}{1 + k_{C,1} * y_{basal,2}}} \right)}{k_{C,2}}$$

(Equation 5.4)

II. Figures list for easy reference

Figure 1: *Analytical modelling approaches used in gene regulation studies.* (A) Thermodynamic or fractional occupancy model of gene expression. (B) Differential equation model of gene expression. (C) Boolean model of gene expression. [18]

Figure 2: *Hepcidin is expressed by hepatocytes as a response to excess iron levels in blood.* Human physiology does not allow for excretion of iron. Therefore strict balance is maintained by controlling the absorption of iron into blood plasma. Hepcidin is controlled by a negative feedback loop. Excess of iron stimulates hepcidin expression that degrades ferroportin thereby blocking iron-absorption from the duodenum resulting in prevention of iron overload in the liver.

Figure 3: *Hepcidin promoter with three major responsive sites.* The proximal STAT-binding site is responsive to phosphorylated-STAT proteins (pSTATs) released on activation of the IL6/JAK/STAT pathway. There are two BMP responsive regions, one at the proximal and one at the distal end. Simultaneous presence of these two similarly responsive regions increases the BMP-responsiveness of the hepcidin promoter. How the binding sites influence each other is the question we attempt to answer with the help of modelling.[51]

Figure 4: *Signal integration at the hepcidin promoter.* Schematic representation of two critical signalling pathways controlling hepcidin expression. SMAD and STAT transcription factors are phosphorylated upon BMP and IL6 stimulation, and bind BMP-responsive elements (BRE) and a STAT-binding site (STATBS) in the hepcidin promoter, respectively. The importance of signalling crosstalk is not clear.

Figure 5: *Plasticity of combinatorial regulation.* Transcription may follow logic models such that presence, absence or combined effect of binding sites or signalling pathways directly influence expression.

Figure 6: *Analysis of transcription factor crosstalk at the promoter level by reporter gene assays.* Luciferase expression is driven by the wildtype (WT) hepcidin promoter (3 kb

upstream of TSS) or promoter mutants lacking one of the transcription factor binding sites (panel B; BRE1m = BRE1 mutated; STATdel = deleted for STATBS). Luciferase activity of each reporter construct (shown on a log₁₀-scale) was measured for increasing doses of IL6 and/or BMP (n = 6).

Figure 7: *Less than multiplicative co-stimulation response in the WT promoter preserved at the level of endogenous hepcidin mRNA expression.* The fold-expression-changes over basal were compared for qPCR and luciferase assays. (A) Scatter plots showing correlation between Luciferase Assay data (n=3) and qPCR data (green and blue are biological replicates, each with technical replicates (n=2)). The axes show log₁₀ of fold-changes over basal expression in unstimulated cells. The respective Pearson Correlation Coefficients(R) for qPCR1 vs. Luciferase Expression and qPCR2 vs. Luciferase Expression support a strong correlation. The blue and green solid lines show linear fits to the data. (B) Co-stimulation modestly increases expression when compared to BMP mono-stimulation. Bargraph showing log₁₀ of fold changes over basal (y- axis) for Luciferase Expression (Blue), and for the two biological replicate qPCR experiments (green and brown; same data as in panel A).

Figure 8: *Co-stimulation response of luciferase reporters reveals multiplicative/sub-multiplicative behavior for multiple stimulus concentrations and promoter mutants.* The x dimension shows the experimentally observed fold-change in expression upon co-stimulation with BMP and IL6. The y dimension shows the product mono-stimulation responses with the same doses of BMP and IL6, respectively. Each data point represents one co-stimulation condition (different concentrations of BMP and IL6 and/or different promoter constructs). The colors of the data points correspond to different promoter constructs (legend). The bisectrix (solid line) marks the expectation for a multiplicative system.

Figure 9: *Modular modelling approach for arriving at gene expression data by integrating systemic behaviour at each level.* This approach provides quantitative understanding of effects involved at the various levels leading to gene expression.

