Excursions in the Theory of Ligand Binding

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
der Universität Mannheim

vorgelegt von
Johannes Wolfgang Robert Martini
aus Kronach

Mannheim, 2014

Dekan:	Professor Dr. Heinz Jürgen Müller, Universität Mannheim
Referent:	Professor Dr. Martin Schlather, Universität Mannheim
Korreferent:	Professor Dr. G. Matthias Ullmann, Universität Bayreuth

Tag der mündlichen Prüfung: 25.03.2014

Abstract

The work on hand deals with different topics within the theory of ligand binding.

The introductory part includes a motivation and basic definitions and presents the mathematical model of equilibrium ligand binding theory, which is based on the Grand Canonical Ensemble of Statistical Mechanics.

The second chapter presents an alternative derivation of the Grand Canonical Partition Function based on a Markov chain model for the ligand binding dynamics of an individual molecule. Moreover, properties of the model are discussed and briefly compared to properties of another processes with the same stationary distribution.

Chapter 3 deals with the decoupled sites representation (DSR, Onufriev et al. (2001)), the underlying mathematical problem and possible generalizations. Moreover, the term "decoupled molecule" is defined and properties of decoupled molecules are discussed.

In Chapter 4, the DSR is transfered –as far as possible– to molecules binding two different types of ligands. Furthermore, the special structure of the system of algebraic equations which has to be solved, if a decoupled molecule shall be calculated is discussed and properties of decoupled molecules are analyzed. Moreover, algorithms to find decoupled molecules are presented.

Chapter 5 transfers results of the algebraic theory of Chapters 3 and 4 to probability theory and thus relates being decoupled to stochastic independence and conditional stochastic independence of certain random variables.

Chapter 6 discusses possible interpretations of complex roots (with imaginary part nonzero) of the binding polynomial and their connection to cooperative ligand binding.

Finally, Chapter 7 presents examples of how ligand binding theory can be used for modeling biological regulatory processes.

Zusammenfassung

Die vorliegende Arbeit behandelt verschiedene Themen innerhalb der Ligandenbindungsteorie.

Der einleitende Teil beinhaltet eine Motivation und grundlegende Definitionen und führt das mathematische Modell der Ligandenbindung im Gleichgewicht, welches auf dem Großkanonischen Ensemble der Statistischem Mechanik beruht, ein.

Das zweite Kapitel stellt eine alternative Herleitung des Großkanonischen Ensembles durch ein Markovkettenmodell der Bindedynamik eines einzelnen Moleküls vor. Außerdem werden Eigenschaften dieses Modells diskutiert und diese mit denen einer anderen ähnlichen Markovkette mit derselben stationären Verteilung verglichen.

Kapitel 3 nimmt sich der decoupled sites representation (DSR, Onufriev et al. (2001)), der Struktur der entsprechenden mathematischen Fragestellung und möglichen Verallgemeinerungen an. Außerdem wird der Begriff des entkoppelten ("decoupled") Moleküls eingeführt und Eigenschaften entkoppelter Moleküle vorgestellt.

In Kapitel 4 wird die DSR –soweit möglich– auf Moleküle mit zwei verschiedenen Typen von Liganden übertragen. Weiterhin wird die spezielle Struktur des Systems algebraischer Gleichungen, welches gelöst werden muss wenn entkoppelte Systeme gesucht werden, diskutiert und Eigenschaften von entkoppelten Molekülen analysiert. Außerdem werden Algorithmen vorgestellt, die es erlauben, die erwähnten algebraischen Systeme zu lösen.

Kapitel 5 überträgt die Resultate der algebraischen Theorie der beiden vorherigen Kapitel in die Wahrscheinlichkeitstheorie und verbindet die Eigenschaft eines Moleküls entkoppelt zu sein mit stochastischer Unabhängigkeit, bzw. bedingter stochastischer Unabhängigkeit bestimmter Zufallsvariablen.

Kapitel 6 diskutiert mögliche Interpretationen komplexer (nicht reeller) Nullstellen des Bindepolynoms und deren Verbindung zu kooperativem Binden des Liganden.

Abschließend stellt Kapitel 7 Beispiele vor, wie Ligandenbindungsteorie auch für die Modellierung biologischer Regulationsprozesse benutzt werden kann.

Preface

The work on hand presents the major part of my scientific research of my time as a research assistant at the Institute of Mathematical Stochastics in Göttigen. The reason for me working on the topic of ligand binding can be found in my last one and a half years of being a student in mathematics at the University of Bayreuth. Having finished the studies in biology, I was looking for ways to combine both disciplines. This search brought me into contact with Prof. G. Matthias Ullmann who leads the Computational Biochemistry Group at the University of Bayreuth. He introduced me to an unsolved problem of ligand binding and thus, I read some theory about this topic from time to time. At the end of my studies of mathematics, I was looking for a position as a research assistant providing the possibility to write a PhD thesis. Luckily, I quickly found an advertisement offering a position as a research assistant for stochastics with applications in natural sciences, especially physics and biology. The job at the University of Göttingen which I applied for, was offered by Prof. Martin Schlather. I was surprised when I read that he graduated as well at the University of Bayreuth, and even more, when I found out that he had also written his Diploma Thesis on Robust Statistics with Prof. Helmut Rieder as advisor. These circumstances show that this work does not only deal with coincidence, but that it also results from it. I got the job and was initially working on stochastic processes in population genetics (allele frequencies). However, talking to Prof. Martin Schlather about my unofficial minor project of ligand binding theory, he encouraged me to put more effort on it and to publish the results. This support and his open mind prepared the ground for this work.

As already mentioned, this work presents the major part of my scientific work of the last three years. Thus, most of the results have already been published (Martini and Ullmann, 2013; Martini, Habeck, and Schlather, 2014; Martini, Schlather, and Ullmann, 2013a,b,c). I would like to emphasize at this point that in this work, bigger parts of all my publications were adopted verbatim. However, some notations, abbreviations and formulas of the text have been amended at several points in order to create a coherent whole text. Moreover, in some chapters, new results were incorporated. For the sake of readability, in particular, to avoid a fragmented text e.g. when the ligand "proton" was substituted by a general "ligand" in Chapter 3, some minor changes are not indicated, but at the beginning of each chapter the publication(s) which the chapter is based on and from which text passages were adopted are mentioned and new results are highlighted. Moreover, concerning the style of this work, it shall be pointed out that even some simple statements are presented in form of a lemma for the sake of a good overview and structure. If the corresponding statement is too obvious or well-known a proof will not be given.

Before we start with the introduction, I would like to express my thankfulness to the people who influenced this work.

I would like to thank Prof. Martin Schlather for being open-minded about every application of mathematics, in particular stochastics, to other sciences as well as for his encouragement and for his constant support.

Moreover, I would like to thank Prof. G. Matthias Ullmann for introducing me to this topic and for his support whenever I asked for it.

I would also like to thank the coauthor of the latest paper (Martini, Habeck, and Schlather, 2014) Michael Habeck for the comfortable working atmosphere and creative discussions.

Furthermore, I am thankful to the following persons with whom I had intense and helpful discussions which also influenced this work: Alexander Malinowski, Kristin Blumenröther, Tobias Dorsch, Mareike Busmann, Timo Aspelmeier and Ulf Fiebig.

Finally, I would like to thank Marco Oesting and Madeline Lips for comments on the text.

Johannes W.R. Martini

Mannheim
January 2014

Contents

1 Introduction

In this introducing chapter, the main vocabulary will be explained and the basics of the corresponding mathematical framework, which can be found in several textbooks (e.g. Ben-Naim, 2001; Cantor and Schimmel, 1980; Wyman and Gill, 1990) will be summarized. This chapter will not present any new ideas or results, except for the way the content is presented, some terms which are not used in literature and the fact that Henderson-Hasselbalch curves will be introduced as a complex-valued function.

The theoretical explanation of the term "titration curve" is in parts adopted verbatim from the publications Martini et al. (2014), Martini et al. (2013c) and Martini and Ullmann (2013).

1.1 Motivation

(Bio-)Chemists come into contact with ligand binding theory in the first semesters of their studies, when pH-dependent proton binding to aminoacids is discussed. The binding properties are usually summarized in a deterministic binding curve (titration curve), which is a rational function in the proton activity ($10^{-\mathrm{pH}}$) and which describes the average number of protons bound to a biomolecule. Titration curves can be investigated from different points of view.

We can regard the curves from a practical point of view and ask how the proton binding might be changed if certain properties of the molecule's structure are altered. This question may be of interest for chemical properties of the substance such as polymerization or breaking properties of a material at a certain pH-value.

We can regard the curves as algebraic objects and ask questions of algebraic geometry and computational algebra: Is it possible –from a mathematical point of view– to construct a molecule with a certain proton binding curve? If two molecules share certain proton binding properties, do other properties have to be identical as well?

We can ask for the origin of these rational curves, and why these curves have this structure. These questions will lead us to an interpretation of the curves as the expectation of a family of distributions. Thus, questions of probability theory and statistical physics can be asked: Which characteristics of the family of probability measures correspond to certain characteristics of the binding curve? How can systems be modeled, which consist of a small number of molecules and which can not be described by the expectation sufficiently well?

And we can ask questions about biological or biochemical processes in which ligand binding is involved: How does proton/electron cotransport work? Which effect has a

change of the concentration of a repressor molecule which binds to DNA on the gene regulation network? Which molecule might be the best target for a drug, if a biochemical pathway shall be blocked?

To put it in a nutshell, ligand binding theory is multifaceted, combines a lot of different disciplines, involves fields of pure and applied mathematics and can be relevant for a lot of real applications.

1.2 Ligand, target molecule and binding site

Even though, the term "ligand" is widely used in biological, biochemical, pharmacological, medical and other scientific literature (see e.g. Brennan et al., 2012; Hameed et al., 2013; Kragh-Hansen, 2013; Leppänen et al., 2013), precise definitions of this term in biochemistry or molecular biology textbooks are rare. Alberts et al. (2008) defines it the following way:

"The substance that is bound by the protein –whether it is an ion, a small molecule, or a macromolecule such as another protein– is referred to as a **ligand** *for that protein".* (Alberts et al., 2008, p. 153, Chapter 3)

Obviously, this description leaves some questions open: Does the molecule a ligand binds to, have to be a protein? Is a peptide or a single amino acid unsatisfying to use the term "ligand" for a potential binding partner? What is the characteristic of the bond? Can a ligand bind covalently to a protein?

The characteristics of the bond are described indirectly in the subsequent sentence explaining selectivity:

"The ability of a protein to bind selectively and with high affinity to a ligand depends on the formation of a set of weak, noncovalent bonds –hydrogen bonds, electrostatic attractions, and van der Waals attractions– plus favorable hydrophobic interactions". (Alberts et al., 2008, p. 153, Chapter 3)

Following this description, in this work a **ligand** will be understood as any particle which can bind non-covalently to a **target molecule** (not necessarily a protein). Moreover, the target molecule and the ligand will be assumed to be of different molecule species. The region of a target molecule to which a ligand can bind will be called a **binding site**. If a target molecule has only one binding site, the choice which molecule is regarded as the ligand and which one as the target molecule will be arbitrary. Usually, the smaller molecule (if the molecules are not of equal size) is considered to be the ligand. Target molecules can have several binding sites for the same type of ligand (e.g. hemoglobin and O_2, see Alberts et al. (2008)). The binding sites are usually assumed to be non-overlapping. However, in the used models, overlapping binding sites can be interpreted as non-overlapping binding sites with infinite interaction energy, which represents exclusionary occupation of the respective binding sites. Possible conformational changes of a target molecule, when certain binding sites are occupied will not be considered in this work.

1.3 Titration curve

Let M be a target molecule species with n binding sites for ligand L. Moreover, let a certain amount of both substances be dissolved in a liquidity. Then a certain target molecule M_1 can exist in 2^n **(micro)states**: Each site can be occupied or unoccupied. Thus, the state of M_1 at a fixed time m can be described by an n-tuple

$$M_{1,m} = k = (k_1, ..., k_n) \in \{0, 1\}^n$$

indicating whether site i is occupied ($k_i = 1$) or not ($k_i = 0$). Since we do not know in which state the molecule M_1 is at time m, and since this information will not give any information about the other target molecules, it is appropriate to describe the system by a probability distribution on the set $\{0, 1\}^n$, stating how likely it will be to find a certain microstate k, if we "draw" a molecule M_j from the solution at random. This distribution on $\{0, 1\}^n$ depends on the concentration of the ligand L. However, the "availability" of ligands for the target molecules does not have to be proportional to the ligand's concentration, since intermolecular forces between the ligand molecules can reduce or increase the availability of the ligand for the target molecules. In extreme examples a higher concentration of the ligand can even reduce the availability for a target molecule, due to high intermolecular attracting forces of the ligand molecules. Since we are interested in a general physical description, similar to the laws for ideal gases, we describe the system by equations based on the ideal case and consequently have to use some sort of idealized concentration, **the chemical (or thermodynamical) activity** λ. The **titration curve of a certain binding site** i is defined as the probability of site i being occupied, dependent on the chemical activity λ. The **overall titration curve** is defined as the sum of the titration curves of all binding sites.

1.3.1 Experimental determination

A well-known traditional way to determine overall titration curves in the case of the ligand being a proton, is an acid-base titration experiment: a known number of target molecules is dissolved at a certain pH value. The number of protons in the solution is changed by adding either a (strong) acid or base. Since, at a higher proton activity, the target molecules will bind more protons, not all of the added protons will be free in the solution. The number of protons which are absorbed by the solute is given by the difference of added protons and the increase of free protons (which can be measured as a change of the pH-value).

For other systems of target molecule and ligand, chemical and physical properties in different states have to be used to determine the overall titration curves or the titration curve of individual sites (e.g. absorption of light in the case of hemoglobin and O_2/CO, see Horecker (1943)). Moreover, sophisticated methods such as Nuclear Magnetic Resonance Spectroscopy can be used to investigate the binding to individual sites.

1.3.2 The mathematical description

The classical derivation of the mathematical framework for ligand binding curves is mainly the following idea (Schellman, 1975): A single target molecule in solution can be regarded as a subsystem within the much larger system of the solution. Assuming that the interaction of these subsystems is negligible but that a molecule can exchange ligand particles as well as energy with the environment which has a fixed average temperature T, we can use the **Grand Canonical Ensemble** of Statistical Mechanics to describe the binding of the ligand to a target molecule.

The Grand Canonical Ensemble relates each microstate the molecule can access with a certain weight. The probability of a microstate is then given by its weight divided by the sum of the weights of all possible microstates. The sum of all weights is also called the **(Grand Canonical) Partition Function** (see Subsection 1.3.3 for a sketch of its derivation). Since a target molecule has only a finite number of binding sites, the Partition Function reduces from a power series to a polynomial: the **binding polynomial**.

To write down this concept precisely in equations, the following notation will be used in this work:

- T denotes the **absolute temperature** in °Kelvin,
- R denotes the **Boltzmann constant**,
- μ denotes the **chemical potential** of the ligand and
- $K := \{0,1\}^n$ denotes the set of all microstates.

Moreover, for a microstate $k = (k_1, ..., k_n) \in K$,

- $G(k)$ denotes its **energy level**, relative to some reference state, usually $\{0\}^n$,
- and $|k| := \sum_{i=1}^{n} k_i$ denotes the **number of occupied sites** in microstate k.
- $\{k \in K \,||k| = i\}$ is called the **macrostate** i.

Note that $G(k)$ describes the energy of molecule M in microstate k, which means it depends on the target molecule M and the ligand we are describing. The use of a second index or a second variable is avoided for the sake of a simplified notation.

Using this notation, the Grand Canonical Ensemble states that the **weight** $\mathrm{wght}(k)$ of a microstate k is given by

$$\mathrm{wght}(k) := \exp\left(\frac{-G(k) + \mu|k|}{RT}\right) \tag{1.1}$$

(see e.g. Ben-Naim, 2001; Landau and Lifschitz, 1987; Reif, 1987). The weight $\mathrm{wght}(k)$ is often also called the **Gibbs factor**. The corresponding **probability** of this microstate is

$$P_{M,\mu}(k) := \frac{\mathrm{wght}(k)}{\Phi(M)}, \tag{1.2}$$

where $\Phi(M) = \sum_{k \in K} \text{wght}(k)$ is the above mentioned Grand Canonical Partition Function of molecule M. Since the activity λ is defined by

$$\lambda := \exp\left(\frac{\mu}{RT}\right) \tag{1.3}$$

(Cohen et al., 2008), we see that, for a molecule with a finite number of ligand binding sites, the Grand Canonical Partition Function is a polynomial (**the binding polynomial**) in the variable activity, if we rewrite Eq. (1.1):

$$\text{wght}(k) = \exp\left(\frac{-G(k) + \mu|k|}{RT}\right) = g(k) \cdot \lambda^{|k|}, \tag{1.4}$$

where $|k| \leq n$ and where

$$g(k) := \exp\left(\frac{-G(k)}{RT}\right) \tag{1.5}$$

is called the **Boltzmann factor** or **microstate constant** of microstate k.

The setup described above presents a family of distributions on $\{0,1\}^n$ which is parameterized by the activity λ. The titration curves of certain binding sites and the overall titration curve are derived by applying operators to the parameterized family of distributions or to image measures under certain maps. For instance, the titration curve of a certain site, which has already been defined previously as the probability of site i being occupied, dependent on λ, is given by the rational function

$$\Psi_i(\lambda) := \frac{E_i(M)}{\Phi(M)} \tag{1.6}$$

with

$$E_i(M) = \sum_{\{k \in K | k_i = 1\}} g(k)\lambda^{|k|} \tag{1.7}$$

denoting the joint weight of all microstates in which site i is occupied. Moreover, defining X_i to be the projection from $\{0,1\}^n$ on the i-th coordinate, X_i is a Bernoulli random variable and the probability of X_i being equal to 1 coincides with its expectation. Thus, we can express Eq. (1.6) also as the expectation $\mathbb{E}_\lambda$ of the random variable X_i:

$$\Psi_i(\lambda) = \frac{E_i(M)}{\Phi(M)} = \mathbb{E}_\lambda X_i. \tag{1.8}$$

The expectation carries the index λ, since the probability measure which is used to calculate it, depends on λ. Correspondingly, the overall titration curve is given by

$$\Psi(\lambda) = \sum_{i=1}^{n} \Psi_i(\lambda) = \frac{\sum_{i=1}^{n} E_i(M)}{\Phi(M)} = \sum_{i=1}^{n} \mathbb{E}_\lambda X_i = \mathbb{E}_\lambda |k|, \tag{1.9}$$

As illustrated, the stochastic setup with a family of measures, parameterized by the activity of the ligand λ, leads to algebraic objects –polynomials and rational functions– as titration curves. The simplest structure of a titration curve of an individual site is the following:

Definition 1 (Henderson-Hasselbalch titration curve)**.** *A titration curve of a certain site i is called Henderson-Hasselbalch titration curve, if a $g \in \mathbb{C}^*$ exists such that*

$$\Psi_i = \frac{g\lambda}{g\lambda + 1}. \tag{1.10}$$

Remark 2. *The reader might wonder why the constant g in Definition 1 can be a complex number. Indeed, one should ask whether this extension makes any sense, since an interpretation of a complex valued probability does not exist. For this reason, at this point, g should be regarded as a positive real number. However, a motivation for this formal extension will we given in Chapter 3 and at least a partial interpretation of complex valued Henderson-Hasselbalch curves as part of a system of several binding sites will be given in Chapter 6.*

1.3.3 Sketches of approaches to derive the Grand Canonical Ensemble

In this subsection we shortly present sketches of two approaches for a derivation of the (Grand) Canonical Ensemble. The purpose of these sketches is to illustrate where the structure of Eqs. (1.1,1.2) comes from and why an exponential function is involved. Another approach, based on a model of the dynamics of a system can be found in Chapter 2.

The concept of the first approach is presented in different physics textbooks and based on a Taylor expansion of a certain function. We will present this concept following the textbook by Reif (1987, pp. 236-239): We start with the **Microcanonical Ensemble** by regarding a closed system with fixed number of particles N and fixed energy level E^0 of the whole system. The system can only access the microstates with this fixed number of particles and the given energy level.

A1 We assume an uniform distribution on the accessible microstates.