Figure 10: Possible promoter states for applying thermodynamic model. Basis of mathematical modelling of signalling and promoter crosstalk

Figure 11: Model selection approach allows for the identification of protein-protein interactions on the promoter. Various model variants were tested for their ability to fit the data in Fig. 6 (Page 58). The minimal model (model 1) assumes that each transcription factor independently activates RNAP (grey arrows), while more complex variants additionally take into account cooperativity among transcription factors (red arrows).

Figure 12: Best-fit of a hepcidin expression model without crosstalk at the level of BMP and IL6 signalling pathways. Luciferase expression was simulated using Eqs. 2.1 and 3.1, and the transcription rate in the model (pbound) were fitted to the data in (using a scaling factor). The best-fit parameter values of this model are given in Inset 2 (Page 98).

Figure 13: Moderate inhibitory crosstalk at the signalling level. (A) - (D) Analysis of crosstalk at the signalling level by immunoblotting against phosphorylated SMAD and STAT. HuH7 cells were stimulated with increasing doses of IL6 in the presence or absence of BMP (A, C) or vice versa (B, D). Actin levels serve as loading controls. Two biological replicates were performed (Replicate 1: panels A and B; Replicate 2: panels C and D) (E) and (F) Quantification of signalling crosstalk. Data points represent mean and standard deviation of densitometric analyses of Western Blots (N =2). Lines are fits of the Hill equation to the data.

Figure 14: Integrative crosstalk model fits the dose-response curves of transcription factor phosphorylation. Solid lines represent model trajectories in comparison to experimentally measured data points (shown as mean +/- std).

Figure 15: *Integrative crosstalk model fits promoter data from luciferase assay.* The thermodynamic promoter model (topology 4 in *Fig. 10*) was coupled to a simple signalling model describing inhibitory crosstalk between phospho-SMAD and phospho-STAT transcription factors (*Fig. 14*). The simulated luciferase activities in the heatmaps agree well with the corresponding data in *Fig. 6*.

Figure 16: *Co-stimulation response of double mutant promoters.* Confirms that the isolated BRE1 and BRE2 behave similarly. This figure shows heatmaps of luciferase activity under co-stimulation conditions along with the range of model predictions from ‘lower’ to ‘upper’. Data points show the mean values and the two rows of model predictions indicate the range of variation.

Figure 17: *Model prediction range.* The shaded corridor in the figure represent model predictions and show measurement-compliant parameter sets with highest and lowest predicted effects. Data and model were normalized to basal luciferase expression in the BRE2mSTATdel construct.

Figure 18: *BRE1 and STATBS cooperativity effect.* Systematic analysis of transcription factor binding site deletion effects confirms cooperativity of BRE1 and STATBS. The impact of binding site deletions was calculated by taking the luciferase activity ratios of different promoters (indicated on the bottom) and expressed as a log₁₀-fold change (y axis). Data points are mean and standard deviation, and model predictions represent the range of measurement-compliant parameter sets.

Figure 19: *Redundancy of BRE2 and STATBS.* Systematic analysis of transcription factor binding site deletion effects supports promoter saturation and redundancy of BRE2 and STATBS. The impact of binding site deletions was calculated by taking the expression ratios of different promoter constructs (as indicated along the y-axis) and expressed as log₁₀ fold-change. Data points are mean and standard deviation, and model predictions represent the range of measurement-compliant parameter sets. The combined deletion of BRE2 and STATBS (red bars) are functionally redundant and act as buffers.

Figure 20: *Redundancy of BRE1 and BRE2.* Systematic analysis of transcription factor binding site deletion effects confirms promoter saturation and redundancy of BRE1 and BRE2. Concepts similar to *Figure 17.* A range of BMP stimulation conditions were considered to ensure visible contribution of both BRE1 and BRE2.

Figure 21: *Systems properties of hepcidin expression:* BMP signalling pathway activity is required for optimal IL6 responsiveness of the hepcidin promoter. The IL6 inducibility, defined as the maximal fold expression change by IL6 (over basal), is shown as a function of phospho-SMAD levels (best-fit WT model). The grey corridor indicates range of phospho-SMAD levels in HuH7 cells. The inset shows how the BMP inducibility of WT and BRE2m promoters is in turn affected by the STAT signalling pathway activity.

Figure 22: *Systems properties of hepcidin expression:* Hepcidin expression (fold over basal) is shown as a function phospho-SMAD level for the WT, BRE1m, and BRE2m promoter (phospho-STAT was assumed zero). The dashed lines indicate the maximal steepness of the WT dose-response. Grey corridor same as in *Fig. 21.*

Figure 23: *Systems properties of hepcidin expression:* The luciferase activity (fold over basal) is plotted as function of the IL6 (blue) or BMP (red concentration) concentration. The figure shows simulations of the best-fit model (*Fig. 14, Page 105*)

Figure 24: *Systems properties of hepcidin expression:* The luciferase activity (fold over basal) is plotted as function of the IL6 (blue) or BMP (red concentration) concentration. The figure contains experimental data ($n = 3-6$) and fits of the Hill equation (solid lines). Dashed lines indicate the maximal steepness of the BMP response.

Figure 25: *Conceptual model of iron homeostasis:* Extended mathematical model describing negative feedback regulation of iron blood levels by hepcidin in vivo. Iron blood levels (Fe_b) are controlled by influx and efflux reactions, and the iron influx rate is proportional to the intestinal iron concentration (species Fe_i).

Figure 26: *Systems properties of hepcidin expression:* Iron homeostasis requires two BMP-responsive elements and is abolished by inflammatory stimulation. The extended model (*Fig. 25*) was used to simulate how iron blood levels respond to changes in the intestinal iron concentration.

Figure 27: *Systems properties of hepcidin expression:* The best-fit model (*Fig. 14, Page 105*) was employed to simulate how increasing IL6 stimulation affects the BMP dose-response curve of the promoter. Dashed lines indicate the maximal slope in the absence of IL6. Grey corridor same as in *Fig. 21*.

Reference

1. Buchler, N.E., U. Gerland, and T. Hwa, *On schemes of combinatorial transcription logic*. Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5136-41.
2. Kitano, H., *Computational systems biology*. Nature, 2002. **420**(6912): p. 206-10.
3. Tyson, J.J., K. Chen, and B. Novak, *Network dynamics and cell physiology*. Nat Rev Mol Cell Biol, 2001. **2**(12): p. 908-16.
4. Westerhoff, H.V. and B.O. Palsson, *The evolution of molecular biology into systems biology*. Nat Biotechnol, 2004. **22**(10): p. 1249-52.
5. Kitano, H., *Systems biology: a brief overview*. Science, 2002. **295**(5560): p. 1662-4.
6. Joyce, A.R. and B.O. Palsson, *The model organism as a system: integrating 'omics' data sets*. Nat Rev Mol Cell Biol, 2006. **7**(3): p. 198-210.
7. Bruggeman, F.J., et al., *Systems biology and the reconstruction of the cell: from molecular components to integral function*. Subcell Biochem, 2007. **43**: p. 239-62.
8. Bruggeman, F.J. and H.V. Westerhoff, *The nature of systems biology*. Trends Microbiol, 2007. **15**(1): p. 45-50.
9. Hornberg, J.J., et al., *Cancer: a Systems Biology disease*. Biosystems, 2006. **83**(2-3): p. 81-90.
10. Kell, D.B., *Systems biology, metabolic modelling and metabolomics in drug discovery and development*. Drug Discov Today, 2006. **11**(23-24): p. 1085-92.
11. Carroll, T.J., S. Riek, and R.G. Carson, *Reliability of the input-output properties of the cortico-spinal pathway obtained from transcranial magnetic and electrical stimulation*. J Neurosci Methods, 2001. **112**(2): p. 193-202.
12. Friedman, N., et al., *Using Bayesian networks to analyze expression data*. J Comput Biol, 2000. **7**(3-4): p. 601-20.
13. de Jong, H., *Modelling and simulation of genetic regulatory systems: a literature review*. J Comput Biol, 2002. **9**(1): p. 67-103.