We extend this ensemble to a model in which the considered system can change its energy state due to contact with a heat-reservoir. This extended model is called the **Canonical Ensemble**. We consider a small system A which exchanges energy with a much larger heat-reservoir A' and use the notation A^0 for the total system which is composed of A and A'. Moreover, let the corresponding energies be denoted by E^0, E and E', with

$$E^0 = E + E',$$

according to the law of energy conservation. If A has a certain fixed state k of energy E, the number of accessible states for the composed system A^0 (with $A = k$) should be equal to the number of possible states for system A' at energy E'. Since the composed system A^0 is a microcanonical system, each microstate of A^0 has the same probability. Consequently, the probability of A being in microstate k is proportional to the number of possible states of A' at energy E':

$$P(A = k) = c\,|\{k'|A' \text{ has energy } E' \text{ in state } k'\}| =: c\,\mathrm{Car}(E'), \tag{1.11}$$

with c denoting a normalization constant, k' a state of system A', and $|\{...\}| = \mathrm{Car}(E')$ the cardinality of the set at energy E'. We are interested in the function $\mathrm{Car}(E')$. We

use the natural logarithm and a Taylor expansion at $E' = E^0$ to receive

$$\ln(\mathrm{Car}(E')) = \ln(\mathrm{Car}(E^0)) + \left(\frac{\partial \ln(\mathrm{Car}(E'))}{\partial E'}\right)_{|E'=E^0} (E' - E^0) + ... \tag{1.12}$$

Moreover, we know from Chapter 3 in Reif (1987) that for

$$\beta(E') := \left(\frac{\partial \ln(\mathrm{Car}(E'))}{\partial E'}\right)$$

the relation

$$\beta(E') = \frac{1}{k\,T(E')}$$

holds with a constant k and the absolute temperature $T(E')$ of the heat reservoir at energy E'. Since $E \ll E^0$, we can assume that

A2 a change of E of the size of its own range is so small compared to E^0 that the temperature of the heat reservoir is not changed.

Assumption A2 implies that β is a constant (independent of E') for the considered scales and that consequently all derivatives of higher order are zero. This gives

$$\ln(\mathrm{Car}(E^0 - E)) = \ln(\mathrm{Car}(E^0)) - \beta E \tag{1.13}$$

and thus

$$\mathrm{Car}(E^0 - E)) = \mathrm{Car}(E^0)\exp(-\beta E). \tag{1.14}$$

This means we can rewrite Eq. (1.11) in the idealized form

$$P(A = k) = C\ \exp(-\beta E), \tag{1.15}$$

with a new constant C since $\mathrm{Car}(E^0)$ is fixed. This procedure can be extended to describe a subsystem exchanging also particles with a bigger system which gives the **Grand Canonical Ensemble**.

A second approach which is also based on Assumption A2 can avoid the Taylor expansion of $\mathrm{Car}(E')$ and instead uses another functional equation. Assuming that

A3 the probability that the system A is in state k with energy E only depends on the temperature of the reservoir

and Assumption A2, Eq. (1.11) gives

$$P(A = k) = c_0\mathrm{Car}(E^0 - E) = c_1\mathrm{Car}(E^1 - E) \tag{1.16}$$

for $E, |(E^1 - E^0)| \ll E^0, E^1$. We define the functions

$$f_0(E) := \frac{\mathrm{Car}(E^0 - E)}{\mathrm{Car}(E^0)} \tag{1.17}$$

and

$$f_1(E) := \frac{\mathrm{Car}(E^1 - E)}{\mathrm{Car}(E^1)}. \tag{1.18}$$

Then

$$f_1(E) = \frac{\mathrm{Car}(E^1 - E)}{\mathrm{Car}(E^1)} = \frac{\mathrm{Car}(E^0 + E^1 - E^0 - E)\mathrm{Car}(E^0)}{\mathrm{Car}(E^0)\mathrm{Car}(E^0 + E^1 - E^0)} = \tag{1.19}$$

$$= \frac{f_0(E^0 - E^1 + E)}{f_0(E^0 - E^1)}.$$

Eq. (1.16) gives

$$\tilde{c}_1 f_1(E) = \tilde{c}_0 f_0(E) \tag{1.20}$$

and thus, with Eq. (1.19)

$$f_0(E^0 - E^1 + E) = \frac{\tilde{c}_0}{\tilde{c}_1} f_0(E) f_0(E^0 - E^1) \tag{1.21}$$

In particular, for $E = 0$ this gives $\frac{\tilde{c}_0}{\tilde{c}_1} = 1$. Thus, f_0 fulfills

$$f_0(F + E) = f_0(F) \cdot f_0(E) \tag{1.22}$$

for $E, F \ll E^0$, which shows that $\ln \circ f_0$ is linear.

A potential target for criticism for both derivations is Assumption A2, since this assumption is obviously an approximation which will never be strictly fulfilled but which provides the desired result of $\mathrm{Car}(E)$ being the exponential function more or less directly.

1.4 Overview

This work presents several topics within the theory of ligand binding. It is largely based on the publications Martini et al. (2014), Martini and Ullmann (2013), Martini et al. (2013a,b,c). Many text passages were adopted verbatim from these publications.

Chapter 2 describes an alternative way to derive the structure of the parameterized family of distributions described by Eqs. (1.2,1.4). The approach is based on a Markov chain model for the ligand binding dynamics of a single molecule. It is shown that the stationary distribution of the Markov chain coincides with the laws of the Grand Canonical Ensemble, if the chemical activity of the ligand is identified with a ratio of the probability of the environment providing a ligand, and the probability of the environment taking a ligand up. The chapter is based on the publication Martini et al. (2014).

Chapter 3 reconsiders the Decoupled Sites Representation (DSR), a theoretical tool originally presented by Onufriev et al. (2001), from a more mathematical point of view. The DSR is a theoretical instrument which allows to regard complex titration curves of biomolecules with several interacting ligand binding sites as composition of isolated, non-interacting sites, each with a standard Henderson-Hasselbalch titration curve. In this chapter, the mathematical framework in which the DSR is embedded is presented and mathematical proofs for several statements in the periphery of the DSR are given. These proofs also identify exceptions. It is highlighted that –to apply the DSR to any arbitrary molecule– it is necessary to extend the set of binding energies from $\mathbb{R}$ to a

stripe within $\mathbb{C}$. An important observation in this context is that even positive interaction energies (repulsion) between the binding sites will not guarantee real-valued binding energies in the decoupled system, at least if the molecule has more than four ligand binding sites. Moreover, it is shown that for a given overall titration curve it is not only possible to find a corresponding system with an interaction energy of zero but with any arbitrary fixed interaction energy. This result also affects practical work as it shows that for any given titration curve, there is an infinite number of corresponding hypothetical molecules. Chapter 3 is based on the publication Martini and Ullmann (2013) but also presents additional results.

In Chapter 4, the Decoupled Sites Representation is transfered to a situation with a target molecule binding two different types of ligands at disjunct sets of binding sites. In particular, the existence of decoupled systems for n_1 and one binding sites for the respective ligand is proven and some difficulties and limits of this transfer are highlighted. A major difference to the DSR for one type of ligand is the loss of uniqueness of the decoupled system. Moreover, properties which all decoupled molecules share are presented. For the case of more than one binding sites for both types of ligands, algorithms are presented which exploit the special structure of the algebraic systems to reduce the problem to the case of only one type of ligand. The corresponding publications are Martini et al. (2013b) and Martini et al. (2013a).

Chapter 5 deals with the stochastic interpretation of the Decoupled Sites Representation for one and for two types of ligands. The algebraic description using polynomials (e.g. the binding polynomial) and rational functions (e.g. titration curves) which is used to characterize systems of molecules and their ligand(s) is translated into stochastics. The shifted point of view facilitates some proofs and physical interpretations. This chapter is based on Martini et al. (2013c).

Chapter 6 discusses the relation between different definitions of cooperative ligand binding and complex valued measures in decoupled systems. It is based in parts on Martini and Ullmann (2013) but also presents other results. The relation between the Hill coefficient and complex roots of the binding polynomial is discussed for molecules with two binding sites. A publication investigating this relation in more detail is planed.

Chapter 7 shows how ligand binding theory, in particular titration curves can be used for modeling biological regulatory and sensing processes. In more detail, simple models for a two component gene regulatory system and for olfactory senses of insects are presented. Chapter 7 shall be the draft for two publications.

2 Modeling the Binding Dynamics of a Molecule With a Finite Number of Binding Sites Using a Markov Chain

The following chapter presents the main part of the paper Martini et al. (2014). Main passages were adopted verbatim from this paper. The comparison of the dynamics of the presented model to the dynamics given by the transition matrix of the common Metropolis-Hastings algorithm in Section 2.4 has not been incorporated in the paper.

2.1 Motivation

Titration experiments as described in Section 1.3 are a classical procedure in chemistry and produce titration curves that characterize the overall binding of L to M. Titration curves for protons binding to aminoacids can be found in nearly every biochemistry textbook and have been studied for a century (Henderson, 1913; Hasselbalch, 1916; Tanford and Kirkwood, 1957; Martini et al., 2013b,a). The mathematical model for titration curves is based on the binding polynomial (bp). As described in the introductory chapter, it is a function of the chemical activity of the ligand and derived as a special case of the Grand Canonical Partition Function (GCPF), if molecule M is regarded as a system that can take up a finite number n of particles (Schellman, 1975; Cantor and Schimmel, 1980; Wyman and Gill, 1990). Its origin in statistical mechanics reemphasizes that it characterizes stochastic properties of a system: It defines a family of distributions over the number of bound ligands, which is parameterized by the chemical activity of the ligand (the temperature is fixed). However, the GCPF describes only the *thermodynamic equilibrium*, a steady state of a system consisting of a large number of molecules, in which every single molecule follows its own dynamics of releasing and binding ligands.

It seems obvious that another approach to derive the well-known laws of equilibrium might be based on modeling the ligand binding dynamics of a single molecule. In this chapter, we derive the GCPF for a system with a finite number of binding sites, starting from modeling the binding dynamics. We use a Markov chain model in discrete time and use some reasonable assumptions about the binding dynamics of the molecule to deduce the transition probabilities. This approach facilitates the understanding of the equilibrium distribution, especially the composition of the probabilities of the microstates and provides an idea of how the chemical activity (which is not necessarily bounded by 1) could be interpreted from a stochastic point of view. Moreover, it also allows us to model the system's way into equilibrium.

2.2 Binding dynamics of a single molecule as a Markov chain model

We will model the binding state of the individual molecule M_1 by a Markov chain on the set of tuples $K := \{0,1\}^n$, with $M_{1,m}$ denoting the state of molecule M_1 after m "time steps". $M_{1,m} = k = (k_1, \ldots, k_n) \in K$ indicates whether a ligand molecule occupies site i ($k_i = 1$) or not ($k_i = 0$). We make the following assumptions concerning ligand binding dynamics to deduce transition probabilities and equilibrium laws:

A1 The time between step m and $m+1$ is so short that the binding state of only one site can change. Using the ℓ_1-Norm

$$|k| := \sum_{i=1}^{n} |k_i|$$

this means $|M_{1,m} - M_{1,m+1}| \leq 1$.

A2 For $k, l \in K$ with $|k - l| = 1$, the probability of a transition $k \mapsto l$ is a product of three factors:

A2-1 the *random choice* of a binding site that may change its binding state,

A2-2 the probability that the *environment* provides a ligand molecule or takes it up (depending on the state of the chosen site) and

A2-3 the probability barrier given by the difference of the energies of microstates k, l of the *target molecule*.

A3 Since the concentration of L is assumed much higher than that of M, we assume that the binding of the ligand to the individual target molecules occurs stochastically independently. This means that the molecules of type M do not interact, and a small reduction or an increase of the number of free ligand molecules, due to an uptake or release by molecules M, does not affect the probability of A2-2.

Assumption A3 guarantees that we can describe the whole system of all target molecules by modeling only one target molecule. In the following, the probability factors of A2 will be specified, which allows us to deduce the matrix of transition probabilities and subsequently its stationary distribution.

Ad **A2-1**: Since the first probability factor describes the choice of a site, there is no need to discriminate between the sites at this point. Consequently, we assume a uniform distribution which means this factor equals $\frac{1}{n}$.

Ad **A2-2**: If the chosen site is not occupied, the second factor of probability A2 is given by probability $\theta_1 \neq 0$, which can be interpreted as the "availability" of the ligand. It incorporates the spatial availability, geometric orientation of the ligand to the binding site and how "costly" it is to decouple the ligand from its environment (e.g. the energy required to remove hydrogen bonds between the ligand and the solvent molecules). In the case of a chosen site being occupied, probability θ_2 characterizes

the barrier of releasing the ligand molecule. In most cases θ_2 can be considered as being equal to 1. However, e.g. in supersaturated solutions or due to weak solubility of the ligand, the release of a ligand molecule might be of energetic disadvantage for the environment. Both factors θ_1 and θ_2 depend on the ligand concentration and describe the energetic state of the environment.

Ad **A2-3** : The third and last factor $p_{k,l}$ models the probability barrier given by the energy difference of the target molecule, when a ligand is released or taken up. In contrast to A2-2, this factor is not assumed to depend on the environment, i.e. on the energy state of the solution. Let $G(k), G(l)$ denote the energy levels of the states k and l. We are looking for a function $p_{k,l} := p(G(k), G(l)) \longrightarrow [0,1]$ with $p_{k,l} = 1$ if $G(l) \leq G(k)$. This means that if the energy level is the same, or is reduced by the transition, there will not be an energy barrier that impedes the transition (expressed as a probability). However, if energy is required, i.e. $G(l) > G(k)$, then $p_{k,l} < 1$. Since $p_{k,l}$ is a probability, it can be represented by

$$p_{k,l} = \min(1, f(G(l) - G(k))) \tag{2.1}$$

for an appropriate nonnegative function $f(x)$ which is different from the zero function and which depends only on the energy differences. Some properties of f are reasonable to assume

$$f(x+y) = f(x)f(y), \tag{2.2}$$

$$f(x) \in (0,1) \text{ if and only if } x \in (0,\infty), \tag{2.3}$$

$$f \text{ is monotone.} \tag{2.4}$$

The first property models that an additional energy barrier represents a second factor: The probability of overcoming a barrier $x + y$ shall be equal to the probability of overcoming x and subsequently y. This characteristic of function f is also required for consistency with possible extensions of this model by incorporating intermediate states. The second property expresses that only a transition that requires energy poses a probability barrier. Monotonicity is reasonable, too.

Lemma 3. *a) A value $\beta \in \mathbb{R}^+$ exists such that $f(x) = \exp(-\beta x)$.*
b) $p_{k,l} < 1 \Longrightarrow p_{l,k} = 1$.

Proof. a) This part is well-known but the proof is sketched for the sake of completeness. Due to Eq. (2.3), we can apply the natural logarithm (on $x \in (0,\infty)$) leading to Cauchy's functional equation, which means $\ln(f)$ is linear, according to Eq. (2.2). Thus, f is an exponential function with $-\beta < 0$, according to Eq. (2.3). Part b) is a direct consequence of a). □

2.2.1 The transition probabilities

With assumptions A1-A3 we obtain, for a certain molecule M_1, a Markov chain $M_{1,m}$ on the set of states $\{0,1\}^n$, with n denoting the number of binding sites. For $|k| < |l|$ and $|l - k| = 1$, the transition probabilities are given by

$$q_{k,l} := P(k \mapsto l) = \frac{1}{n}\theta_1 p_{k,l}, \tag{2.5}$$

where θ_1 denotes the "availability" of the ligand. If $|k| > |l|$, $|l-k| = 1$:

$$q_{k,l} = \frac{1}{n}\theta_2 p_{k,l} \tag{2.6}$$

with θ_2 denoting the "resistance". The probability of staying in the present state l is:

$$q_{l,l} = 1 - \sum_{k \neq l} q_{l,k} =$$

$$= 1 - \frac{1}{n}\left(\theta_1 \sum_{\{k\,||k|>|l|,|k-l|=1\}} p_{l,k} \;+\; \theta_2 \sum_{\{k\,||k|<|l|,|k-l|=1\}} p_{l,k}\right) \tag{2.7}$$

Example 4. *For a molecule with two binding sites for ligand L, we use the notation* $0 := (0,0), 1 := (0,1), 2 := (1,0), 3 := (1,1)$ *as a new composite index. The matrix of transition probabilities is*

$$\begin{pmatrix} 1 - \frac{1}{2}(\theta_1(p_{0,1} + p_{0,2})) & \frac{1}{2}\theta_1 p_{0,1} & \frac{1}{2}\theta_1 p_{0,2} & 0 \\ \frac{1}{2}\theta_2 p_{1,0} & 1 - \frac{1}{2}(\theta_1 p_{1,3} + \theta_2 p_{1,0}) & 0 & \frac{1}{2}\theta_1 p_{1,3} \\ \frac{1}{2}\theta_2 p_{2,0} & 0 & 1 - \frac{1}{2}(\theta_1 p_{2,3} + \theta_2 p_{2,0}) & \frac{1}{2}\theta_1 p_{2,3} \\ 0 & \frac{1}{2}\theta_2 p_{3,1} & \frac{1}{2}\theta_2 p_{3,2} & 1 - (\frac{1}{2}\theta_2(p_{3,1} + p_{3,2})) \end{pmatrix}$$

2.2.2 Aperiodicity, connectivity and detailed balance

We know that the Markov chain with transition probabilities defined by Eqs. (2.5-2.7) is aperiodic and connected. Aperiodicity can clearly be seen because the system can return to its initial state within one time step, which means it remains in this state, or in two time steps by going there and back. Connectivity is also obvious since every state can be reached. Consequently, the Markov chain has a unique stationary distribution π to which the system's distribution will converge and which we will characterize. If the matrix fulfills the detailed balance condition, we will be able to calculate the stationary distribution quickly, according to the procedure described in the following lemma.

Lemma 5. *Let* $Q = (q_{i,j})_{i,j\in\{1,\ldots,n\}}$ *be a transition matrix on a connected space. Moreover, let* $\pi = (\pi_1, \ldots, \pi_n)$ *denote its unique stationary distribution fulfilling the detailed balance condition. Then the stationary distribution can be calculated in the following way:*

1) *Choose a reference state* k*, and define* $\pi_k = 1$*.*

2) *Calculate the ratios* $\frac{\pi_i}{\pi_k}$ *of all pairs* $\{i,k\}$ *with* $q_{i,k} \neq 0$ *by* $\frac{\pi_i}{\pi_k} = \frac{q_{k,i}}{q_{i,k}}$*.*

3) *If* $q_{i,k} = 0$ *choose any path* $(i, \ldots, k)$ *with probability greater than zero and calculate the pairwise ratio.*

4) *Normalize the distribution.*

Proof. First note that $\pi_i \neq 0 \ \forall i \in \{1, \ldots, n\}$, since the space is connected. The described procedure gives the stationary distribution because the detailed balance condition means

$$\pi_i q_{i,k} = \pi_k q_{k,i},$$

which gives the ratio $\frac{\pi_i}{\pi_k}$ if $q_{i,k} \neq 0$. Since the space is connected, a path from i to k exists with probability greater than zero. Thus, if $q_{i,k} = 0$, we can calculate the ratios pairwise "along the path" to calculate the ratio $\frac{\pi_i}{\pi_k}$. □

In other words, Lemma 5 states that for a given reference state, the ratio of the probabilities of the stationary distributions are identical to the ratios of the expected flux between two states (pairwise) along any path. This statement is actually one direction of Kolmogorov's criterion (Kelly, 2011). Even though it is not obvious that the matrix of Example 4 satisfies the detailed balance equation, we will use the procedure of Lemma 5 and show that the obtained distribution is stationary, for the special case of two binding sites.

Example 6. *We use the same abbreviations for the different states as in Example 4. We calculate the probabilities of the stationary distribution the following way:*

$$\pi(0) \propto 1$$

$$\pi(1) \propto \frac{\frac{1}{2}\theta_1 p_{0,1}}{\frac{1}{2}\theta_2 p_{1,0}} = \frac{\theta_1 p_{0,1}}{\theta_2 p_{1,0}}$$

$$\pi(2) \propto \frac{\frac{1}{2}\theta_1 p_{0,2}}{\frac{1}{2}\theta_2 p_{2,0}} = \frac{\theta_1 p_{0,2}}{\theta_2 p_{2,0}}$$

$$\pi(3) \propto \frac{\frac{1}{2}\theta_1 p_{2,3} \cdot \pi(2)}{\frac{1}{2}\theta_2 p_{3,2}} = \frac{\theta_1^2 p_{0,2} p_{2,3}}{\theta_2^2 p_{3,2} p_{2,0}},$$

where $\propto$ means proportional to (of course with the same factor for all equations).

For the weights of Example 6 it is not obvious that we would obtain the same probability distribution if we compared $\pi(3)$ with $\pi(1)$:

$$\pi(3) = \frac{\theta_1^2 p_{0,1} p_{1,3}}{\theta_2^2 p_{3,1} p_{1,0}}.$$

To see that the weights do not depend on the choice of the path, Lemma 7 which is an obvious result from the definition of $p_{i,j}$ is helpful. It will also be used subsequently to show that our model satisfies the detailed balance condition for any number of binding sites.