14. Friedman, J., T. Hastie, and R. Tibshirani, *Regularization Paths for Generalized Linear Models via Coordinate Descent*. J Stat Softw, 2010. **33**(1): p. 1-22.
15. El Samad, H., et al., *Stochastic modelling of gene regulatory networks*. International Journal of Robust and Nonlinear Control, 2005. **15**(15): p. 691-711.
16. Coulon, A., O. Gandrillon, and G. Beslon, *On the spontaneous stochastic dynamics of a single gene: complexity of the molecular interplay at the promoter*. BMC Syst Biol, 2010. **4**: p. 2.
17. Gianchandani, E.P., et al., *Matrix formalism to describe functional states of transcriptional regulatory systems*. PLoS Comput Biol, 2006. **2**(8): p. e101.
18. Ay, A. and D.N. Arnosti, *Mathematical modelling of gene expression: a guide for the perplexed biologist*. Crit Rev Biochem Mol Biol, 2011. **46**(2): p. 137-51.
19. Yuh, C.H., H. Bolouri, and E.H. Davidson, *Cis-regulatory logic in the endo16 gene: switching from a specification to a differentiation mode of control*. Development, 2001. **128**(5): p. 617-29.
20. Jaeger, J., et al., *Dynamic control of positional information in the early Drosophila embryo*. Nature, 2004. **430**(6997): p. 368-71.
21. Fakhouri, W.D., et al., *Deciphering a transcriptional regulatory code: modelling short-range repression in the Drosophila embryo*. Mol Syst Biol, 2010. **6**: p. 341.
22. Vellela, M. and H. Qian, *Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlogl model revisited*. J R Soc Interface, 2009. **6**(39): p. 925-40.
23. Schlitt, T. and A. Brazma, *Current approaches to gene regulatory network modelling*. BMC Bioinformatics, 2007. **8 Suppl 6**: p. S9.
24. Bintu, L., et al., *Transcriptional regulation by the numbers: models*. Curr Opin Genet Dev, 2005. **15**(2): p. 116-24.
25. Segal, E., et al., *Predicting expression patterns from regulatory sequence in Drosophila segmentation*. Nature, 2008. **451**(7178): p. 535-40.

26. von Hippel, P.H., et al., *Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects*. Proc Natl Acad Sci U S A, 1974. **71**(12): p. 4808-12.
27. Zhou, X. and Z. Su, *tCal: transcriptional probability calculator using thermodynamic model*. Bioinformatics, 2008. **24**(22): p. 2639-40.
28. Granek, J.A., O. Kayikci, and P.M. Magwene, *Pleiotropic signalling pathways orchestrate yeast development*. Curr Opin Microbiol, 2011. **14**(6): p. 676-81.
29. Kaplan, T., et al., *Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early Drosophila development*. PLoS Genet, 2011. **7**(2): p. e1001290.
30. Chen, C.C. and S. Zhong, *Inferring gene regulatory networks by thermodynamic modelling*. BMC Genomics, 2008. **9 Suppl 2**: p. S19.
31. Reinitz, J., S. Hou, and D.H. Sharp, *Transcriptional Control in <i>Drosophila</i>. Complexus, 2003. **1**(2): p. 54-64.*
32. Goodwin, B.C., *Oscillatory behavior in enzymatic control processes*. Adv Enzyme Regul, 1965. **3**: p. 425-38.
33. Griffith, J.S., *Mathematics of cellular control processes. II. Positive feedback to one gene*. J Theor Biol, 1968. **20**(2): p. 209-16.
34. Griffith, J.S., *Mathematics of cellular control processes. I. Negative feedback to one gene*. J Theor Biol, 1968. **20**(2): p. 202-8.
35. Wong, P., S. Gladney, and J.D. Keasling, *Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose*. Biotechnol Prog, 1997. **13**(2): p. 132-43.
36. von Dassow, G., et al., *The segment polarity network is a robust developmental module*. Nature, 2000. **406**(6792): p. 188-92.
37. Sanchez, L. and D. Thieffry, *A logical analysis of the Drosophila gap-gene system*. J Theor Biol, 2001. **211**(2): p. 115-41.