Lemma 7.

$$\frac{p_{i,j}}{p_{j,i}} = f(G(j) - G(i)) \tag{2.8}$$

This statement will be used to prove Proposition 8.

Proposition 8. *For every number of binding sites n, the matrix of transition probabilities defined by Eqs. (2.5-2.7) is detailed balanced with respect to its stationary distribution.*

Proof. We use Kolmogorov's criterion (Kelly, 2011), which (in simple words) states that a stochastic matrix and its stationary distribution fulfill the detailed balance condition if and only if the probability for "walking on a closed path" is independent of the direction. More precisely, this means, the matrix $(q_{i,j})$ fulfills the detailed balance condition if and only if

$$q_{k,i_1}q_{i_1,i_2}\cdots q_{i_{r-1},i_r}q_{i_r,k} = q_{k,i_r}q_{i_r,i_{r-1}}\cdots q_{i_2,i_1}q_{i_1,k} \tag{2.9}$$

for any path $(k, i_1, i_2, \ldots, i_r, k)$ and any $r \in \mathbb{N}$. We show that the matrix of transition probabilities defined by Eqs. (2.5-2.7) satisfies Eq. (2.9) and firstly identify $q_{k,l} = P(k \mapsto l)$. Let a closed path $(k = i_0, i_1, i_2, \ldots, i_r, i_{r+1} = k)$ be given. First note that if a path includes a step which changes the state of more than one binding site, both directions will have probability zero, since $q_{j,l} = 0 = q_{l,j}$ if $|j - l| > 1$. The probability of all other transitions from j to l with $|j - l| \leq 1$ are nonzero, since all factors which the probabilities $q_{j,l}$ are built of are nonzero. Moreover, if at a certain step, the state is not changed, the factor $q_{j,j}$ cancels out on both sides. Thus, without loss of generality, every step of the path changes the state, that is $i_j \neq i_{j+1} \forall j \in \{0, \ldots, r\}$. Since every probability $q_{i,j}$ includes the factor $\frac{1}{n}$ on both sites, it cancels out. Moreover, since the path is closed, the power of the factor θ_1 on one side of the equation is equal to the power of θ_2 (we return to the initial state, every ligand which is taken up has to be released afterwards). Using the other "direction" of the path every factor θ_1 of the left side will be substituted by a factor θ_2. However, since both factors have the same power, they all cancel. The remaining factors are given by $p_{i,j}$ and we see, that the matrix $(q_{i,j})$ satisfies Eq. (2.9) if and only if

$$\frac{p_{k,i_1}p_{i_1,i_2}\cdots p_{i_{r-1},i_r}p_{i_r,k}}{p_{k,i_r}p_{i_r,i_{r-1}}\cdots p_{i_2,i_1}p_{i_1,k}} = 1,$$

which is true since Lemma 7 states that the left site is equal to $f(0) = 1$. □

Remark 9. *In our model, the probability of a transition from k to l, with $|k - l| = 1$ is composed of a uniform proposal distribution on the states of the "neighborhood" and of an acceptance rate given by $\theta_1 p_{i,j}$ or $\theta_2 p_{i,j}$, depending on the state of the chosen site. Even though this structure resembles the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970), our model does not coincide with this algorithm: The factor θ_i is not part of the proposal distribution, since otherwise, the proposal probabilities do not sum up to one. Consequently, the acceptance probability is different to the one commonly used for the Metropolis-Hastings algorithm, since it is bounded by θ_i.*

2.2.3 The stationary distribution

Proposition 10. *The stationary distribution on the set of states is given by a normalized version of*

$$P(l) = \left(\frac{\theta_1}{\theta_2}\right)^{|l|} f(G(l) - G(\{0\}^n)). \tag{2.10}$$

Proof. We know that the Markov chain fulfills the detailed balance condition. Using Lemma 5 with the reference state $\{0\}^n$, we receive Eq. (2.10). □

Since we assumed the molecules to bind ligands independently (A3), the distribution of the states within the solution in equilibrium will be close to the stationary distribution of a single molecule, due to the Law of Large numbers, if the number of molecules is sufficiently large.

2.2.4 Activation energies

In the model presented in Section 2.2.1 we did not incorporate activation energy barriers. However, an extension of our model is straightforward: Assuming, that an activation energy barrier between two states i, j is a "symmetric" barrier, given by an instable transition intermediate state $e_{i,j}$, we can rewrite Eq. (2.1):

$$p_{i,j} = \min(1, f(G(e_{i,j}) - G(i))) \cdot \min(1, f(G(j) - G(e_{i,j}))). \tag{2.11}$$

Assuming that $e_{i,j}$ has an energy level higher than those of states i, j (activation energy, instable state), the second factor equals 1. This gives for the ratios

$$\frac{p_{i,j}}{p_{j,i}} = \frac{\min(1, f(G(e_{i,j}) - G(i)))}{\min(1, f(G(e_{i,j}) - G(j)))} = f(G(j) - G(i)) \tag{2.12}$$

This result shows that we can add any additional "symmetric" probability barriers and the stationary distribution will be unchanged.

2.3 Comparison to the Grand Canonical Partition Function

As mentioned in the introduction, the Grand Canonical Partition Function (or "binding polynomial" for a finite number of binding sites) is usually formulated as a function in the variable "chemical activity" which is denoted by λ:

$$\sum_{k \in K} f(G(k) - G(\{0\}^n))\lambda^{|k|}. \tag{2.13}$$

It coincides with the stationary distribution of our model if we identify $\frac{\theta_1}{\theta_2} =: \lambda$. Thus, chemical activity might be interpreted as the ratio between "availability" and "resistance", in our model.

2.4 Comparison of the dynamics defined by the presented transition matrix to that induced by the commonly used matrix for the Metropolis-Hastings algorithm

As already mentioned in Section 2.1, the presented model can also be useful for modeling the way of a system into equilibrium. However, since there is an infinite number of transition matrices with the same stationary distribution, it is neither clear which one is really appropriate to model the dynamics of ligand binding, nor which characteristics can be used to decide which transition matrix is "best". For this reason, the two different models which were mentioned in this chapter –the presented model and the transition matrix of the common Metropolis-Hastings algorithm– will be compared in an example to illustrate differences between the corresponding binding dynamics.

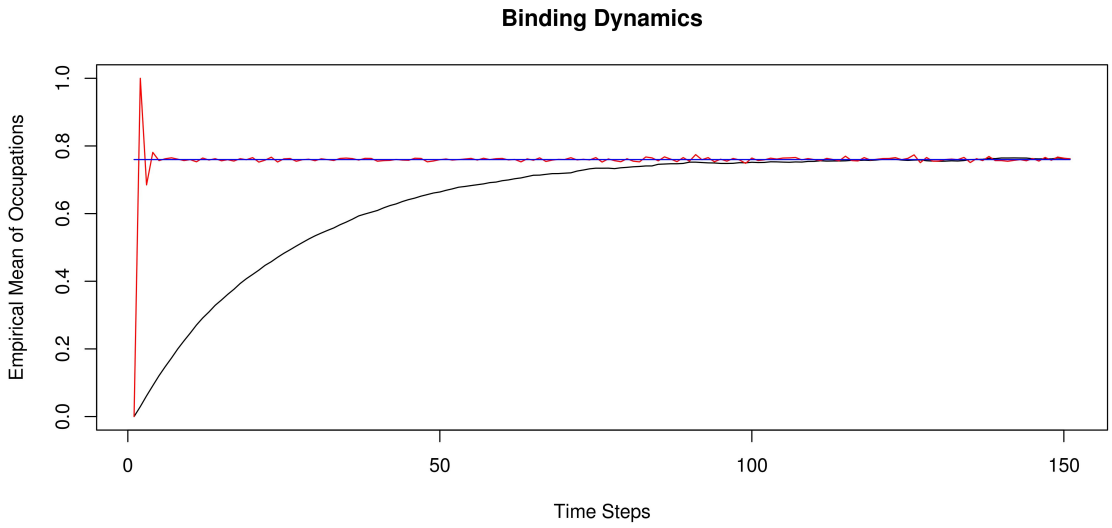

Figure 2.1: Simulated dynamics of a system consisting of 10^4 target molecules, each with one binding site, starting at time $t = 1$ from the state in which all molecules are unoccupied, over 150 time steps. The configuration $\lambda = 10^{-1.5}, g_1 = 10^2$ was chosen. **Black line:** Dynamics with the presented model (P_1), **red line:** Dynamics with the common Metropolis-Hastings matrix (P_2), **blue line:** Equilibrium.

Example 11. *In this example the way into equilibrium of a system consisting of n target molecules, each with one binding site for its ligand, is simulated starting from the state in which all of them are unoccupied. We choose the reference 0 and $f(G(1))$ to be $g_1 \geq 1$ (the occupied state has a lower energy level than the unoccupied state), arbitrarily. The stationary distribution on the two possible states is given by*

$$P(0) = \frac{1}{g_1\lambda + 1} \qquad \text{and} \qquad P(1) = \frac{g_1\lambda}{g_1\lambda + 1}, \tag{2.14}$$

with 0 denoting the unoccupied and 1 the occupied state. Assuming that $\theta_2 = 1$ and thus $\theta_1 = \lambda$, the transition matrix of the presented model is

$$P_1 := (p_{i,j})_{i,j\in\{1,2\}} = \begin{pmatrix} 1-\lambda & \lambda \\ \frac{1}{g_1} & 1-\frac{1}{g_1} \end{pmatrix} \tag{2.15}$$

where $p_{i,j}$ denotes the probability of a transition from state $i-1$ to state $j-1$. The matrix of the common choice for the Metropolis-Hastings algorithm, with a proposal distribution which always proposes the other state, is

$$P_2 := \begin{pmatrix} 1-\min(1, g_1\lambda) & \min(1, g_1\lambda) \\ \min(1, \frac{1}{g_1\lambda}) & 1-\min(1, \frac{1}{g_1\lambda}) \end{pmatrix}. \tag{2.16}$$

To point out the major differences between the dynamics defined by the matrices P_1 and P_2, an example was calculated with $\lambda = 10^{-1.5}, g_1 = 10^2, n = 10^4$ target molecules and

$m = 150$ *time steps. How the system reaches equilibrium, following the two different transition matrices, is illustrated in Fig. 2.1.*

One thing which can be observed in this example is that the system described by the commonly used transition matrix of the Metropolis-Hastings algorithm, is close to equilibrium after five time steps. This characteristic is of advantage for the usual purpose of "drawing" from its stationary distribution. However, for describing the dynamics of real phenomena it might not be appropriate, since this fact makes a scaling of the discrete time difficult: No matter how short a time step is, the system is quickly close to equilibrium. A second characteristic of the dynamics of a system described by P_2 is the huge overrun after the first time step. Both characteristics are results of the same property of P_2: a negative second eigenvalue with comparatively small absolute value. A small absolute second eigenvalue characterizes the convergence rate, which will be illustrated in Example 13, after a helpful lemma.

Lemma 12. *Let $P \in R^{n\times n}$ be a stochastic matrix and $\eta = (\eta_1, ..., \eta_n)$ be a left-eigenvector for the eigenvalue e: Then*

$$e = 1 \qquad or \qquad \sum_{i=1}^{n} \eta_i = 0.$$

Proof. Since P is stochastic

$$\sum_{i=1}^{n} \eta_i = \eta \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \eta P \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = e\eta \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = e \sum_{i=1}^{n} \eta_i.$$

□

Example 13. *Let us regard the example of a stochastic matrix $P \in \mathbb{R}^{2\times 2}$ of a connected Markov chain, with stationary distribution π and a second eigenvector η with corresponding eigenvalue $e_\eta \neq 1$. Since $e_\eta \neq 1$, we know from Lemma 12 that the sum of the entries of η is 0. For a starting distribution $v = (v_1, v_2)$ we know that coefficients $\alpha_1, \alpha_2 \in \mathbb{R}$ exist such that $v = \alpha_1\pi + \alpha_2\eta$, since the span $\langle \pi, \eta \rangle = \mathbb{R}^2$. The equation*

$$1 = v_1 + v_2 = \alpha_1 \underbrace{(\pi_1 + \pi_2)}_{=1} + \alpha_2 \underbrace{(\eta_1 + \eta_2)}_{=0}$$

shows that $\alpha_1 = 1$. The convergence of the vector $vP^n \to \pi$ is illustrated by the equation

$$vP^n = (\pi + \alpha_2\eta)P^n = \pi + \alpha_2 e_\eta^n \eta$$

in which the summand $\alpha_2 e_\eta^n \eta$ converges to 0 as $n \to \infty$. The speed of the convergence is determined by the absolute value of e_η.

Example 13 illustrated –for the case of matrices in $\mathbb{R}^{2\times 2}$– why the absolute value of the second eigenvalue is a characteristic of the speed of convergence to the stationary distribution (for analogous statements for larger matrices and more general statements see e.g. Seneta (2006)). We will compare some properties of the dynamics described by P_1 and P_2 and their dependence on the chemical activity in Proposition 14. In the following, let e_i denote the second eigenvalue of P_i, respectively.

Proposition 14. *a) If $\lambda = g_1^{-1}$, then $e_2 = -1$ and the system described by P_2 is periodic. It will not generally converge to its stationary distribution.*

b) If $g_1\lambda > 1$, then $e_2 = -\frac{1}{g_1\lambda}$ and if $g_1\lambda < 1$, then $e_2 = -g_1\lambda$. In particular, e_2 is always smaller than or equal to zero.

For P_1, let $\theta_2 = 1$ and thus $\lambda = \theta_1 \leq 1$. Then the following statements hold:

c) If $g_1 = \lambda = 1$, then $e_1 = -1$ and the system described by P_1 is periodic. It will not generally converge to its stationary distribution.

d) If $g_1 \geq 1$, then $e_1 = 1 - \frac{1}{g_1} - \lambda$ and if $g_1 < 1$, then P_1 is identical to P_2 and $e_1 = e_2 = -g_1\lambda$.

e) $e_1 \geq e_2$ $\forall \lambda, g_1$.

f) Let $0 < \lambda < 1$. Moreover, let $[a, b] \subset (0, 1]$ denote the interval in $(0, 1]$, on which $|e_1| \geq |e_2|$ (as functions of λ). Then, for $g_1 \to \infty : a \to 0$ and $b \to 1$.

Proof. a) If $\lambda = g_1^{-1}$, P_2 has the entry 1 on its secondary diagonal, and 0 on its diagonal. Its characteristic polynomial is $P_2(t) = t^2 - 1$, which has the roots $t_{1,2} \in \{\pm 1\}$. The lengths of all closed paths can be divided by 2, consequently the greatest common divisor of the lengths of closed paths is 2 and the Markov chain is periodic. Starting with any distribution will lead to the same distribution within two steps. The distribution will not converge.

b) If $g_1\lambda > 1$, the characteristic polynomial is given by $P_2(t) = t(t - 1 + \frac{1}{g_1\lambda}) - \frac{1}{g_1\lambda}$. Calculate its roots. Analogously for $g_1\lambda < 1$.

c) If $g_1 = \lambda = 1$, we find the matrix described in a).

d) $\theta_2 = 1$ implies $\lambda = \theta_1 \leq 1$. If $g_1 \geq 1$, the characteristic polynomial of this matrix is $P_1(t) = (1 - \lambda - t)(1 - \frac{1}{g_1} - t) - \frac{\lambda}{g_1}$. Its roots are given by $t_{1,2} \in \{1, 1 - \frac{1}{g_1} - \lambda\}$. If $g_1 < 1$, the matrix P_1 coincides with P_2 (Note here that Eq. (2.15) is P_1 for the case of $g_1 \geq 1$).

e) We know from d) that if $g_1 < 1$, the matrices are identical and consequently $e_1 = e_2$. For $g_1 \geq 1$, we distinguish between three cases: $g_1\lambda < 1$, $g_1\lambda > 1$ and $g_1\lambda = 1$.
So let $g_1\lambda < 1$: What has to be shown is

$$-g_1\lambda \leq 1 - \frac{1}{g_1} - \lambda$$

If $\lambda = 0$ this relation is true. For $\lambda > 0$ we can rewrite this inequality as

$$0 \leq \underbrace{1 - \frac{1}{g_1}}_{\geq 0} + \underbrace{(g_1 - 1)}_{\geq 0}\lambda.$$

The right side describes a linear equation with nonnegative slope which is nonnegative if $\lambda \in (0, 1]$. Consequently, the statement is true for this case.
Let $g_1\lambda > 1$: What has to be shown is

$$-\frac{1}{g_1\lambda} \leq 1 - \frac{1}{g_1} - \lambda \text{ which is equivalent to } 0 \leq g_1\lambda - \lambda - g_1\lambda^2 + 1.$$

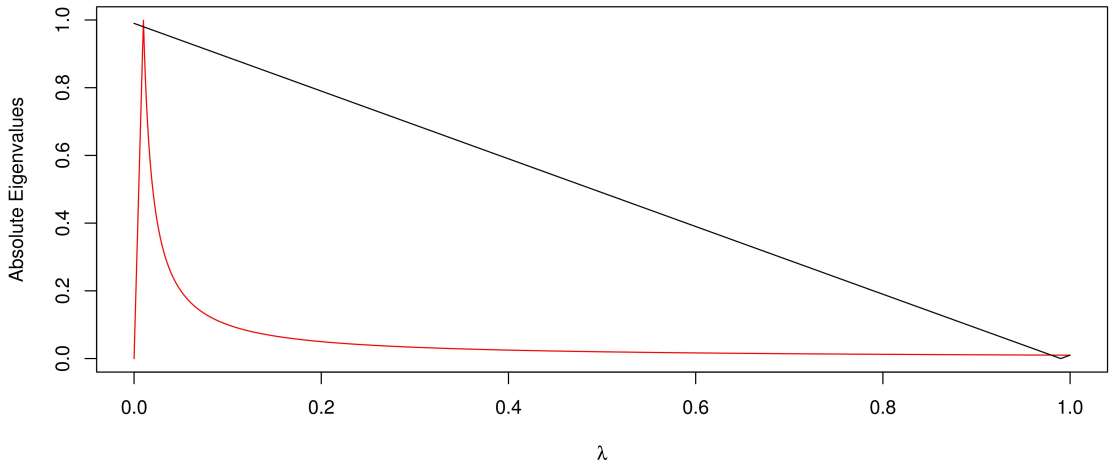

Figure 2.2: The absolute second eigenvalues of P_2 **(red line)** and P_1 **(black line)** as a function of λ for fixed $g_1 = 10^2$ (Example 15). Note that e_2 is always smaller than or equal to 0, which means e_2 is the negative of the red line.

The right side is a polynomial of degree two in the variable λ. Its leading coefficient is negative, it has value 1 at $\lambda = 0$ and 0 at $\lambda = 1$. Thus, it is nonnegative in between.
Let $g_1\lambda = 1$. Then $e_2 = -1$ which is the smallest second eigenvalue a stochastic matrix can have.

f) Let g_1 grow. Then, for g_1 large enough, $g_1\lambda > 1$ and $0 \leq 1 - \frac{1}{g_1} - \lambda$. What has to be shown is that the set on which

$$\frac{1}{g_1\lambda} \leq 1 - \frac{1}{g_1} - \lambda \text{ which is equivalent to } 0 \leq g_1\lambda - \lambda - g_1\lambda^2 - 1$$

tends to $(0, 1]$. The right side is a polynomial of degree two in the variable λ. Its leading coefficient is negative and the interval on which the inequality is true is bounded by the polynomial's roots. Its roots are given by

$$-\frac{1}{2g_1}(-g_1 + 1 \pm \sqrt{g_1(g_1 - 6) + 1}).$$

We can rewrite this to

$$\frac{1}{2} - \frac{1}{2g_1} \pm \sqrt{\frac{1}{4} - \frac{6}{4g_1} + \frac{1}{4g_1^2}}$$

Consequently, for growing g_1, the roots tend to $\{0, 1\}$, respectively.

□

Proposition 14 showed that the presented model has several advantages compared to the commonly used matrix for the Metropolis-Hastings algorithm if the dynamics of

a system into equilibrium shall be modeled: In the case of $g_1 \neq 1$, it will converge to the stationary distribution for any value $\lambda \in [0,1]$. In contrast, P_2 produces a periodic Markov chain at an important value of λ: the pK_a-value $\lambda = g_1^{-1}$, which is used in chemistry as a characteristic of the titration curve. Moreover, it was shown, that the dynamics described by P_2 always tend to create fluctuations due to the negative second eigenvalue. Concerning the convergence rate, part f) of Proposition 14 showed, that for growing g_1, the interval on which $|e_2| \leq |e_1|$ grows, too. As a final illustrating example of this chapter, the interval on which $|e_2| \leq |e_1|$ will be calculated exactly, for the setup of Example 11.