38. Bakker, B.M., et al., *Glycolysis in bloodstream form Trypanosoma brucei can be understood in terms of the kinetics of the glycolytic enzymes*. J Biol Chem, 1997. **272**(6): p. 3207-15.
39. Kholodenko, B.N., et al., *Quantification of short term signalling by the epidermal growth factor receptor*. J Biol Chem, 1999. **274**(42): p. 30169-81.
40. Teusink, B., et al., *Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry*. Eur J Biochem, 2000. **267**(17): p. 5313-29.
41. Teusink, B., et al., *The danger of metabolic pathways with turbo design*. Trends Biochem Sci, 1998. **23**(5): p. 162-9.
42. Bakker, B.M., et al., *Network-based selectivity of antiparasitic inhibitors*. Mol Biol Rep, 2002. **29**(1-2): p. 1-5.
43. Babitt, J.L., et al., *Bone morphogenetic protein signalling by hemojuvelin regulates hepcidin expression*. Nat Genet, 2006. **38**(5): p. 531-9.
44. Mleczko-Sanecka, K., et al., *SMAD7 controls iron metabolism as a potent inhibitor of hepcidin expression*. Blood, 2010. **115**(13): p. 2657-65.
45. Corradini, E., et al., *Bone morphogenetic protein signalling is impaired in an HFE knockout mouse model of hemochromatosis*. Gastroenterology, 2009. **137**(4): p. 1489-97.
46. Feder, J.N., et al., *A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis*. Nat Genet, 1996. **13**(4): p. 399-408.
47. Camaschella, C., et al., *The gene TFR2 is mutated in a new type of haemochromatosis mapping to 7q22*. Nat Genet, 2000. **25**(1): p. 14-5.
48. Andriopoulos, B., Jr., et al., *BMP6 is a key endogenous regulator of hepcidin expression and iron metabolism*. Nat Genet, 2009. **41**(4): p. 482-7.
49. Meynard, D., et al., *Regulation of TMPRSS6 by BMP6 and iron in human cells and mice*. Blood, 2011. **118**(3): p. 747-56.

50. Casanovas, G., et al., *Bone morphogenetic protein (BMP)-responsive elements located in the proximal and distal hepcidin promoter are critical for its response to HJV/BMP/SMAD*. J Mol Med (Berl), 2009. **87**(5): p. 471-80.
51. Truksa, J., et al., *Bone morphogenetic proteins 2, 4, and 9 stimulate murine hepcidin 1 expression independently of HFE, transferrin receptor 2 (TFR2), and IL-6*. Proc Natl Acad Sci U S A, 2006. **103**(27): p. 10289-93.
52. De Domenico, I., et al., *The molecular mechanism of hepcidin-mediated ferroportin down-regulation*. Mol Biol Cell, 2007. **18**(7): p. 2569-78.
53. Nemeth, E., et al., *IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin*. Journal of Clinical Investigation, 2004. **113**(9): p. 1271-1276.
54. Bolondi, G., et al., *Altered hepatic BMP signalling pathway in human HFE hemochromatosis*. Blood Cells Mol Dis, 2010. **45**(4): p. 308-12.
55. Xia, Y., et al., *Hemojuvelin regulates hepcidin expression via a selective subset of BMP ligands and receptors independently of neogenin*. Blood, 2008. **111**(10): p. 5195-204.
56. Wrighting, D.M. and N.C. Andrews, *Interleukin-6 induces hepcidin expression through STAT3*. Blood, 2006. **108**(9): p. 3204-9.
57. Island, M.L., et al., *A new mutation in the hepcidin promoter impairs its BMP response and contributes to a severe phenotype in HFE related hemochromatosis*. Haematologica, 2009. **94**(5): p. 720-4.
58. Peyssonnaud, C., et al., *Regulation of iron homeostasis by the hypoxia-inducible transcription factors (HIFs)*. J Clin Invest, 2007. **117**(7): p. 1926-32.
59. Fleming, R.E., *Hepcidin activation during inflammation: make it STAT*. Gastroenterology, 2007. **132**(1): p. 447-9.
60. Verga Falzacappa, M.V., et al., *STAT3 mediates hepatic hepcidin expression and its inflammatory stimulation*. Blood, 2007. **109**(1): p. 353-8.