Example 15. *In this example, it will be demonstrated, for which values of λ, the absolute value of e_1 is equal or greater than that of e_2 in the setup of Example 11. The probability constant was given by $g_1 = 10^2$. Thus, we have to distinguish between the cases $\lambda < 10^{-2}$ and $\lambda > 10^{-2}$. Recall that if $\lambda = 10^{-2}$, the system described by P_2 will not converge, according to Proposition 14 a).*
Let $\lambda < 10^{-2}$ and thus $g_1\lambda < 1$. We are interested in the values of λ which satisfy

$$100\lambda \leq \frac{99}{100} - \lambda.$$

This is true if and only if $\lambda \leq \frac{99}{10100}$ which holds for the bigger part of the interval $(0, 10^{-2})$.
Let $\lambda > 10^{-2}$, then $g_1\lambda > 1$ and we are interested in the values of λ which satisfy

$$\frac{1}{100\lambda} \leq \left| \frac{99}{100} - \lambda \right|.$$

This is true on the closed interval $[a, b]$ with a, b the roots of $P(\lambda) = -100\lambda^2 + 99\lambda - 1$. This gives the approximated interval of $[0.0102, 0.9798] \cup \{1\}$. For the comparison of the absolute second eigenvalues, see Fig. 2.2.

3 The Decoupled Sites Representation

The following chapter is essentially the paper Martini and Ullmann (2013) with Section 3.7 being based on Martini et al. (2014). At certain points the text was altered and some arguments in proofs were changed. Moreover, new content was added: In Section 3.2.1 binding and interaction energies are introduced in a different way, Proposition 19 was extended by the reverse statement, and Definition 21, Section 3.4 and Propositions 29 and 32 were added.

3.1 Motivation

The investigation of activity-dependent average binding of ligands to binding sites of a molecule in equilibrium is a key field of biochemistry, since many biological processes are based on or regulated by non-covalent binding processes. For instance, the binding of protons to proteins can change the charge distribution across the macromolecule and thus affect the catalytic center of an enzyme, the affinity to the substrate (or to another type of ligand), and the target molecule's tertiary structure (Garcia-Moreno, 1995). Moreover, electron or proton transport chains in oxidation processes and photosynthesis can be described by binding properties of the carrier proteins (Becker et al., 2007; Medvedev and Stuchebrukhov, 2006; Till et al., 2008).

As mentioned in the introduction, the titration curve of a molecule with only one binding site is described by a Henderson-Hasselbalch (HH) curve (Hasselbalch, 1916; Henderson, 1913, Eq. (1.10)). The mathematical description of titration curves of individual sites of macromolecules is much more complicated than that of small molecules with only one binding site. The more complicated shape of the titration curve of macromolecules is a result of interaction between the different sites (Bombarda and Ullmann, 2010; Onufriev and Ullmann, 2004; Tanford and Kirkwood, 1957). This interaction can lead to titration curves exhibiting strong deviations from the classical Henderson-Hasselbalch shape (Ackers et al., 1983; Bashford and Karplus, 1991; Bombarda and Ullmann, 2010; Onufriev and Ullmann, 2004).

The Decoupled Sites Representation (DSR) was developed to find a simple expression for the complex shape of overall titration curves and of titration curves of the individual sites (Onufriev et al., 2001; Onufriev and Ullmann, 2004). The main result is that for any given macromolecule with n ligand binding sites and interaction between these sites, a hypothetical molecule with n non-interacting binding sites exists which possesses the same overall titration curve. "Existence" refers here to the mathematical point of view, in terms of a tuple of n binding energies and not in terms of a certain chemical structure. Even though Onufriev et al. (2001) described how to calculate the corresponding decoupled system, the mathematical structure of the problem was not investigated deeply. In this chapter, we investigate the mapping of binding energies and

interaction energies of a molecule to its binding polynomial. A first result is that all molecules sharing the same binding polynomial form an algebraic variety. Due to the special structure of the equations defining the sub-variety consisting of the molecule without interaction, no methods of algebraic geometry are required to prove the existence and uniqueness (with respect to an equivalence relation) of the decoupled system. Moreover, we show that the DSR can be generalized: For a given binding polynomial, it is not only possible to find a system without interaction with this binding polynomial, but it is possible to find systems with any interaction energy t with the same binding polynomial as long as the interaction energy is the same between all sites. We call this extension the generalized DSR, which shows that the DSR is not a special result of the "lack of interaction", but a consequence of identical interaction energies.

3.2 A molecule and its binding polynomial

3.2.1 Molecules with several ligand binding sites

In general, the binding properties of a ligand to its target molecule in equilibrium can be described by the relative energy levels of the microstates of the target molecule (see Chapter 2). For instance, for a molecule with three binding sites, the binding behavior in equilibrium is described by the relation of the energies of the eight possible microstates. Choosing the completely unoccupied state as the reference, we call the energies of microstates with only one site occupied the binding energy G_i of site i:

$$G((0,0,0)) = 0 \qquad G((1,0,0)) =: G_1 \qquad G((0,1,0)) =: G_2 \qquad G((0,0,1)) =: G_3.$$

The difference between the energy level of a state with the first and the second sites occupied $(1,1,0)$ and the sum of the two energies of the corresponding microstates with single occupation is called the interaction energy $W_{1,2}$ (Ben-Naim, 2001; Cantor and Schimmel, 1980; Wyman and Gill, 1990):

$$G((1,1,0)) - G_1 - G_2 =: W_{1,2}.$$

Thus, all states with exactly two occupied sites define all pairwise interaction energies. In the case of a target molecule with three or more binding sites, interaction terms of higher order can be required to describe the system completely:

$$G((1,1,1)) - G_1 - G_2 - G_3 - W_{1,2} - W_{1,3} - W_{2,3} =: W_{1,2,3}.$$

However, in the following chapter, a model which incorporates only pairwise interaction terms is underlying. This means that all interaction energies of higher order are assumed to be zero and the energy levels of all microstates are determined by the energies of the microstates with not more than two bound ligands. This reduction facilitates notation and the major statements of this chapter are likewise true for the generalized model with interaction energies of higher order.

Let us regard a target molecule M with n binding sites for ligand L. The ligand binding properties of the molecule can be characterized by n binding energies $G_1, ..., G_n$ and $\frac{n(n-1)}{2}$ pairwise interaction energies $W_{1,2}, ..., W_{1,n}, ..., W_{n-1,n}$, where $W_{i,j}$ is the interaction energy of the i-th and j-th ligand binding site. Since titration properties are

determined by these energies, every other system of ligand binding sites with identical binding and interaction energies, is regarded as being equal to M, since the average overall titration as well as the titration curve of every individual site are identical. Thus, every molecule can be identified with at least one element

$$M \in \mathbb{R}^{\frac{n(n+1)}{2}}. \tag{3.1}$$

(We will show later that, to use the DSR as a general tool, it is necessary to extend the domain of energies from $\mathbb{R}$ to a stripe within $\mathbb{C}$.) Moreover, there is no natural order of the binding sites within a molecule. Thus, one has to notice that one and the same molecule can be identified with several tuples which is illustrated by Example 16 .

Example 16. *Let $n = 3$. Then the tuple*

$$(G_1, G_2, G_3, W_{1,2}, W_{1,3}, W_{2,3}) = (1, 2, 3, 1, 2, 3)$$

and

$$(G_1, G_2, G_3, W_{1,2}, W_{1,3}, W_{2,3}) = (2, 3, 1, 3, 1, 2)$$

belong to the same molecule.

This is a result of the disorderliness of the binding sites: One can number the binding sites in any order. However, it is important that the interaction energies are permuted accordingly. This property motivates the definition of an equivalence relation.

Definition 17. *Let*

$$a = (G_1^a, G_2^a, ...G_n^a, ...W_{n-1,n}^a)$$

and

$$b = (G_1^b, G_2^b, ...G_n^b, ...W_{n-1,n}^b) \in \mathbb{R}^{\frac{n(n+1)}{2}}.$$

Then a is equivalent to b (Notation: $a \sim b$) if and only if a permutation σ of $(1, ..., n)$ exists such that

$$a = (G_{\sigma(1)}^b, G_{\sigma(2)}^b, ...G_{\sigma(n)}^b, W_{\sigma(1),\sigma(2)}^b, W_{\sigma(1),\sigma(3)}^b, ..., W_{\sigma(n-1),\sigma(n)}^b). \tag{3.2}$$

This means that every molecule can be identified with exactly one element

$$M \in {}^{\mathbb{R}^m}\!/_{\sim} \tag{3.3}$$

with $m = \frac{n(n+1)}{2}$.

Remark 18. *a) The reader might ask why the new notation with binding and interaction energies was introduced: Of course it is as well possible to choose an order for the 2^n microstates and to describe the molecule by the tuple of the corresponding 2^n energy levels (or reduced to a model with only pairwise interaction which means describing the energy levels of states with only one or two sites occupied). However, the distinction between "binding" and "interaction" energies is used in literature (e.g. Ben-Naim, 2001; Onufriev et al., 2001) and it is of advantage for descriptions and definitions.*
b) Concerning the equivalence relation: At several points in this work, a tuple notation will be used. However, in most cases a tuple has to be interpreted as a representative of an equivalence class, which means that the corresponding statement is true for the equivalence class (the molecule).

Having defined the set of molecules, we investigate how an element M is mapped to its binding polynomial, which is an element of $\mathbb{R}[\lambda]$ (the polynomial ring in one variable λ and coefficients in $\mathbb{R}$).

3.2.2 The binding polynomial

Recall, that the binding polynomial of a molecule M was defined by

$$\Phi(M) := \sum_{k \in K} \exp\left(\frac{-G(k)}{RT}\right) \lambda^{|k|} \tag{3.4}$$

in Subsection 1.3.2 (Schellman, 1975; Wyman and Gill, 1990). Recall that $g(k) = \exp\left(\frac{-G(k)}{RT}\right)$ is called the **Boltzmann factor** or –in this work– also the **probability constant** of state k. Analogously, $g_i := \exp\left(\frac{-G_i}{RT}\right)$ is called the **binding constant** of site i and $w_{i,j} := \exp\left(\frac{-W_{i,j}}{RT}\right)$ the **interaction constant** of sites i, j.
Due to the reduction to only pairwise interaction terms, the energy of a microstate $k = (k_1, ..., k_n)$ is given by

$$G(k) := \sum_{i=1}^{n} k_i G_i + \sum_{i=1}^{n} \sum_{j>i}^{n} k_i k_j W_{i,j}. \tag{3.5}$$

Analogously, the probability constant of microstate k (see Section 1.3.2), writes

$$g(k) = \prod_{i=1}^{n} \left(\exp\left(\frac{-k_i G_i}{RT}\right) \prod_{i<j} \exp\left(\frac{-k_i k_j W_{i,j}}{RT}\right) \right). \tag{3.6}$$

Since, the structure of the energies of a microstate (in Eq. (3.5)) translates to products for the constants (in Eq. (3.6)), we can factor the map

$$\Phi : \ {}^{\mathbb{R}^m}\!/_{\sim} \ \longrightarrow \ \mathbb{R}[\lambda]$$

$$M \mapsto \Phi(M)$$

which maps the tuple of energies to the binding polynomial into $\Phi = \Phi_2 \circ \Phi_1$ with

$$\Phi_1 : \ {}^{\mathbb{R}^m}\!/_{\sim} \ \longrightarrow \ {}^{\mathbb{R}^{+m}}\!/_{\sim} \tag{3.7}$$

$$(G_1, ...G_n, ...W_{n-1,n}) \mapsto \left(\exp\left(-\frac{G_1}{RT}\right), .., \exp\left(-\frac{G_n}{RT}\right), ..., \exp\left(-\frac{W_{n-1,n}}{RT}\right) \right)$$

mapping the energies to binding and interaction constants, and

$$\Phi_2 : \ {}^{\mathbb{R}^{+m}}\!/_{\sim} \ \longrightarrow \ \mathbb{R}[\lambda] \tag{3.8}$$

$$(g_1, g_2, ..., g_n, ..., w_{n-1,n}) \mapsto \sum_{k \in K} \left[\prod_{i=1}^{n} \left(g_i^{k_i} \prod_{i<j} w_{i,j}^{k_i k_j} \right) \lambda^{|k|} \right].$$

mapping the tuple of constants to the polynomial. Analogously to the notation of energies and probability constants of microstates, the notation with lowercase letters is used to indicate that we are considering the image of Φ_1:

$$(g_1, ..., g_n, ..., w_{n-1,n}) := \Phi_1(G_1, ...G_n, ...W_{n-1,n}).$$

Moreover, $\mathbb{R}[\lambda]$ denotes the polynomial ring in the variable λ and coefficients in $\mathbb{R}$. Note that Φ_1 is well defined since

$$a \sim b \iff \Phi_1(a) \sim \Phi_1(b).$$

Φ_1 being a bijection, we can work with its image values –that is every molecule M is described by $(g_1^M, ..., g_n^M, ..., w_{n-1,n}^M)$– and concentrate on the investigation of Φ_2:
The map is well defined as for two representatives of the same equivalence class

$$a \sim b \implies \Phi_2(a) = \Phi_2(b).$$

Moreover, we can rewrite Φ_2 without using microstates $k \in K$ but with focus on the coefficients of $\Phi(M)$:

$$\Phi_2(g_1..., w_{n-1,n}) =$$
$$\prod_{i\in N}\left(g_i \prod_{i<j} w_{i,j}\right)\lambda^n + \sum_{p\in N}\left[\prod_{i\in N, i\neq p}\left(g_i \prod_{i<j, j\neq p} w_{i,j}\right)\right]\lambda^{n-1} +$$
$$\sum_{\substack{(p_1,p_2)\in N^2 \\ p_1\neq p_2}}\left[\prod_{p_1\neq i\neq p_2}\left(g_i \prod_{\substack{i<j \\ p_1\neq j\neq p_2}} w_{i,j}\right)\right]\lambda^{n-2} + ...$$
$$... + \sum_{i\in N} g_i\lambda + 1 \tag{3.9}$$

with $N := \{1, ...n\}$.

In this section, up to now, we defined a molecule as an equivalence class and investigated the map, which calculates the corresponding binding polynomial, superficially. A question which might arise at this point is why the binding polynomial is an important quantity. Why is the map to the binding polynomial considered, even though it just represents the sum of the weights of the microstates? An answer is given by Proposition 19 which states that there is a one to one correspondence between binding polynomials and overall titration curves.

Proposition 19. *Let M be a molecule with*

$$\Phi(M) = a_n\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + ... + a_1\lambda + 1.$$

Then its overall titration curve is given by

$$\Psi(\lambda) = \mathbb{E}_\lambda|k| = \frac{na_n\lambda^n + (n-1)a_{n-1}\lambda^{n-1} + (n-2)a_{n-2}\lambda^{n-2} + ... + a_1\lambda}{\Phi(M)}. \tag{3.10}$$

Moreover, for any overall titration curve $\mathbb{E}_\lambda|k|$, the corresponding binding polynomial is unique.

Proof. For the first part of the statement, a counting or combinatorial argument can be found in the paper by Martini and Ullmann (2013), on which this chapter is based. An even simpler argument is its stochastic definition: The coefficient a_i is the sum of all probability constants $g(k)$ with $|k| = i$. Consequently, the probability of $|\cdot|$ being equal to i is:

$$P_{M,\lambda}(|k| = i) = \frac{a_i \lambda^i}{\Phi(M)}.$$

The definition of the expectation gives Eq. (3.10).
For the second statement, let two titration curves be identical for every nonnegative value of λ:

$$\frac{n a_n \lambda^n + ... + a_1 \lambda}{a_n \lambda^n + ... + a_1 \lambda + 1} = \frac{n b_n \lambda^n + ... + b_1 \lambda}{b_n \lambda^n + ... + b_1 \lambda + 1} \tag{3.11}$$

which is equivalent to

$$\begin{aligned}(n a_n \lambda^n + (n-1) a_{n-1} \lambda^{n-1} + ... + a_1 \lambda)(b_n \lambda^n + b_{n-1} \lambda^{n-1} + ... + b_1 \lambda + 1) = \\ = (n b_n \lambda^n + (n-1) b_{n-1} \lambda^{n-1} + ... + b_1 \lambda)(a_n \lambda^n + a_{n-1} \lambda^{n-1} + ... + a_1 \lambda + 1).\end{aligned} \tag{3.12}$$

Each side of Eq. (3.12) represents a polynomial $\sum_{i=0}^{2n} c_i \lambda^i$ of degree $2n$. Since Eq. (3.12) is true for any nonnegative λ, the coefficients on both sites have to be equal:

$$c_k = \sum_{i=1}^{n} i a_i b_{k-i} = \sum_{i=1}^{n} i b_i a_{k-i} \tag{3.13}$$

where $a_i = b_i = 0$ for $i > n$ and $i < 0$. Moreover, $a_0 = b_0 = 1$. This gives

$$c_1 = 1 \cdot a_1 \cdot 1 = 1 \cdot b_1 \cdot 1.$$

The fact that $a_i = b_i$ for all $i \in \{0, ..., n\}$ is shown by induction.
For this, let us assume that we know $a_i = b_i$ for all $i \in \{0, ..., k\}$ we show that this implies $a_{k+1} = b_{k+1}$:

$$\begin{aligned}c_{k+1} = \sum_{i=1}^{n} i a_i b_{k+1-i} = \sum_{i=1}^{k+1} i a_i b_{k+1-i} = \\ = \sum_{i=1}^{n} i b_i a_{k+1-i} = \sum_{i=1}^{k+1} i b_i a_{k+1-i} = \\ = (k+1) a_{k+1} + \sum_{i=1}^{k} i a_i b_{k+1-i} = (k+1) b_{k+1} + \sum_{i=1}^{k} i b_i a_{k+1-i}.\end{aligned}$$

Since $b_i = a_i$ for all $i \in \{0, ..., k\}$ the second summand on both sites are equal and consequently $a_{k+1} = b_{k+1}$. □

Note here, that the second part of Proposition 19 states the following: By definition, the titration curve is the expected value of the distribution of the sum S as a function of λ. Due to the parametrization of the family of distributions of S, the whole distribution of S can be calculated for every activity λ from its expectations. In other words: The family of expectations determines the corresponding family of measures on the macrostates $\{0, ..., n\}$.

Remark 20. *From now on all molecules will be described by their image with respect to Φ_1, which will also be indicated by the use of lowercase letters.*

3.3 The Decoupled Sites Representation

At first, we will define what the term "decoupled sites" means and show equivalences to other definitions in Proposition 26.

Definition 21. *A molecule $L = (g_1, ..., g_n, w_{1,2}, ..., w_{n-1,n})$ is called* **decoupled** *or said to have* **decoupled sites** *if $w_{i,j} = 1 \; \forall i, j$.*

Onufriev et al. (2001) presented the main result of the Decoupled Sites Representation (DSR), saying that for every molecule $M = (g_1^M, ..., w_{n-1,n}^M)$ exactly one decoupled system $L = (g_1, ..., g_n, 1, .., 1)$ exists such that

$$\Phi_2(M) = \Phi_2(L).$$

However, it is possible that this might require the use of complex energies and probability constants with a nonzero imaginary part which is illustrated by Example 22.

Example 22. *Let M be a molecule with two interacting binding sites and*

$$(g_1^M, g_2^M, w_{1,2}^M) = (1, 1, 2).$$

Another molecule $L = (g_1, g_2, 1)$ with the same binding polynomial solves, according to Eq. (3.9), the system

$$g_1 g_2 = g_1^M g_2^M w_{1,2}^M = 2$$

$$g_1 + g_2 = g_1^M + g_2^M = 2.$$

Solving these equations gives the unique solution $L = (g_1, g_2, 1) = (1 + i, 1 - i, 1)$ which is equal to $(1 - i, 1 + i, 1)$ since we are dealing with equivalence classes.

This example shows that if the DSR shall be valid for all molecules, it is necessary to allow $g_i, w_{i,j} \in \mathbb{C} \setminus \{0\} =: \mathbb{C}^*$. This set would be appropriate as it guarantees the existence of a decoupled system with the same bp, and additional bps with coefficients in $\mathbb{C} \setminus \mathbb{R}$ do not have to be considered. However, complex numbers with imaginary part $iy \neq 0$ pose a problem for physical interpretation as it might not be regarded as binding "energy". Some ideas on possible interpretations of this phenomenon can be found in Chapter 6.