61. Andrews, N.C., *Disorders of iron metabolism*. N Engl J Med, 1999. **341**(26): p. 1986-95.
62. Ohtake, T., et al., *Hepcidin is down-regulated in alcohol loading*. Alcohol Clin Exp Res, 2007. **31**(1 Suppl): p. S2-8.
63. Girelli, D., et al., *Reduced serum hepcidin levels in patients with chronic hepatitis C*. J Hepatol, 2009. **51**(5): p. 845-52.
64. Hentze, M.W., et al., *Two to tango: regulation of Mammalian iron metabolism*. Cell, 2010. **142**(1): p. 24-38.
65. Bluthgen, N. and S. Legewie, *Robustness of signal transduction pathways*. Cell Mol Life Sci, 2012.
66. Delaby, C., et al., *Presence of the iron exporter ferroportin at the plasma membrane of macrophages is enhanced by iron loading and down-regulated by hepcidin*. Blood, 2005. **106**(12): p. 3979-84.
67. Bintu, L., et al., *Transcriptional regulation by the numbers: applications*. Curr Opin Genet Dev, 2005. **15**(2): p. 125-35.
68. Heinrich, P.C., et al., *Principles of interleukin (IL)-6-type cytokine signalling and its regulation*. Biochem J, 2003. **374**(Pt 1): p. 1-20.
69. Alon, U., *Network motifs: theory and experimental approaches*. Nat Rev Genet, 2007. **8**(6): p. 450-61.
70. Kuhlman, T., et al., *Combinatorial transcriptional control of the lactose operon of Escherichia coli*. Proc Natl Acad Sci U S A, 2007. **104**(14): p. 6043-8.
71. Setty, Y., et al., *Detailed map of a cis-regulatory input function*. Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7702-7.
72. Mayo, A.E., et al., *Plasticity of the cis-regulatory input function of a gene*. PLoS Biol, 2006. **4**(4): p. e45.
73. Raveh-Sadka, T., M. Levo, and E. Segal, *Incorporating nucleosomes into thermodynamic models of transcription regulation*. Genome Res, 2009. **19**(8): p. 1480-96.