Having illustrated the motivation of a formal extension of the domain of energies, we have to adapt the definitions. In Eq. (3.8) the domain and image space of the map Φ_2 has to be changed to

$$\Phi_2 : \quad {\mathbb{C}^*}^m \big/ \sim \quad \longrightarrow \quad \mathbb{C}[\lambda] \tag{3.14}$$

where the definition of the equivalence relation "$\sim$" in Eq. (3.2) is not affected. However, one has to think about the appropriate domain of Φ_1. As $\exp : \mathbb{C} \longrightarrow \mathbb{C}^*$ is not bijective it is necessary to change Eq. (3.7) to

$$\Phi_1 : \quad D^m \big/ \sim \quad \longrightarrow \quad {\mathbb{C}^*}^m \big/ \sim \tag{3.15}$$

with $D := \{x + iy | x \in \mathbb{R}, y \in [-\pi, \pi[\,\} \subset \mathbb{C}$. Consequently, Eq. (3.3) rewrites

$$M \in {}^{D^m}\!/\!_{\sim}. \tag{3.16}$$

To simplify notation we will use

$$\mathbb{D} := {}^{D^m}\!/\!_{\sim} \qquad \text{and} \qquad \mathbb{H} := {}^{\mathbb{C}^{*^m}}\!/\!_{\sim}.$$

With this framework we can express the DSR as a proposition. If not mentioned explicitly, a molecule will always be understood as an element of $\mathbb{H}$, that is as an equivalence class on the set of binding and interaction constants.

Proposition 23 (The Decoupled Sites Representation)**.** *Let $M = (g_1^M, ..., g_n^M, ..., w_{n-1,n}^M)$ be a molecule. Then a unique molecule $L = (g_1, ..., g_n, 1, ..., 1)$ exists, such that*

$$\Phi_2(M) = \Phi_2(L).$$

Moreover, the binding constants of L are given by $g_i = -\frac{1}{r_i}$ with $(r_i)_{i=1,...,n}$ the roots of $\Phi_2(M)$.

The proof of Proposition 23 presented in Martini and Ullmann (2013) was based on the properties of Φ_2. The argumentation was the following: We define $O_n := \{(g_1, g_2, ..., g_n, 1, ..., 1) \in \mathbb{H}\}$, the set of all molecules without interaction, and show that Φ_2 is injective on O_n. Moreover, we show that

$$\Phi_2(O_n) = \Phi_2(\mathbb{H}) =: \mathrm{Im}(\Phi_2),$$

i.e. the image set of the restriction is still the same. This means that

$$\Phi_{2,O_n} : O_n \longrightarrow \mathrm{Im}(\Phi_2)$$

is also surjective to $\mathrm{Im}(\Phi_2)$, thus bijective and consequently an inverse map

$$\Phi_{2,O_n}^{-1} : \mathrm{Im}(\Phi_2) \longrightarrow O_n$$

exists. This gives

$$L = \Phi_{2,O_n}^{-1} \circ \Phi_2(M). \tag{3.17}$$

Bijectivity of Φ_{2,O_n} guarantees existence of the inverse map and thus uniqueness of L. For the details of this argumentation see Martini and Ullmann (2013).

However, in this work, some arguments shall be outsourced by using Vieta's formulas of which a proof is included in the argumentation described above.

Proof of Proposition 23. Let a polynomial $\Phi_2(M) = a_n\lambda^n + ... + a_1\lambda + 1$ be given. Eq. (3.9) shows how the coefficients are composed of the binding and interaction constants. In particular, if we look for a decoupled molecule $L = (g_1, ..., g_n, 1, ..., 1)$, its binding constants have to satisfy the following system of equations:

$$a_n = \prod_{i \in N} g_i$$

$$a_{n-1} = \sum_{k \in N} \prod_{\substack{i \in N \\ i \neq k}} g_i$$

$$\vdots$$

$$a_1 = \sum_{k \in N} g_k$$

Since the binding constants are the variables, this system of equations describes an affine algebraic variety. However, due to its very special structure, no methods from algebraic geometry are required. Vieta's formulas state that the solutions are given by $-\frac{1}{r_i}$ with r_i denoting the roots of $\Phi(M)$. Uniqueness is given, because of the introduced equivalence relation: All permutations of the binding constants with interaction constants equal to one describe the same molecule. □

The second statement of the DSR, which says that the binding constants of the decoupled system are given by the negative inverses of the roots of the binding polynomial, underlines once more that the domain of the binding constants has to be extended, if the DSR shall be valid for every molecule. Another important point is that uniqueness of the decoupled system for every binding polynomial, is not a result of the lack of interaction, but of the reduced number of variables, since the interaction constants were fixed to one. Thus, for instance, we can fix the interaction constants to any other value of $w_{i,j} = t \ \forall \ i, j$ and find a hypothetical molecule which has the predefined binding polynomial. This statement is a generalization of the DSR.

Proposition 24 (Generalized DSR)**.** *Let $M = (g_1^M, ..., g_n^M, ..., w_{n-1,n}^M)$ be a molecule. Then $\forall\, t \in \mathbb{C}^*$ a unique system $L = (g_1, ..., g_n, t, ..., t)$ exists such that*

$$\Phi_2(M) = \Phi_2(L).$$

Proof. Since we fixed all interaction constants to the same value, the proof of Proposition 23 can be transfered directly. Using Eq. (3.9) gives the following system which the binding constants of molecule L have to solve:

$$a_n = t^{\frac{n(n-1)}{2}} \prod_{i \in N} g_i$$

$$a_{n-1} = t^{\frac{(n-1)(n-2)}{2}} \sum_{k \in N} \prod_{\substack{i \in N \\ i \neq k}} g_i$$

$$\vdots$$

$$a_1 = \sum_{k \in N} g_k$$

The power of t can be brought to the left site, which gives new coefficients and the system of equations of Propostion 23. □

Remark 25. *The generalized DSR described in Proposition 24 is a generalization of the DSR to any value t of all interaction constants $w_{i,j}$. However, note that all interaction constants are fixed to the same value. This allowed to factor out a power of t in each equation and to use the proof of Propostion 23. A further generalization to any value of $(w_{1,2}, w_{1,3}, ...) = (t_1, t_2, ...)$ with $t_i \neq t_j$ is also possible, but we cannot use the proof of Propostion 23, since we cannot factor out a power of t. Since fixing $(w_{1,2}, w_{1,3}, ...) = (t_1, t_2, ...)$ leads to a system of n equations and n variables one might assume that the existence of a solution is obvious. However this is not clear since the system does not consist of (non-contradictory) linear equations. Thus, methods of algebraic geometry or computer algebra have to be used. We will present a proof for the case $n = 3$ later.*

3.4 Properties of decoupled molecules

Using the second part of Proposition 23, we can prove the following statements:

Proposition 26. *Let M be a molecule.*

a) Then the following statements are equivalent:

i) M is decoupled.

ii) The energy $G(k)$ of a microstate $k = (k_1, ..., k_n)$ is the sum of the energies of the corresponding states with only one site occupied:

$$G(k) = \sum_{i=1}^{n} k_i G_i.$$

iii) The probability constant $g(k)$ of microstate k is the product of the corresponding constants of the states with only one site occupied:

$$g(k) = \prod_{i=1}^{n} g_i^{k_i}.$$

b) Let $M = (g_1, ..., g_n, 1, ..., 1)$ be a decoupled molecule. Then each binding site of M has a Henderson-Hasselbalch titration curve described by

$$\Psi_i = \frac{g_i \lambda}{g_i \lambda + 1}.$$

c) If M is a molecule whose titration curves of the individual sites are all of Henderson-Hasselbalch shape, then a decoupled molecule M' exists, which has the same titration curve at each individual site. Moreover, M and M' share the same binding constants.

Proof. a) *i)* $\Leftrightarrow$ *ii)*: Definition of $G(k)$.
ii) $\Leftrightarrow$ *iii)*: Definition of $g(k)$, bijective map Φ_1.

b) Let the tuple $M = (g_1, ..., g_n, 1, ..., 1)$ be a decoupled system and a representative from its equivalence class. Moreover, let

$$M_{-i} := (g_1, ..., g_{i-1}, g_{i+1}, ..., g_n, 1, ..., 1)$$

denote the tuple with $n-1$ binding sites describing M as if site i was missing. Then,

$$P_{M,\lambda}(X_i = 1) = \mathbb{E}_\lambda X_i \overset{\text{Eq. (1.8)}}{=} \frac{E_i(M)}{\Phi(M)} \overset{\text{Eqs. (1.7,3.6)}}{=} \frac{g_i\lambda\Phi(M_{-i})}{\Phi(M)} \overset{\text{Prop. 23}}{=}$$

$$= \frac{g_i\lambda \prod\limits_{k\neq i}(\lambda + \frac{1}{g_k}) \prod\limits_{k\neq i} g_k}{\prod\limits_k(\lambda + \frac{1}{g_k}) \prod\limits_k g_k} = \frac{\lambda}{\lambda + \frac{1}{g_i}} = \frac{g_i\lambda}{g_i\lambda + 1}.$$

c) Let a_l denote the sum of the products of all l-element subsets of $\{g_1, ..., g_n\}$. The overall titration curve is the sum of all titration curves of individual sites:

$$\Psi(\lambda) = \sum_{i=1}^{n} \Psi_i \overset{\text{HH shape}}{=} \sum_{i=1}^{n} \frac{g_i\lambda}{g_i\lambda + 1} \tag{3.18}$$

Note here that it is not clear up to now, whether the g_i's are the binding constants of molecule M. They are given by the HH curves of the binding sites. Eq. (3.18) can be rewritten to

$$\frac{\sum\limits_{i=1}^{n}\left(g_i\lambda \prod\limits_{k\neq i}(g_k\lambda + 1)\right)}{\prod\limits_k(g_k\lambda + 1)} = \frac{n\prod\limits_k(g_k\lambda + 1) - \sum\limits_i \prod\limits_{k\neq i}(g_k\lambda + 1)}{\prod\limits_k(g_k + 1)} =$$

$$\frac{n(a_n\lambda^n + a_{n-1}\lambda^{n-1} + ... + 1) - a_{n-1}\lambda^{n-1} - 2a_{n-2}\lambda^{n-2} - ... - (n-1)a_1\lambda - n}{a_n\lambda^n + a_{n-1}\lambda^{n-1} + ... + 1} =$$

$$= \frac{na_n\lambda^n + (n-1)a_{n-1}\lambda^{n-1} + ... + a_1\lambda}{a_n\lambda^n + a_{n-1}\lambda^{n-1} + ... + 1}.$$

Since the overall titration curve is of the special form which is described in Proposition 19, the denominator is the binding polynomial. Starting from this polynomial, decoupling returns the initial binding constants. This means the molecule M shares its titration curves of each site with the decoupled molecule $M' = (g_1, ..., g_n, 1, ..., 1)$.

What remains to be shown is that the binding constants of M coincide with the g_i's. This is the case, since the definition of the titration curve of a certain binding site gives:

$$\frac{\sum\limits_{\{k\in K|k_i=1\}} g_M(k)\lambda^{|k|}}{\Phi_2(M)} = \frac{\sum\limits_{\{k\in K|k_i=1\}} g_{M'}(k)\lambda^{|k|}}{\Phi_2(M)}$$

and thus

$$\sum_{\{k\in K|k_i=1\}} g_M(k)\lambda^{|k|} = \sum_{\{k\in K|k_i=1\}} g_{M'}(k)\lambda^{|k|}.$$

Since the polynomials are identical if and only if every coefficient is identical, in particular, this implies that

$$g_i = g_i^M,$$

which is the first coefficient.

□

Remark 27. *The initial intention of formulating Proposition 26 was to show the equivalence of a molecule being decoupled and possessing HH curves at each binding site. However, it turned out to be difficult to show that if each binding site has a HH titration curve the interaction has to be trivial, in the case of more than three binding sites. This difficulty shall be illustrated in Example 28, and proofs for the equivalence of a molecule being decoupled and all sites exhibiting HH titration curves will be given for the cases of two and of three binding sites in Subsections 3.5 and 3.6.*

Example 28. *Let*

$$M = (g_1^M, g_2^M, g_3^M, g_4^M, w_{1,2}^M, w_{1,3}^M, w_{1,4}^M, w_{2,3}^M, w_{2,4}^M, w_{3,4}^M)$$

be a molecule with HH-curves at every binding site:

$$\Psi_i(\lambda) = \frac{g_i\lambda}{g_i\lambda + 1}.$$

Analogously, to the proof of Proposition 26 c), we sum up the titration curves, calculate the binding polynomial and see that $M' = (g_1, ..., g_n, 1, ..., 1)$ is decoupled and shares all titration curves of the individual sites with M. Analogously to the proof of Proposition 26 c), comparing coefficients of the numerator of the titration curves of the individual sites gives $g_i = g_i^M$. The remaining coefficients give conditions on the interaction constants. In the case of four binding sites, the second coefficients of the four titration curve give the following equations:

$$g_2w_{1,2} + g_3w_{1,3} + g_4w_{1,4} = g_2 + g_3 + g_4$$

$$g_1w_{1,2} + g_3w_{2,3} + g_4w_{2,4} = g_1 + g_3 + g_4$$

$$g_1w_{1,3} + g_2w_{2,3} + g_4w_{3,4} = g_1 + g_2 + g_4$$

$$g_1w_{1,4} + g_2w_{2,4} + g_3w_{3,4} = g_1 + g_2 + g_3$$

Moreover, the third coefficients give the equations

$$g_2g_3w_{1,2}w_{1,3}w_{2,3} + g_2g_4w_{1,2}w_{1,4}w_{2,4} + g_3g_4w_{1,3}w_{1,4}w_{3,4} = g_2g_3 + g_2g_4 + g_3g_4$$

$$g_1g_3w_{1,2}w_{1,3}w_{2,3} + g_1g_4w_{1,2}w_{1,4}w_{2,4} + g_3g_4w_{2,3}w_{2,4}w_{3,4} = g_1g_3 + g_1g_4 + g_3g_4$$

$$g_1g_2w_{1,2}w_{1,3}w_{2,3} + g_1g_4w_{1,3}w_{1,4}w_{3,4} + g_2g_4w_{2,3}w_{2,4}w_{3,4} = g_1g_2 + g_1g_4 + g_2g_4$$

$$g_1g_2w_{1,2}w_{1,4}w_{2,4} + g_1g_3w_{1,3}w_{1,4}w_{3,4} + g_2g_3w_{2,3}w_{2,4}w_{3,4} = g_1g_2 + g_1g_3 + g_2g_3$$

and the fourth coefficients state:

$$w_{1,2}w_{1,3}w_{1,4}w_{2,3}w_{2,4}w_{3,4} = 1.$$

To show that M itself is decoupled we would have to show that

$$(w_{1,2}, w_{1,3}, w_{1,4}, w_{2,3}, w_{2,4}, w_{3,4}) = (1, 1, 1, 1, 1, 1)$$

is the unique solution of this system for every choice of binding constants. However, this is not obviously the case, since the equations are polynomials in several variables.

3.5 Special considerations of the case n=2

At first, we will improve the statement of Proposition 26 b) and c) for the case of two binding sites, followed by an example for Proposition 24 for $n = 2$.

Proposition 29. *Let M be a molecule with two binding sites. Then M is decoupled if and only if both binding sites exhibit a HH titration curve.*

Proof. "⇒": see Proposition 26.
"⇐": As explained in Example 28, $w_{1,2}$ has to satisfy $g_1 g_2 = g_1 g_2 w_{1,2}$ and consequently $w_{1,2} = 1$ is the unique solution. □

Example 30. *Let M be a molecule with two interacting binding sites described by $M = (a, b, c)$. Then a system $(d, e, 4)$ exists with the same bp*

$$\Phi_2(a, b, c) = abc\lambda^2 + (a + b)\lambda + 1.$$

A system $(d, e, 4)$ with the same bp has to solve the equations:

$$4de = abc, \qquad d + e = a + b$$

Thus, d has to solve

$$d^2 - (a + b)d + \frac{abc}{4} = 0.$$

For example, the system $(a, b, c) = (1, 1, 2)$ shares its binding polynomial

$$P(\lambda) = 2\lambda^2 + 2\lambda + 1$$

with the system $(1 + \frac{1}{\sqrt{2}}, 1 - \frac{1}{\sqrt{2}}, 4)$. Its decoupled system is given by $(1 - i, 1 + i, 1)$.

Moreover, it is not only possible to fix the interaction constant $w_{1,2}$, but also a binding constant g_i to find a unique solution.

Proposition 31. *Let $M = (a, b, 1)$ be a system without intrinsic interaction. Then for every $d \in \mathbb{C}^* \setminus \{a + b\}$ a system $L = (d, a + b - d, \frac{ab}{d(a+b-d)})$ exists, which has the same binding polynomial. Moreover, all systems sharing the same binding polynomial are of this shape.*

Proof. Calculating the bp of L proves the first result. The argumentation to prove the second statement is the following: Let L be a molecule possessing the same bp, then $L = (g_1^L, g_2^L, w_{1,2}^L)$ has to solve the following equations:

$$g_1^L g_2^L w_{1,2}^L = ab$$

$$g_1^L + g_2^L = a + b.$$

Defining $d := g_1^L$ proves the second result. □

In Proposition 31, we showed that for a given binding polynomial

$$\Phi_2(a, b, 1) = a_2\lambda^2 + a_1\lambda + 1$$

and any binding constant $d \in \mathbb{C}^* \setminus \{a+b\}$, we can find a molecule $M = (d, e, w)$ possessing this bp. Note that it is not possible to find a system $L = (d, a+b-d, \frac{ab}{d(a+b-d)})$ if $d = a+b$, as this leads to a division by zero. Moreover, there is no $t \in \mathbb{C}$ such that

$$e^t = 0 = a + b - d.$$

The exception $d = a + b$ can be interpreted the following way: If both molecules share the same bp, the equation

$$a + b = d + e$$

is true. Thus, if d is changed, this equation will allow to adapt e, accordingly. If d comes close to $a+b$ and thus e close to zero, the distribution of the binding energies on the two binding sites of the corresponding molecule will become extremely asymmetric. A small absolute value of the binding constant e has to be compensated by a great interaction constant w. The exception $d = a + b$ corresponds to the limit case of $e = 0$ ($G_e = \infty$) and $w = \infty$ ($W = -\infty$). It can be interpreted physically as molecule with only one ligand binding site since an infinite amount of energy is required to bind the ligand to the second binding site.

3.6 Special considerations of the case n=3

Analogously to Proposition 29, we will improve Proposition 26 for the special case of $n = 3$:

Proposition 32. *Let M be a molecule with three binding sites. Then M is decoupled if and only if all three binding sites exhibit a HH titration curve.*

Proof. "⇒": see Proposition 26.
"⇐": As explained in Example 28, $(w_{1,2}, w_{1,3}, w_{2,3})$ has to satisfy

$$g_2 w_{1,2} + g_3 w_{1,3} = g_2 + g_3$$

$$g_1 w_{1,2} + g_3 w_{2,3} = g_1 + g_3$$

$$g_1 w_{1,3} + g_2 w_{2,3} = g_1 + g_2.$$

This linear system can be rewritten to

$$\begin{pmatrix} 0 & w_{1,2} & w_{1,3} \\ w_{1,2} & 0 & w_{2,3} \\ w_{1,3} & w_{2,3} & 0 \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix}$$

which gives

$$\underbrace{\begin{pmatrix} 0 & w_{1,2}-1 & w_{1,3}-1 \\ w_{1,2}-1 & 0 & w_{2,3}-1 \\ w_{1,3}-1 & w_{2,3}-1 & 0 \end{pmatrix}}_{=:A} \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{3.19}$$

Since $(g_1, g_2, g_3) \neq (0, 0, 0)$ Eq. (3.19) can only be true if

$$\det(A) = 2(w_{1,2}-1)(w_{1,3}-1)(w_{2,3}-1) = 0.$$

However, since $g_i \neq 0$, one of the factors of $\det(A)$ being equal to zero implies that the other interaction constants are equal to one, too. Thus, $(w_{1,2}, w_{1,3}, w_{2,3}) = (1, 1, 1)$ is the unique solution to the system, which also satisfies the remaining equation

$$w_{1,2}w_{1,3}w_{2,3} = 1.$$

□

The next proposition generalizes Proposition 24 even further for the case of three binding sites.

Proposition 33. *Let $M = (a, b, c, 1, 1, 1)$ be a molecule. Then for every $(t_1, t_2, t_3) \in {\mathbb{C}^*}^3$ a system $L = (g_1, g_2, g_3, t_1, t_2, t_3)$ exists such that*

$$\Phi_2(a, b, c, 1, 1, 1) = \Phi_2(g_1, g_2, g_3, t_1, t_2, t_3). \tag{3.20}$$

Proof. Eq. (3.20) means (g_1, g_2, g_3) is a solution to the system

$$\begin{aligned} f_1 &: \quad abc = t_1t_2t_3g_1g_2g_3 \\ f_2 &: \quad ab + ac + bc = t_1g_1g_2 + t_2g_1g_3 + t_3g_2g_3 \\ f_3 &: \quad a + b + c = g_1 + g_2 + g_3 \end{aligned}$$

We will show that a solution exists, independently of the choice of (a, b, c) and (t_1, t_2, t_3). To this end, we regard equations f_1, f_2, f_3 as polynomials in

$$\mathbb{C}[g_1, g_2, g_3, a, b, c, t_1, t_2, t_3].$$

We used the computational algebra system Magma to calculate the Gröbner basis (w.r.t. the lexicographic order: $g_1 > g_2 > g_3 > a > b > c > t_1 > t_2 > t_3$) of the corresponding ideal $\langle f_1, f_2, f_3 \rangle$. This Gröbner basis consists of 11 polynomials. The second elimination ideal is generated by the last polynomial which is of degree six in g_3 with constant term $a_0 \neq 0$ (as $a, b, c, t_1, t_2, t_3 \neq 0$). This means that for any choice of $a, b, c, t_i \in \mathbb{C}^*$ we will find six solutions of g_3 (with multiplicity). The Extension Theorem (Cox et al., 2008, p. 165) tells us that those partial solutions can be extended to solutions to the first elimination ideal if the leading coefficients of the generators (regarded as polynomials in g_2) of the first elimination ideal do not all vanish at the partial solution. Looking at the second polynomial, we see that the leading coefficient is t_1. This means, in the present situation of $t_1 \neq 0$, all solutions can be extended. Regarding the first polynomial of the Gröbner basis shows that the leading coefficient is $1 \neq 0$. This means those solutions can be extended further to full solutions of the whole system. Obviously, this leads to six simultaneous solutions of equations f_1, f_2, f_3 (with multiplicity). For additional information about the Gröbner basis, the Extension and the Elimination Theorem see Cox et al. (2005, 2008). □

We will give an example.

Example 34. *Let $M = (2, -\frac{1}{2}, 4, 1, 1, 1)$ be a system without interaction. We look for molecules $L = (g_1, g_2, g_3, 2, 4, 3)$ with the same bp and use the computer algebra system Maxima to calculate the solutions of the system*

$$-\frac{1}{6} = g_1g_2g_3$$

$$5 = 2g_1g_2 + 4g_1g_3 + 3g_2g_3$$
$$\frac{11}{2} = g_1 + g_2 + g_3.$$

The six solutions for (g_1, g_2, g_3) *are approximately given by the set*

$$\{(5.313833028641072, -0.10698947820485, 0.29315667609982),$$
$$(4.928390901432182, 0.62565997888068, -0.054051184028884),$$
$$(0.60206626291414, 4.953815261044177, -0.055881066899869),$$
$$(0.31463862460357, -0.10021762785637, 5.285578747628083),$$
$$(-0.083397327599447, 5.199003322259136, 0.38439389576735),$$
$$(-0.075531559612602, 0.4287288758266, 5.146802325581396)\}.$$

All molecules described by the different solutions have the same binding polynomial and the same interaction energies but different binding energies. Note that the precision is necessary to see that the binding polynomials are identical.

A nice by-product of Proposition 23 is the following potential algorithm for the calculation of a molecule (d, e, f, t_1, t_2, t_3) with a given bp in a different way.

Proposition 35. *Let* $M = (a, b, c, 1, 1, 1)$ *be a molecule without interaction. Moreover, let* $L = (d, e, f, t_1, t_2, t_3)$ *be a molecule possessing the same bp. Then a permutation* $\sigma \in \mathcal{S}_3$ *exists such that* (d, e, f) *is a fixed point of the following potential algorithm:*

$$d_1 := 1, \qquad e_1 := 1, \qquad f_1 := 1$$
$$P_i(\lambda) := \frac{abc}{t_1t_2t_3}\lambda^3 + \left(\frac{ab+ac+bc}{t_1} + d_if_i\left(1-\frac{t_2}{t_1}\right) + e_if_i\left(1-\frac{t_3}{t_1}\right)\right)\lambda^2 + (a+b+c)\lambda + 1$$
$$(\lambda_{i,1}, \lambda_{i,2}, \lambda_{i,3}) := \textit{Roots of } P_i$$
$$(d_{i+1}, e_{i+1}, f_{i+1}) := \left(-\frac{1}{\lambda_{i,\sigma(1)}}, -\frac{1}{\lambda_{i,\sigma(2)}}, -\frac{1}{\lambda_{i,\sigma(3)}}\right).$$

Moreover, every fixed point (h, i, j) *of this potential algorithm satisfies*

$$\Phi_2(a, b, c, 1, 1, 1) = \Phi_2(h, i, j, t_1, t_2, t_3).$$

Proof. (d, e, f) is a fixed point for the right permutation σ if $\left(-\frac{1}{d}, -\frac{1}{e}, -\frac{1}{f}\right)$ are the roots of

$$P(\lambda) := \frac{abc}{t_1t_2t_3}\lambda^3 + \left(\frac{ab+ac+bc}{t_1} + df\left(1-\frac{t_2}{t_1}\right) + ef\left(1-\frac{t_3}{t_1}\right)\right)\lambda^2 + (a+b+c)\lambda + 1.$$

According to Proposition 23, $\left(-\frac{1}{d}, -\frac{1}{e}, -\frac{1}{f}\right)$ are the roots of the polynomial

$$Q(\lambda) := def\lambda^3 + (de + df + ef)\lambda^2 + (d + e + f)\lambda + 1.$$

To show that (d, e, f) is a fixed point, we have to show that $P(\lambda) = Q(\lambda)$ which means

$$\frac{abc}{t_1t_2t_3} = def \tag{3.21}$$

$$\left(\frac{ab+ac+bc}{t_1} + df\left(1-\frac{t_2}{t_1}\right) + ef\left(1-\frac{t_3}{t_1}\right)\right) = (de+df+ef) \tag{3.22}$$

$$(a+b+c) = (d+e+f). \tag{3.23}$$

Correctness of Eqs. $(3.21-3.23)$ is a result of

$$\Phi_2(a,b,c,1,1,1) = \Phi_2(d,e,f,t_1,t_2,t_3).$$

For the second statement, let (h,i,j) be a fixed point. Then $(-\frac{1}{h},-\frac{1}{i},-\frac{1}{j})$ are the roots of the corresponding polynomial. This means (h,i,j) fulfills Eqs. $(3.21-3.23)$ and consequently

$$\Phi_2(a,b,c,1,1,1) = \Phi_2(h,i,j,t_1,t_2,t_3).$$

□

Remark 36. *a) Two open questions concerning this potential algorithm are whether it always converges to a fixed point (attraction of a solution) and which of the six possible solutions will be found. The attraction of the solution is not obvious since a small perturbation of the coefficients may have a huge effect on the roots of the polynomial and thus the roots of a polynomial at a certain step might not be close to the roots of the polynomial of the step before.*

b) Moreover, up to now it is not clear whether this algorithm is useful, since solutions can also be calculated by solving the corresponding system of algebraic equations under the use of a computer algebra program such as Maxima or Magma. However, this algorithm exploits the special structure of the problem and it will be extended to other problems which can not be solved with standard procedures, in Chapter 4.

Example 37. *We have implemented this potential algorithm using the computer algebra system Maxima to calculate a solution of Example 34. Let $M = (2,-\frac{1}{2},4,1,1,1)$ be a system without interaction. We look for a system $L = (d,e,f,2,4,3)$ with the same binding polynomial. The algorithm described in Proposition 35 gives, for 1000 iteration steps:*

$$(d,e,f) = (4.92839118225807, 0.62565999882172, -0.054051181079786).$$

This represents the second solution of Example 34.

These results draw the following picture: For a given binding polynomial $P(\lambda)$ and interaction constants t_1,t_2,t_3 there are in general six different corresponding molecules possessing the bp $P(\lambda)$. If two interaction constants are equal, e.g. $t_1 = t_2$, the role of g_2 and g_3 in the system of equations given by the coefficients and the map Eq. (3.9) is identical. This means if $(g_1,g_2,g_3,w_1,w_2,w_3) = (d,e,f,t_1,t_1,t_3)$ is a solution, then (d,f,e,t_1,t_1,t_3) is a solution, too. Due to the equivalence relation Eq. (3.2) both solutions are equivalent which means that there are not more than three different solutions. In the case of $t_1 = t_2 = t_3$ all solutions coincide, resulting in the uniqueness of the solution which is the statement of the (generalized) DSR.

3.7 Decoupled sites in the model of Chapter 2

We will shortly highlight what decoupled sites, for every fixed chemical activity, mean for the presented model of the molecule's ligand binding dynamics in Chapter 2. Proposition 26 states that a molecule is decoupled if and only if the energy $G(k)$ satisfies

$$G(k) = k_1 \cdot G_1 + k_2 \cdot G_2 + \ldots + k_n \cdot G_n \tag{3.24}$$

for any microstate $k = (k_1, \ldots, k_n)$. Eq. (3.24) directly implies that the energy difference between two neighboring states, e.g.

$$(k_1, \ldots, k_{m-1}, 0, k_{m+1}, k_n) \quad \text{and} \quad (k_1, \ldots, k_{m-1}, 1, k_{m+1}, k_n),$$

only depends on the site m which has a different binding state. Due to the structure of the transition probabilities (Eqs. (2.5-2.7)) this means that the probability of changing the occupation state of a certain site does not depend on the state of the other binding sites. Thus, in the presented model, decoupled sites translate into a transition matrix $q_{i,l}$ that satisfies a certain kind of stationarity: $q_{i,l} = \tilde{q}_{i-l}$ for a distribution $\tilde{q}$ and any pair i, l. Note that, in a decoupled molecule, the rows of the transition matrix cannot be identical, i.e. the transition distribution depends on the state.

4 The Decoupled Sites Representation For Two Different Types of Ligands

The following chapter deals with the transfer of the DSR to molecules that bind two different types of ligands. It presents results of the papers Martini et al. (2013a) and Martini et al. (2013b) and the text is in main parts adopted verbatim from these publications, except for notational adaption to the previous chapters. Note that the premise that all coefficients of the binding polynomial are positive will be added in Proposition 55. This precondition was not used in the paper (Martini et al., 2013a). However, the proof presented in the paper, is not totally satisfying since the last argument that the appearing polynomial of degree two has at least one root which is nonzero is not obvious. The added premise is not a problem, since all "real" molecules will have this feature, and it may be required to prove the DSR for other cases, as well.

4.1 Motivation

In this chapter, the Decoupled Sites Representation will be transfered to molecules with two different types of ligands. In particular, this transfer is of interest since proteins involved in the transport of substances across membranes often bind several ligands. Examples are the active or passive co-transport of different ligands driven by the concentration gradients at the membrane. Moreover, the DSR may also be applied to receptors with different types of ligands. Transporter molecules as well as receptors have been objects of investigations in the last years (e.g. Becker et al., 2007; Gnacadja, 2011; Till et al., 2008).

From a mathematical point of view, the transfer of the DSR is not trivial since one has to deal with polynomials of the polynomial ring in two variables $\mathbb{C}[\lambda, \kappa]$, which means it is not straightforward to calculate the binding polynomials "roots" to find a decoupled system.

In the following, we will summarize the mathematical basics of ligand binding in equilibrium, for molecules with different types of ligands. We use a model in which every microstate energy is the sum of binding and pairwise interaction energies, analogously to the model presented in Chapter 3. Again, the main results of this chapter are equally true for the extended model with interaction terms of higher order.

4.2 Two different types of ligands and one binding polynomial

4.2.1 Molecules with binding sites for two different types of ligands

At first, we transfer the setup described in Section 3.2 to the situation of two different types of ligands. We assume that the ligands do not share binding sites, which means

that there are two disjunct sets of binding sites which can only be occupied by one type of ligand. Analogously to Eqs. (3.1-3.3), we receive the following framework: The equilibrium binding properties of a molecule M with n_1 binding sites for ligand L_1 (sites $1, 2, ..., n_1$) and n_2 binding sites for ligand L_2 (sites $A_1, A_2, ..., A_{n_2}$) are described by an $m := \frac{(n_1+n_2)(n_1+n_2+1)}{2}$-tuple

$$(g_1^M, ..., g_{n_1}^M, g_{A_1}^M, ..., g_{A_{n_2}}^M, w_{1,2}^M, ..., w_{1,A_{n_2}}^M, ..., w_{A_1,A_2}^M, ..., w_{A_{n_2-1},A_{n_2}}^M)$$

$$= M \in {}^{\mathbb{C}^{*m}}\!/_{\sim}. \tag{4.1}$$

Note here, that we directly start from the set of binding and interaction constants, that is from the image of Φ_1 and forget about its domain (Φ_1 is bijective). Since Φ_1 will not be considered, we will use the notation Φ for the map Φ_2 of Chapter 3. To describe a molecule with two types of ligands by Eq. (4.1) appropriately, the equivalence relation of Definition 17 has to be adapted, which is illustrated by Example 38.

Example 38. *Let S and M be two molecules with one binding site for ligand L_1 and one binding site for ligand L_2:*

$$M = \left(1, 2, \frac{1}{2}\right) \qquad S = \left(2, 1, \frac{1}{2}\right).$$

Using the equivalence relation of Definition 17 with $n = n_1 + n_2$ gives

$$M \sim S.$$

However, the role of the ligands which bind to S and M is permuted and, thus, the molecules should not be regarded as equal. For this reason we adapt Definition 17.

Definition 39. *Let $\mathbb{C}^{*m}$ with $m = \frac{(n_1+n_2)(n_1+n_2+1)}{2}$ be the set of all tuples describing molecules with n_1 binding sites for ligand L_1 and n_2 binding sites for ligand L_2. Moreover, let*

$$M = \left(g_1^M, ..., g_{n_1}^M, g_{A_1}^M, ..., g_{A_{n_2}}^M, w_{1,2}^M, ..., w_{1,A_{n_2}}^M, ..., w_{A_{n_2-1},A_{n_2}}^M\right)$$

and

$$N = \left(g_1^N, ..., g_{n_1}^N, g_{A_1}^N, ..., g_{A_{n_2}}^N, w_{1,2}^N, ..., w_{1,A_{n_2}}^N, ..., w_{A_{n_2-1},A_{n_2}}^N\right).$$

Then M is equivalent to N (Notation: $M \sim N$) if and only if two permutations σ_1 of $(1, ..., n_1)$ and σ_2 of $(1, ..., n_2)$ exist such that

$$M = \left(g_{\sigma_1(1)}^N, ..., g_{A_{\sigma_2(n_2)}}^N, w_{\sigma_1(1),\sigma_1(2)}^N, ..., w_{\sigma_1(1),A_{\sigma_2(n_2)}}^N, ..., w_{A_{\sigma_2(n_2-1)},A_{\sigma_2(n_2)}}^N\right). \tag{4.2}$$

To simplify notation we will henceforth write g_i for g_i^M if it is clear to which molecule the binding constant belongs to. Moreover, we use the microstate notation and

$${}^{\mathbb{C}^{*m}}\!/_{\sim} =: \mathbb{G}_{n_1,n_2}$$

with " $\sim$ " of Definition 39.

4.2.2 A binding polynomial in two variables

Analogously to Eq. (3.4), we define the binding polynomial in the ligand activities λ and κ of a molecule M, with n_1 binding sites for ligand L_1 and n_2 binding sites for ligand L_2 by

$$\Phi(M) = \sum_{k \in K} g(k) \lambda^{S_1(k)} \kappa^{S_2(k)} \tag{4.3}$$

with $S_1(k) := \sum_{i=1}^{n_1} k_i$ and $S_2(k) := \sum_{i=1}^{n_2} k_{A_i}$ denoting the number of bound ligands of both types. Moreover, as in the previous chapters, $g(k)$ denotes the microstate constant of state k. Analogously to Eqs. (1.6-1.8), the average amount of bound ligand to site i in equilibrium is given by

$$\Psi_i = \frac{\sum\limits_{\{k \in K | k_i = 1\}} g(k) \lambda^{S_1(k)} \kappa^{S_2(k)}}{\sum\limits_{k \in K} g(k) \lambda^{S_1(k)} \kappa^{S_2(k)}} =: \frac{E_i(M)}{\Phi(M)}. \tag{4.4}$$

Eq. (4.4) leads to the following overall titration curves for ligands L_1 and L_2.

$$\Psi_{L_1} = \frac{\sum\limits_{i=1}^{n_1} E_i(M)}{\Phi(M)} \tag{4.5}$$

$$\Psi_{L_2} = \frac{\sum\limits_{i=A_1}^{A_{n_2}} E_i(M)}{\Phi(M)} \tag{4.6}$$

4.3 On decoupling molecules with two types of ligands

To get an idea, and to point out some problems with the transfer of the DSR, we give two simple examples of hypothetical molecules whose bp can be calculated easily.

Example 40. *Let $M = (g_1^M, g_A^M, w_{1,A}) = (\frac{1}{2}, 2, \frac{1}{3})$ be a molecule with one binding site for each type of ligand. Then:*

$$\Phi(M) = \frac{1}{3}\lambda\kappa + \frac{1}{2}\lambda + 2\kappa + 1$$

$$E_1(M) = \frac{1}{3}\lambda\kappa + \frac{1}{2}\lambda$$

$$E_A(M) = \frac{1}{3}\lambda\kappa + 2\kappa.$$

Moreover, we see here that it is not possible to decouple this system as the map

$$\Phi : (g_1, g_A, w_{1,A}) \mapsto (g_1 g_A w_{1,A}, g_1, g_A)$$

which gives the coefficients of the polynomial, is injective. Thus, it is impossible to find a molecule $(g_1, g_A, 1)$ with the same binding polynomial.