74. Thompson, J.F., L.S. Hayes, and D.B. Lloyd, *Modulation of firefly luciferase stability and impact on studies of gene regulation*. *Gene*, 1991. **103**(2): p. 171-7.
75. Bachmair, A., D. Finley, and A. Varshavsky, *In vivo half-life of a protein is a function of its amino-terminal residue*. *Science*, 1986. **234**(4773): p. 179-86.
76. Sherman, M.S. and B.A. Cohen, *Thermodynamic state ensemble models of cis-regulation*. *PLoS Comput Biol*, 2012. **8**(3): p. e1002407.
77. Horvath, C.M., *STAT proteins and transcriptional responses to extracellular signals*. *Trends Biochem Sci*, 2000. **25**(10): p. 496-502.
78. Massague, J., J. Seoane, and D. Wotton, *SMAD transcription factors*. *Genes Dev*, 2005. **19**(23): p. 2783-810.
79. Hengl, S., et al., *Data-based identifiability analysis of non-linear dynamical models*. *Bioinformatics*, 2007. **23**(19): p. 2612-8.
80. Ferrell, J.E., Jr. and E.M. Machleder, *The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes*. *Science*, 1998. **280**(5365): p. 895-8.
81. Wang, R.H., et al., *A role of SMAD4 in iron metabolism through the positive regulation of hepcidin expression*. *Cell Metab*, 2005. **2**(6): p. 399-409.
82. Babitt, J.L., et al., *Modulation of bone morphogenetic protein signalling in vivo regulates systemic iron balance*. *J Clin Invest*, 2007. **117**(7): p. 1933-9.
83. Papanikolaou, G., et al., *Hepcidin in iron overload disorders*. *Blood*, 2005. **105**(10): p. 4103-5.
84. Kartikasari, A.E., et al., *Secretion of bioactive hepcidin-25 by liver cells correlates with its gene transcription and points towards synergism between iron and inflammation signalling pathways*. *Biochim Biophys Acta*, 2008. **1784**(12): p. 2029-37.
85. Maes, K., et al., *In anaemia of multiple myeloma, hepcidin is induced by increased bone morphogenetic protein 2*. *Blood*, 2010. **116**(18): p. 3635-44.
86. Biggar, S.R. and G.R. Crabtree, *Cell signalling can direct either binary or graded transcriptional responses*. *EMBO J*, 2001. **20**(12): p. 3167-76.

87. Paulsen, M., et al., *Negative feedback in the bone morphogenetic protein 4 (BMP4) synexpression group governs its dynamic signalling range and canalizes development.* Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10202-7.
88. Zinzen, R.P., et al., *Combinatorial binding predicts spatio-temporal cis-regulatory activity.* Nature, 2009. **462**(7269): p. 65-70.
89. Castoldi, M., et al., *The liver-specific microRNA miR-122 controls systemic iron homeostasis in mice.* J Clin Invest, 2011. **121**(4): p. 1386-96.
90. Pietrangelo, A., *Hepcidin in human iron disorders: therapeutic implications.* J Hepatol, 2011. **54**(1): p. 173-81.
91. Bridle, K., et al., *Hepcidin is down-regulated in alcoholic liver injury: implications for the pathogenesis of alcoholic liver disease.* Alcohol Clin Exp Res, 2006. **30**(1): p. 106-12.
92. Nishina, S., et al., *Hepatitis C virus protein and iron overload induce hepatic steatosis through the unfolded protein response in mice.* Liver Int, 2010. **30**(5): p. 683-92.

Books consulted for general reading

93. Eils, Roland; Kriete, Andres: Computational Systems Biology
94. Alberghina, Lila; Westerhoff, Hans V.: Systems Biology: Definitions and Perspectives
95. Koch, Ina; Reisig, Wolfgang;; Schreiber, Falk: Modeling in Systems Biology
96. Allman, Elizabeth; Rhodes, John A.: Mathematical Modeling in biology
- An introduction
97. McKean, Joseph; Craig, Allen T.: Introduction to Mathematical Statistics
98. Richard, J; Sundar Rao, P.S.S.: Introduction to Biostatistics and Research Methods
99. Draper, Norman R.; Smith, Harry: Applied Regression Analysis (Wiley Series in Probability and Statistics)

100. Strogatz, Steven H.: Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry and Engineering (Studies in Nonlinearity)
101. Kaplan, Daniel; Glass, Leon: Understanding Nonlinear Dynamics: A Primer on Stability, Chaos and Fractals (Texts in Applied Mathematics)
102. Alberts, Bruce; Johnson, Alexander; Walter, Peter; Lewis, Julian; Raff, Martin; Roberts, Keith: Molecular Biology of the Cell