Example 41. *Let* $M = (g_1^M, g_2^M, g_A^M, w_{1,2}^M, w_{1,A}^M, w_{2,A}^M) = (\frac{1}{2}, 2, 3, \frac{1}{2}, 2, \frac{1}{3})$ *be a molecule with two binding sites for ligand* L_1 *and one for ligand* L_2*. Then,*

$$\Phi(M) = \lambda^2\kappa + \frac{1}{2}\lambda^2 + 5\lambda\kappa + \frac{5}{2}\lambda + 3\kappa + 1.$$

In this situation of two and one bindings sites Φ *is a map*

$$\Phi : \mathbb{G}_{2,1} \longrightarrow \mathbb{C}^{*5}.$$

Here, the image space $\mathbb{C}^{*5}$ *represents the polynomials in two variables with* 6 *coefficients including the constant term which equals 1. Thus, it should not be injective and decoupling might be possible. However, intuitively, it is clear, that not all interaction energies can be trivial, as this would reduce the domain to* $\mathbb{C}^{*3}/\sim$*. Looking for another molecule with the same binding polynomial means searching for a solution* $(g_1, g_2, g_A, w_{1,2}, w_{1,A}, w_{2,A})$ *to the system*

$$\begin{aligned} 1 &= g_1 g_2 g_A w_{1,2} w_{1,A} w_{2,A} \\ \frac{1}{2} &= g_1 g_2 w_{1,2} \\ 5 &= g_1 g_A w_{1,A} + g_2 g_A w_{2,A} \\ \frac{5}{2} &= g_1 + g_2 \\ 3 &= g_A. \end{aligned} \tag{4.7}$$

As g_A *is fixed, the solutions to system (4.7), that is all molecules with bp* $\Phi(M)$*, form an algebraic variety* $V \subset \mathbb{C}^{*5}$ *defined by four polynomials. This means, under certain conditions on the polynomials,* $dim(V) = 1$*. The systems without interaction between the binding sites for the same type of ligand are given by*

$$V \cap \{(g_1, g_2, w_{1,2}, w_{1,A}, w_{2,A}) \in \mathbb{C}^{*5} | w_{1,2} = 1\}.$$

We used the computer algebra program Maxima to calculate the solutions. In this special situation we receive the following tuples sharing the same binding polynomial but with non-interacting sites for ligand L_1*:*

$$(2.2807762, 0.2192236, 3, 1, 0.6288467, 1.0601419),$$

$$(0.2192236, 2.2807762, 3, 1, 1.0601419, 0.6288467),$$

$$(2.2807762, 0.2192236, 3, 1, 0.1018987, 6.5424460),$$

$$(0.2192236, 2.2807762, 3, 1, 6.5424460, 0.1018987).$$

Note that, as we are dealing with equivalence classes, the first and the second pair of the solutions coincide. Calculating the same with $w_{1,A} = 1$ *or* $w_{2,A} = 1$ *does not give any solution. Remarkably, fixing* $g_1 = 1$ *or* $g_2 = 1$ *is solvable, however we will not investigate this phenomenon further, as we are interested in decoupling the system, which means setting interaction constants to 1.*

Example 41 leads to the conjecture that it is possible to decouple the binding sites for the same type of ligand. However, it is not generally possible to decouple different binding sites for different types of ligands which was illustrated by Example 40. For this reason we call a system decoupled if its binding sites for the same type of ligand do not interact directly:

Definition 42. *A molecule* $N = (g_1, ..., g_{n_1}, g_{A_1}, ..., g_{A_{n_2}}, w_{1,2}, ..., w_{A_{n_2-1},A_{n_2}})$ *with binding sites for two different ligands is called* **decoupled** *if*

$$w_{i,j} = 1 \qquad \forall\{i,j\} \subset \{1,2,...,n_1\},\ \forall\{i,j\} \subset \{A_1, A_2, ..., A_{n_2}\}$$

A very important point which is illustrated in Example 41 is the loss of uniqueness of the decoupled system which is given in the case of one type of ligand (Proposition 23). We formulate the DSR for two types of ligands the following way:

Conjecture 43. *Let*

$$M = (g_1^M, ..., g_{n_1}^M, g_{A_1}^M, ..., g_{A_{n_2}}^M, w_{1,2}^M, ..., w_{A_{n_2-1},A_{n_2}}^M)$$

be a molecule with n_1 *binding sites for ligand type* L_1 *and* n_2 *binding sites for ligand type* L_2*. Then at least one decoupled molecule* N *exists, such that*

$$\Phi(M) = \Phi(N).$$

Since we did not find a general proof for Conjecture 43 we will investigate the case $n_2 = 1$ first. The problem with proving this conjecture generally, is the following: One could use Hilbert's weak Nullstellensatz and show that the ideal generated by the polynomials (analogously to Eqs. (4.7)) does not contain unity. Then the existence of a solution would be guaranteed. However, to use this approach, one has to calculate the ideal generated by the polynomials (e.g. the corresponding Gröbner basis) without writing down the polynomials explicitly as n_1 and n_2 are not fixed. Another similar argumentation –with the same problem– would be to calculate a Gröbner basis to find partial solutions in an elimination ideal and to extend theses solutions to full solutions of the system under the use of the Extension Theorem. This approach has already been used to prove Proposition 33. Another approach would be the use of a higher-dimensional analog of the Bezout-Theorem. Yet, this would only give a statement for varieties in projective space. The most promising idea might be to use the special structure of the polynomials to give a proof constructively by reducing the problem to the proof of the DSR for one type of ligand (Proposition 23). We will compare the approach of calculating the Gröbner basis and using the Extension and Elimination Theorems to the concept of exploiting the special structure of the algebraic systems to prove Conjecture 43 for $n_2 = 1$ in the next section. Moreover, we will investigate which unique properties all decoupled molecules share. The fact that a decoupled system is not unique was illustrated by Example 41. However, Proposition 44 shows that at least the binding constants are unique (except for permutations):

Proposition 44. *Let*

$$\Phi(M) = a_{n_1,n_2}\lambda^{n_1}\kappa^{n_2} + a_{n_1-1,n_2}\lambda^{n_1-1}\kappa^{n_2} + ...a_{0,n_2}\kappa^{n_2} + a_{n_1,n_2-1}\lambda^{n_1}\kappa^{n_2-1}... + 1$$

be a bp of a molecule with n_1 binding sites for ligand L_1 and n_2 binding sites for ligand L_2. Let

$$N = (g_1^N, ..., g_{n_1}^N, g_{A_1}^N, .., g_{A_{n_2}}^N, 1, ..., w_{n_1,A_{n_2}}^N, 1..., 1)$$

and

$$K = (g_1^K, ..., g_{n_1}^K, g_{A_1}^K, .., g_{A_{n_2}}^K, 1, ..., w_{n_1,A_{n_2}}^K, 1..., 1)$$

be two different corresponding decoupled systems. Then two permutations σ_1 of $\{0, ..., n_1\}$ and σ_2 of $\{0, ..., n_2\}$ exist, such that

$$(g_i^N)_{i=1}^{n_1} = (g_{\sigma_1(i)}^K)_{i=1}^{n_1}$$

and

$$(g_{A_i}^N)_{i=1}^{n_2} = (g_{A_{\sigma_2(i)}}^K)_{i=1}^{n_2}.$$

Proof. A decoupled system is in the preimage of $\Phi(M)$ with respect to the map $M \mapsto \Phi(M)$. In particular, it has to solve the subsystem of equations given by its coefficients $\{a_{i,0}\}_{i=1}^{n_1}$. As this subsystem is free of the binding constant variables $\{g_{A_i}\}_{i=1}^{n_2}$ of ligand L_2 it represents the case of the DSR for one type of ligand. Consequently, according to Proposition 23, the set $\{g_i\}_{i=1}^{n_1}$ can be calculated from the roots of

$$a_{n_1,0}\lambda^{n_1} + a_{n_1-1,0}\lambda^{n_1-1} + ... + 1,$$

which shows $(g_i^N)_{i=1}^{n_1} = (g_{\sigma_1(i)}^K)_{i=1}^{n_1}$. The same is true for the subsystem of equations given by $\{a_{0,i}\}_{i=1}^{n_2}$ which gives the second result. □

4.4 Molecules with n to one binding sites

4.4.1 The Decoupled Sites Representation

At first we will prove Conjecture 43 for the case $(n_1, n_2) = (2, 1)$ to compare the approach of calculating the Gröbner basis and using the Elimination and Extension Theorems to the exploitation of the special structure of the varieties we are dealing with. Note that Proposition 46 includes the statement of Proposition 45. However, we will illustrate the different approaches for proving Conjecture 43 on the basis of the proof of Proposition 45.

Proposition 45. *Let*

$$M = (g_1^M, g_2^M, g_A^M, w_{1,2}^M, w_{1,A}^M, w_{2,A}^M)$$

be a molecule with two binding sites for ligand L_1 and one binding site for ligand L_2. Then a molecule

$$N = (g_1, g_2, g_A, 1, w_{1,A}, w_{2,A})$$

exists such that

$$g_A = g_A^M \quad \textit{and} \quad \Phi(M) = \Phi(N).$$

Proof. Let

$$\Phi(M) = a_{2,1}\lambda^2\kappa + a_{2,0}\lambda^2 + a_{1,1}\lambda\kappa + a_{1,0}\lambda + a_{0,1}\kappa + 1$$

be the binding polynomial of molecule M. A molecule $N = (g_1, g_2, g_A, 1, w_{1,A}, w_{2,A})$ is a solution to the algebraic system

$$\begin{aligned} g_1^M g_2^M g_A^M w_{1,2}^M w_{1,A}^M w_{2,A}^M &= a_{2,1} = & g_1 g_2 g_A w_{1,A} w_{2,A} \\ g_1^M g_2^M w_{1,2}^M &= a_{2,0} = & g_1 g_2 \\ g_1^M g_A^M w_{1,A}^M + g_2^M g_A^M w_{2,A}^M &= a_{1,1} = & g_1 g_A w_{1,A} + g_2 g_A w_{2,A} \\ g_1^M + g_2^M &= a_{1,0} = & g_1 + g_2 \\ g_A^M &= a_{0,1} = & g_A \end{aligned} \tag{4.8}$$

We regard these equations as polynomials in

$$\mathbb{C}[g_1, g_2, g_A, w_{1,A}, w_{2,A}, a_{2,1}, a_{2,0}, a_{1,1}, a_{1,0}, a_{0,1}]$$

and use the computational algebra system Magma to calculate the Gröbner basis GB of the corresponding ideal (w.r.t. the lexicographic order $g_1 > g_2 > g_A > w_{1,A} > ... > a_{0,1}$, see Cox et al. (2008)). Note that in this situation it is not enough to see that (the reduced) Gröbner basis is unequal to $\{1\}$, since this only implies that a solution to the system exists. However this does not show that a solution exists for any choice of $(g_1^M, g_2^M, g_A^M, w_{1,2}^M, w_{1,A}^M, w_{2,A}^M) \in \mathbb{C}^{*6}$. For the following argumentation it is important to see that $a_{2,1}, a_{2,0}, a_{0,1}$ are nonzero. The last of 17 polynomials of the Gröbner basis is

$$\begin{aligned} P_{17} = w_{2,A}^4 a_{2,0}^2 a_{0,1}^2 - w_{2,A}^3 a_{2,0} a_{1,1} a_{1,0} a_{0,1} - 2 w_{2,A}^2 a_{2,1} a_{2,0} a_{0,1} + \\ + w_{2,A}^2 a_{2,1} a_{1,0}^2 a_{0,1} + w_{2,A}^2 a_{2,0} a_{1,1}^2 - w_{2,A} a_{2,1} a_{1,1} a_{1,0} + a_{2,1}^2 \end{aligned}$$

and it defines the fourth elimination ideal (Elimination Theorem). Since its leading coefficient is nonzero, we will find four solutions for $w_{2,A}$. Moreover, since its constant term is nonzero, $w_{2,A} = 0$ is not a solution. As the leading coefficient of the 16th polynomial (regarded as a polynomial in $w_{1,A}$)

$$\begin{aligned} P_{16} = w_{1,A} a_{2,1} a_{2,0} a_{0,1} + w_{2,A}^A a_{2,0}^2 a_{0,1}^2 - w_{2,A}^2 a_{2,0} a_{1,1} a_{1,0} a_{0,1} - \\ - 2 w_{2,A} a_{2,1} a_{2,0} a_{0,1} + w_{2,A} a_{2,1} a_{1,0}^2 a_{0,1} + w_{2,A} a_{2,0} a_{1,1}^2 - a_{2,1} a_{1,1} a_{1,0} \end{aligned}$$

is $a_{2,1} a_{2,0} a_{0,1}$ and will not vanish in the solutions of $P_{17} = 0$, all solutions of $w_{2,A}$ can be extended to solutions of the third elimination ideal (Extension Theorem). To extend the solutions to the second elimination ideal, consider the tenth polynomial with leading coefficient 1. Moreover, to see that the partial solutions can be extended further, consider the sixth and the first polynomial with leading coefficients $w_{2,A}^2 a_{2,0} a_{0,1}$ and 1, respectively. Thus, this procedure leads to four solutions of the full system (only two of them are different w.r.t. the equivalence relation of Definition 39). For more information on the Elimination and Extension Theorems see Cox et al. (2005, 2008). □

Note that the proof of Proposition 45 also showed that

$$\Phi(\mathbb{G}_{2,1}) \supset \{a_{2,1}\lambda^2\kappa + a_{2,0}\lambda^2 + a_{1,1}\lambda\kappa + a_{1,0}\lambda + a_{0,1}\kappa + 1 | a_{i,j} \in \mathbb{C}^*\},$$

which means all polynomials of this shape have a preimage w.r.t. Φ. To calculate the energies of the decoupled molecule N in Proposition 45 one can use a computational algebra system (Magma, Maxima) to solve Eqs. (4.8) or use some special properties of this system: In general, not only for this choice of n_i, the coefficients $(a_{i,0})_{i=1,...,n_1}$ define a system of algebraic equations which allows to calculate $(g_1, ..., g_{n_1})$ (Proposition 23). Analogously, $(a_{0,i})_{i=1,...,n_2}$ give $(g_A, ..., g_{A_{n_2}})$. In system (4.8), with the same argument, $w_{1,A}$ and $w_{2,A}$ are given by

$$w_{1,A} = -\frac{1}{g_1 \lambda_{z_i}}$$

$$w_{2,A} = -\frac{1}{g_2 \lambda_{z_j}}$$

where λ_{z_i} are the roots of

$$\frac{a_{2,1}}{g_A}\lambda^2 + \frac{a_{1,1}}{g_A}\lambda + 1.$$

This calculation can also be used to prove Proposition 45 and will be used in the following to prove the more general case of $(n_1, 1)$ binding sites.

Proposition 46. *Let M be a molecule with n_1 binding sites for Ligand L_1 and one binding site for ligand L_2 (which is denoted as site A). Then a decoupled molecule $N = (g_1, ..., g_{n_1}, g_A, 1, ..., w_{n_1,A})$ exists with*

$$\Phi(M) = \Phi(N).$$

Proof. Let

$$\Phi(M) = a_{n_1,1}\lambda^{n_1}\kappa + a_{n_1,0}\lambda^{n_1} + ... + a_{1,1}\lambda\kappa + a_{1,0}\lambda + a_{0,1}\kappa + 1$$

be the binding polynomial of M. Again, a decoupled molecule N is a point of the algebraic variety V in the variables $g_1, ..., g_A, w_{1,A}, ..., w_{n_1,A}$ defined by the coefficients $a_{i,j}$ and the corresponding equations. The equations of coefficients $a_{n_1,0}, ..., a_{1,0}$ are free from the variables $w_{i,j}$ since we are looking for a decoupled system. Thus, Proposition 23 gives

$$(g_1, ..., g_{n_1}) = \left(-\frac{1}{\lambda_{z_1}}, ..., -\frac{1}{\lambda_{z_{n_1}}}\right)$$

with λ_{z_i} the roots of

$$a_{n_1,0}\lambda^{n_1} + a_{n_1-1,0}\lambda^{n_1-1} + ... + a_{1,0}\lambda + 1. \tag{4.9}$$

Moreover, $a_{0,1}$ gives g_A. The remaining equations can be rewritten

$$\frac{a_{n_1,1}}{g_A} = \prod_{i=1}^{n_1} g_i w_{i,A}$$

$$\frac{a_{n_1-1,1}}{g_A} = \sum_{j=1}^{n_1} \prod_{i=1, i\neq j}^{n_1} g_i w_{i,A}$$

$$\vdots$$

$$\frac{a_{1,1}}{g_A} = \sum_{i=1}^{n_1} g_i w_{i,A}.$$

Thus, the products $g_i w_{i,A}$ are determined by the roots of

$$\frac{a_{n_1,1}}{g_A}\lambda^{n_1} + \frac{a_{n_1-1,1}}{g_A}\lambda^{n_1-1} + ... + \frac{a_{1,1}}{g_A}\lambda + 1. \tag{4.10}$$

Consequently, the interaction energies $w_{i,A}$ can be calculated as the binding energies g_i are already known. □

4.4.2 The maximal number of decoupled molecules and properties they share

The proof of Proposition 46 also shows how many different decoupled molecules exist at most.

Corollary 47. *Let M be a molecule with n_1 binding sites for Ligand L_1 and one binding site for ligand L_2. Then at most $n_1!$ different decoupled molecules exist.*

Proof. The proof of Proposition 46 shows that at most $(n_1!)^2$ tuples exist which correspond to the different permutations of the roots of polynomial (4.9) and polynomial (4.10) and which solve the system. However, $n_1!$ tuples represent the same molecule. Thus, the maximal number of different decoupled molecules is $n_1!$. □

Example 48. *We illustrate the binding curves of individual sites of a system with two binding sites for ligand L_1 (activity λ) and one binding site for ligand L_2 (activity κ)and its corresponding decoupled systems. We used other hypothetical binding and interaction constants than in Example 41 to observe titration curves which can be distinguished by eye. To this end, let*

$$M = (g_1, g_2, g_A, w_{1,2}, w_{1,A}, w_{2,A}) = (900, 900, 300, 10^{-4}, 1000, 2000)$$

be a molecule. Its decoupled molecules are given by

$$N = (1799.955, 0.04500113, 300, 1, 1500.004, 1333.33)$$

$$K = (1799.955, 0.04500113, 300, 1, 0.03333491, 59997167).$$

The titration curves of the individual sites of the molecules are illustrated in Fig. 4.1.

An interesting observation is the fact that in the titration curves of the individual sites of the decoupled molecules of Example 48, the area of transition between 0.1 and 0.9 probability of occupation is comparatively small. However, it is difficult to quantify this characteristic. Moreover, regarding the titration curves of the decoupled molecules, it seems that the titration curve of site 1 of molecule K is a composition of the left part of site 1 and the right part of site 2 of molecule N. Analogously, the curve of site 2 of molecule K seems to be composed of the remaining parts of the curves of sites $1, 2$ of molecule N. This observation leads to the conjecture, that there is a unique set of "bricks" all decoupled molecules are built of. We will investigate this in the following. As the titration curves of individual sites are sums of the probabilities of the microstates in which the individual site is occupied, the constants of the microstates,

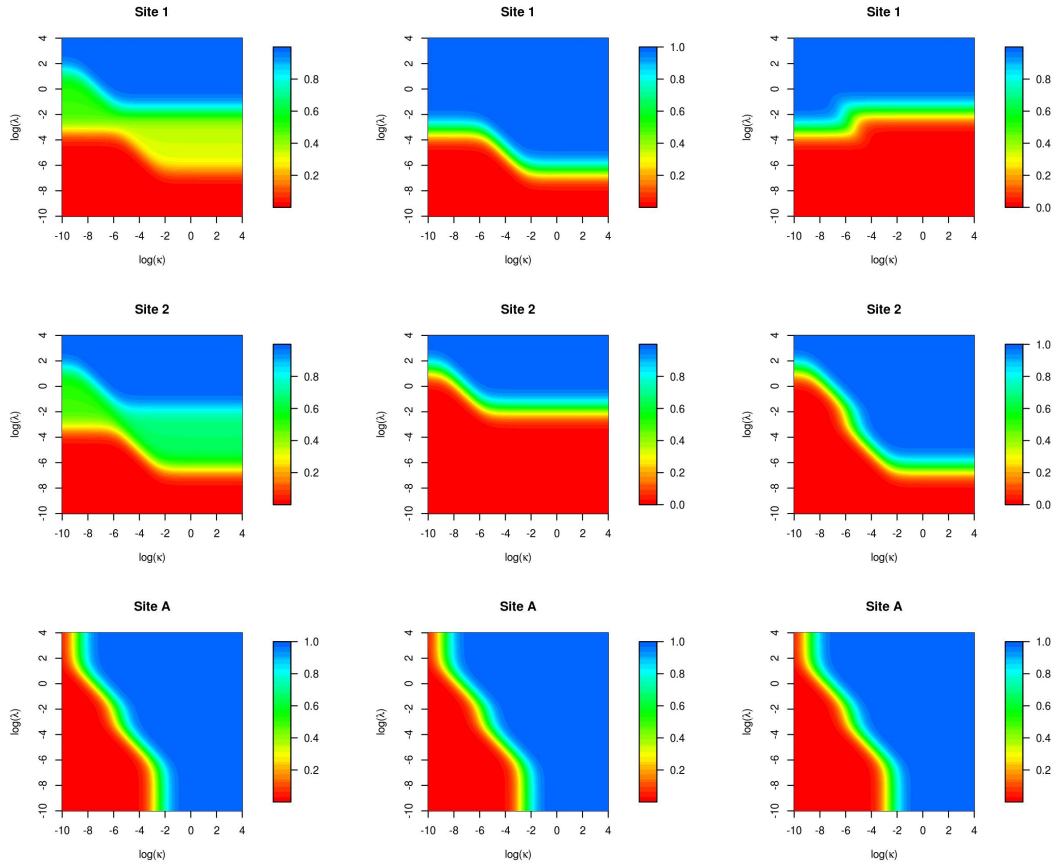

Figure 4.1: Activity dependent ligand binding to each site of the original molecule M (left column) and of the decoupled molecules N (middle column) and K (right column) of Example 48. A logarithmic scale of the ligand activities λ and κ is used. The probability of occupation is encoded by the colors, according to the colorbars.

Table 4.1: Constants of all microstates of the different molecules of Example 48: $M = (900, 900, 300, 10^{-4}, 1000, 2000)$ and the corresponding decoupled molecules N, K. The binding sites 1 and 2 for the first type of ligand are described by the first and the second entry of the microstate. The third entry of the microstate corresponds to site A, the binding site for the second ligand.

Microstate	M	N	K
$(0,0,0)$	1	1	1
$(0,0,1)$	300	300	300
$(0,1,0)$	900	0.04500113	0.04500113
$(1,0,0)$	900	1799.955	1799.955
$(1,1,0)$	81	81	81
$(0,1,1)$	$5.4 \cdot 10^8$	18000.4	$8.09982 \cdot 10^8$
$(1,0,1)$	$2.7 \cdot 10^8$	$8.09982 \cdot 10^8$	18000.4
$(1,1,1)$	$4.86 \cdot 10^{10}$	$4.86 \cdot 10^{10}$	$4.86 \cdot 10^{10}$

which are listed in Table 4.1 can give information about this observation. We see here, that in the decoupled systems of Example 48, the probabilities of the two events in which one ligand of type L_1 and the the second ligand are bound, are permuted. These probabilities are the unique "bricks" all decoupled molecules are built of. Before we formulate this as a proposition, we define the term macrostate for molecules with two types of ligands:

Definition 49. *Let M be a molecule with (n_1, n_2) binding sites. It is said to be in macrostate (i, j), $i \leq n_1, j \leq n_2$ if –in its current microstate– exactly i ligands of type L_1 and exactly j ligands of type L_2 are bound.*

Proposition 50. *Let M be a molecule with n_1 binding sites for ligand L_1 and one binding site for ligand L_2. Moreover, let the order of the sites in the decoupled molecules be fixed to the same permutation. Then the following statements hold:*

a) For every microstate k with unoccupied site A, all decoupled molecules share the same microstate constant $g(k)$.

b) For every macrostate $(i, 1)$ with occupied site A and i occupied sites for ligand L_1, $\binom{n_1}{i}$ numbers exist such that for any decoupled molecule the tuple of its constants of microstates belonging to this macrostate is a permutation of these numbers.

c) The permutation of microstate constants of macrostate $(1, 1)$ fixes the permutations of the microstate constants of all other macrostates $(i, 1)$.

d) Every decoupled molecule can be identified one to one with a permutation of the microstate constants of macrostate $(1, 1)$.

Proof. a) Let k be a microstate in which site A is unoccupied. Then its constant is the product of the binding constants of the sites which are occupied. According to Proposition 44 all decoupled molecules share the same binding constants which

gives the first statement since the permutation of $\{g_1, ..., g_{n_1}\}$ was assumed fixed, previously.

b) Let k be a microstate in which site A is occupied. Eq. (1.5), together with the assumption of only pairwise interaction, states that its constant is the product of the binding constants of all occupied sites and their interaction constants. As the interaction constants of any pair of binding sites for the same type of ligand are 1, this reduces to

$$g(k) = g_A \prod_{i=1}^{n_1} g_i^{k_i} w_{i,A}^{k_i}. \tag{4.11}$$

Since the decoupled systems share the binding constant g_A and since the products $(g_i w_{i,A})_{i=1,...,n_1}$ correspond to the permutations of the roots of polynomial (4.10), the microstate constants of different decoupled systems belonging to the same macrostate are permutations.

c) Let a permutation of the microstate constants belonging to macrostate $(1,1)$ be chosen. Then all interaction constants are determined as the microstate constants are given by a product of $g_i g_A w_{i,A}$ and g_i, g_A are known. Thus, the molecule is known and all other constants are determined.

d) Let N, K be two different decoupled molecules. Then their permutation of the microstate constants of macrostate $(1,1)$ differs, as otherwise $N = K$, due to identical binding and interaction constants (injectivity). Conversely, every permutation of the microstate constants solves the system described by polynomial (4.10) (surjectivity).

□

Remark 51. *In Proposition 50 we used the term permutation for permutations of numbers. This means that different permutations of the symmetric group $\mathcal{S}_n$ can be regarded as equal if some numbers are equal.*

4.4.3 The existence of point-wise decoupled systems for fixed activity of the second ligand

Regarding the one-dimensional titration curve of site 1 of molecule K of Example 48 for fixed ligand activity $-\log(\kappa) = 6$, we see that it is not of classical Henderson-Hasselbalch form (Fig. 4.2). This means that even though the ligand binding sites have a trivial interaction constant in the decoupled molecule, the sites interact. This result may be counterintuitive since the binding sites for ligand L_1 do not interact directly. However, a secondary interaction of the binding sites for ligand L_1 results from the interaction with the second ligand. Let $M = (g_1, g_2, g_A, w_{1,2}, w_{1,A}, w_{2,A})$ be a molecule with two binding sites for ligand L_1 (sites $1, 2$) and one binding site for ligand L_2 (site A). We investigate the following question: Which conditions on the binding and interaction constants are necessary to let the titration curves of the individual sites Ψ_1 and Ψ_2 be of HH shape for all κ?

Proposition 52. *Let $M = (g_1, g_2, g_A, w_{1,2}, w_{1,A}, w_{2,A})$ be a molecule.*
Then the one-dimensional titration curves of sites 1 and 2 are of HH shape for all κ if and only if $w_{1,2} = 1$ and ($w_{1,A} = 1$ or $w_{2,A} = 1$).

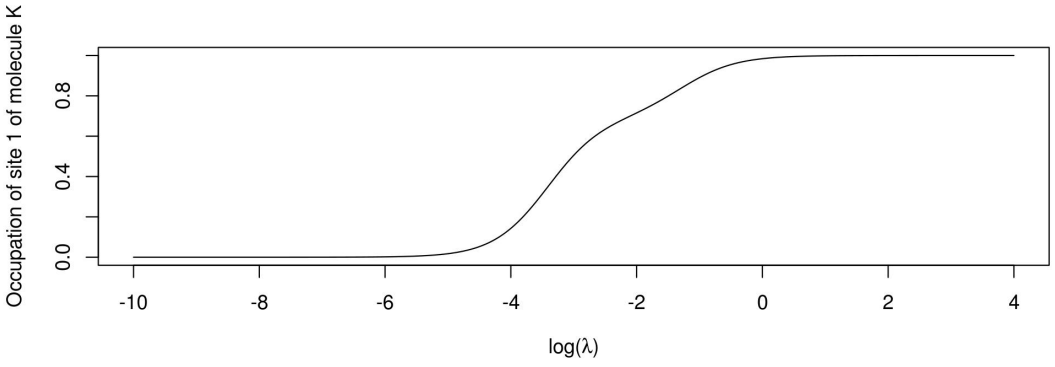

Figure 4.2: Activity dependent ligand binding to site 1 of molecule K of Example 48 for fixed activity of the second ligand: $6 = -\log(\kappa)$.

Proof. "$\Longrightarrow$" Let M be a molecule as described. Its bp is given by

$$\Phi(M) = g_1 g_2 g_A w_{1,2} w_{1,A} w_{2,A} \lambda^2 \kappa + g_1 g_2 w_{1,2} \lambda^2 +$$

$$(g_1 g_A w_{1,A} + g_2 g_A w_{2,A}) \lambda \kappa + (g_1 + g_2) \lambda + g_A \kappa + 1.$$

Since the sites 1 and 2 are of HH shape for all κ, they are in particular decoupled for $\kappa = 0$. This implies $w_{1,2} = 1$, according to Proposition 29. Thus, the titration curve of site 1 has the shape

$$\frac{(g_1 g_2 g_A w_{1,A} w_{2,A} \kappa + g_1 g_2) \lambda^2 + (g_1 g_A w_{1,A} \kappa + g_1) \lambda}{\Phi(M)}. \tag{4.12}$$

We know that for all fixed κ, site 1 has a HH titration curve. This implies in particular that Eq. (4.12) can be rewritten (for fixed κ) as

$$\frac{g_1'(\kappa) \lambda}{g_1'(\kappa) \lambda + 1}. \tag{4.13}$$

where $g_1'(\kappa)$ depends on κ but not on λ. Equality of Eq. (4.12) and Eq. (4.13) shows that a factor $a(\kappa) \in \mathbb{C}[\lambda]$ exists ($a(\kappa)$ has to be of degree one), such that

$$a(\kappa) g_1'(\kappa) \lambda = (g_1 g_2 g_A w_{1,A} w_{2,A} \kappa + g_1 g_2) \lambda^2 + (g_1 g_A w_{1,A} \kappa + g_1) \lambda \tag{4.14}$$

and

$$a(\kappa)(g_1'(\kappa) \lambda + 1) = \Phi(M). \tag{4.15}$$

Regarding these polynomials as elements of $\mathbb{C}[\lambda]$, we see that the constant term of $\Phi(M)$ is given by $g_A \kappa + 1$. Moreover, the constant term of $(g_1'(\kappa)\lambda + 1)$ is 1. This implies that $a(\kappa)$ has constant term $g_A \kappa + 1$. Since $g_1'(\kappa)$ is independent of λ, Eq. (4.14) implies

$$g_1'(\kappa) = \frac{(g_1 g_A w_{1,A} \kappa + g_1)}{g_A \kappa + 1} \tag{4.16}$$

$$a(\kappa) = \frac{(g_A \kappa + 1) g_2 (g_A w_{1,A} w_{2,A} \kappa + 1)}{(g_A w_{1,A} \kappa + 1)} \lambda + (g_A \kappa + 1). \tag{4.17}$$

The same arguments for site 2 show that its titration curve is of shape of Eq. (4.13) with

$$g_2'(\kappa) = \frac{(g_2 g_A w_{2,A}\kappa + g_2)}{g_A\kappa + 1}. \tag{4.18}$$

As the overall titration curve is the sum of the individual curves we have necessarily

$$\frac{g_1'(\kappa)\lambda}{g_1'(\kappa)\lambda + 1} + \frac{g_2'(\kappa)\lambda}{g_2'(\kappa)\lambda + 1} = \tag{4.19}$$

$$= \frac{2(g_1 g_2 g_A w_{1,A} w_{2,A}\kappa + g_1 g_2)\lambda^2 + (g_1 g_A w_{1,A} + g_2 g_A w_{2,A}\kappa + g_1 + g_2)\lambda}{\Phi(M)}.$$

Hence, a $b(\kappa)$ must exist which is independent of λ such that

$$(g_1'(\kappa)\lambda + 1)(g_2'(\kappa)\lambda + 1)b(\kappa) = \Phi(M). \tag{4.20}$$

Again, a comparison of the constant term of the polynomials gives $b(\kappa) = (g_A\kappa + 1)$. Thus, comparing the leading coefficients of the polynomials of Eq. (4.20) yields

$$\frac{(g_1 g_A w_{1,A}\kappa + g_1)(g_2 g_A w_{2,A}\kappa + g_2)}{g_A\kappa + 1} = g_1 g_2 g_A w_{1,A} w_{2,A}\kappa + g_1 g_2 \tag{4.21}$$

which gives

$$(w_{1,A} + w_{2,A})\kappa = (1 + w_{1,A} w_{2,A})\kappa \qquad \forall\kappa$$

and thus

$$w_{1,A}(1 - w_{2,A}) = 1 - w_{2,A}. \tag{4.22}$$

Eq. (4.22) shows that $w_{2,A} \neq 1$ implies $w_{1,A} = 1$.

"$\Longleftarrow$" Without loss of generality, let $w_{1,A} = 1 = w_{1,2}$. Then, using notation of Eqs. (4.16, 4.17) we can rewrite

$$a(\kappa)(g_1'(\kappa)\lambda + 1) = \Phi(M)$$

and

$$a(\kappa)g_1'(\kappa)\lambda = (g_1 g_2 g_A w_{2,A}\kappa + g_1 g_2)\lambda^2 + (g_1 g_A\kappa + g_1)\lambda.$$

Thus, the 1-dimensional titration curve for fixed κ (Eq. (4.12)) reduces to

$$\frac{(g_1 g_2 g_A w_{2,A}\kappa + g_1 g_2)\lambda^2 + (g_1 g_A\kappa + g_1)\lambda}{\Phi(M)} = \frac{g_1'(\kappa)\lambda}{g_1'(\kappa)\lambda + 1} = \frac{g_1\lambda}{g_1\lambda + 1}. \tag{4.23}$$

Moreover, the titration curve of site 2 for fixed κ is given by

$$\frac{(g_1 g_2 w_{2,A}\kappa + g_1 g_2)\lambda^2 + (g_2 g_A w_{2,A}\kappa + g_2)\lambda}{\Phi(M)} =$$

$$= \frac{(g_2 g_A w_{2,A}\kappa + g_2)\lambda(g_1\lambda + 1)}{(g_1\lambda + 1)a(\kappa)} = \frac{g_2'(\kappa)\lambda}{g_2'(\kappa)\lambda + 1},$$

with $g_2'(\kappa)$ of Eq. (4.18). □

Proposition 52 can be interpreted the following way: The different binding sites for the same type of ligand have HH shape, for all activities of the second ligand if and only if they are not "connected". To avoid different interpretations, we give an exact definition of this term:

Definition 53. *Let M be a fixed tuple with binding sites $1, ..., n_1$ for ligand L_1 and binding sites $A_1, .., A_{n_2}$ for Ligand L_2. Moreover, let i,j $\in \{1, ..., n_1, A_1, ...A_{n_2}\}$. The sites i and j are called connected if a path $I = \{(i, k_1), (k_1, k_2), ..., (k_p, j)\}$ exists with $w_i \neq 1 \; \forall i \in I$.*

Using this definition we conjecture that Proposition 52 can be generalized to molecules with more binding sites.

Conjecture 54. *Let M be a molecule with binding sites $1, ..., n_1$ for ligand L_1 and binding sites $A_1, .., A_{n_2}$ for ligand L_2. Then the one-dimensional titration curves of the binding sites ligand L_1 are of HH shape for all κ if and only if all binding sites for ligand L_1 are pairwise not connected.*

4.5 Molecules with n to two binding sites

We prove the DSR for the special case of two binding sites for both ligands, and present an iterative approach to calculate decoupled systems with $(n, 2)$ bindings sites. Exemplarily, this is used subsequently to decouple a molecule with $(4, 2)$ binding sites.

4.5.1 Two to two binding sites

We formulate the DSR as proposition for the case $n_1 = n_2 = 2$.

Proposition 55. *Let*

$$M = (g_1^M, g_2^M, g_A^M, g_B^M, w_{1,2}^M, w_{1,A}^M, w_{1,B}^M, w_{2,A}^M, w_{2,B}^M, w_{A,B}^M)$$

be a molecule with two binding sites for each type of ligand whose binding polynomial has real, positive coefficients, only. Then a decoupled molecule

$$N = (g_1, g_2, g_A, g_B, 1, w_{1,A}, w_{1,B}, w_{2,A}, w_{2,B}, 1)$$

exists, with

$$\Phi(M) = \Phi(N).$$

Proof. We will use the special structure of the algebraic equations we are dealing with to prove Proposition 55. The binding polynomial of M is

$$\Phi(M) = a_{2,2}\lambda^2\kappa^2 + a_{2,1}\lambda^2\kappa + a_{2,0}\lambda^2 + a_{1,2}\lambda\kappa^2 + a_{1,1}\lambda\kappa + a_{1,0}\lambda + a_{0,2}\kappa^2 + a_{0,1}\kappa + 1.$$

We seek for a molecule

$$N = (g_1, g_2, g_A, g_B, 1, w_{1,A}, w_{1,B}, w_{2,A}, w_{2,B}, 1)$$

with the same binding polynomial. The bp gives a system of eight equations corresponding to its coefficients

$$\begin{aligned}
a_{2,2} &= g_1g_2g_Ag_Bw_{1,A}w_{1,B}w_{2,A}w_{2,B}\\
a_{2,1} &= g_1g_2g_Aw_{1,A}w_{2,A} + g_1g_2g_Bw_{1,B}w_{2,B}\\
a_{2,0} &= g_1g_2\\
a_{1,2} &= g_1g_Ag_Bw_{1,A}w_{1,B} + g_2g_Ag_Bw_{2,A}w_{2,B}\\
a_{1,1} &= g_1g_Aw_{1,A} + g_1g_Bw_{1,B} + g_2g_Aw_{2,A} + g_2g_Bw_{2,B}\\
a_{1,0} &= g_1 + g_2\\
a_{0,2} &= g_Ag_B\\
a_{0,1} &= g_A + g_B
\end{aligned} \tag{4.24}$$

The binding energies g_i can be calculated using the equations given by the coefficients with only one type of ligand, according to Proposition 23:

$$(g_1, g_2) = \left(-\frac{1}{\lambda_{z_1}}, -\frac{1}{\lambda_{z_2}}\right)$$

with λ_{z_i} denoting the roots of

$$a_{2,0}\lambda^2 + a_{1,0}\lambda + 1.$$

Analogously, (g_A, g_B) can be calculated using $a_{0,2}, a_{0,1}$. Thus, in general for any choice of (n_1, n_2), the subsystem of equations given by the coefficients $a_{i,0}$ and $a_{0,j}$, is enough to calculate $(g_i)_{i=1,\dots,n_1+n_2}$. With the same argument equations $a_{2,2}$ and $a_{2,1}$ give the products $g_Aw_{1,A}w_{2,A}$ and $g_Bw_{1,B}w_{2,B}$ by calculating the roots of

$$\frac{a_{2,2}}{g_1g_2}\lambda^2 + \frac{a_{2,1}}{g_1g_2}\lambda + 1. \tag{4.25}$$

Analogously, $a_{2,2}$ and $a_{1,2}$ give $g_1w_{1,A}w_{1,B}$ and $g_2w_{2,A}w_{2,B}$. This means we have already found (g_i) solving the subsystem $\{a_{0,j}, a_{j,0}\}$, and products $(g_iw_{i,A}w_{i,B})_{i=1,2}$, $(g_jw_{1,j}w_{2,j})_{j=A,B}$ such that all equations, except for $a_{1,1}$ are solved. The remaining question is whether the products can be factorized such that all required conditions are fulfilled. As we know the binding constants we can rewrite the conditions on the products to

$$\begin{aligned}
w_{1,A}w_{1,B} &= b_1\\
w_{2,A}w_{2,B} &= b_2\\
w_{1,A}w_{2,A} &= b_A\\
w_{1,B}w_{2,B} &= b_B\\
g_1g_Aw_{1,A} + g_2g_Bw_{2,B} + g_1g_Bw_{1,B} + g_2g_Aw_{2,A} &= a_{1,1}
\end{aligned} \tag{4.26}$$

If $(w_{i,j})$ solving system (4.26) exist then the whole system will be solved. Rearranging the first four equations of system (4.26) shows that we can solve them simultaneously for any choice of $w_{1,A} \neq 0$ if and only if $\frac{b_Ab_B}{b_2b_1} = 1$. However, this is true as the b_is are derived from the roots of polynomials and fulfill in particular

$$a_{2,2} = w_{1,A}w_{1,B}w_{2,A}w_{2,B}\prod_{i=1}^{B} g_i = b_1b_2\prod_{i=1}^{B} g_i = b_Ab_B\prod_{i=1}^{B} g_i.$$

Consequently, we can solve the first four equations to receive expressions depending on $w_{1,A}$, only, and plug them into the last equation which gives a polynomial of degree two. Unfortunately, contrarily to the statement of the proof presented in Martini et al. (2013a), it is not obvious (or even not correct) that the polynomial has to have a root which is nonzero. However, this would be required if we want to stop here, since $w_{1,A} = 0$ does not solve system (4.26). We will prove the statement that a solution exists by a contradiction: Assume now, that a solution to the system does not exist. Regarding the polynomial

$$\left(g_1 g_A + g_2 g_B \frac{b_B}{b_1}\right) w_{1,A}^2 - a_{1,1} w_{1,A} + g_1 g_B b_1 + g_2 g_A b_A, \tag{4.27}$$

we see, that if the leading term vanishes, we receive a linear equation ($a_{1,1} \neq 0$), which has a solution which is nonzero if the constant term does not vanish. Conversely, if the constant term equals zero, we will receive a linear equation, with a nonzero solution, if the leading term is nonzero. Since we assume that a solution does not exist, we know that the leading coefficient as well as the constant coefficient have to be zero:

$$g_1 g_A b_1 + g_2 g_B b_B = 0 \tag{4.28}$$

and

$$g_1 g_B b_1 + g_2 g_A b_A = 0. \tag{4.29}$$

Recall, that a permutation of the negative inverses of the roots of a polynomial with positive coefficients was chosen for $g_1 b_1 := N_1$ and $g_2 b_2 := N_2$ (analogously for $g_A b_A =: N_A$ and $g_B b_B =: N_B$). Assuming that a solution to the system does not exist implies that Eqs. (4.28, 4.29) are true for permutations of N_1, N_2 or N_A, N_B:

$$g_A N_1 + g_2 N_B = 0 \text{ and } g_B N_1 + g_2 N_A = 0 \tag{4.30}$$

$$g_A N_2 + g_2 N_B = 0 \text{ and } g_B N_2 + g_2 N_A = 0 \tag{4.31}$$

$$g_A N_1 + g_2 N_A = 0 \text{ and } g_B N_1 + g_2 N_B = 0 \tag{4.32}$$

Eqs. (4.30,4.31) imply that $N_1 = N_2$, which means that they are real valued and consequently positive, since they are the negative inverses of roots of a polynomial with positive coefficients (if the imaginary part was nonzero, they would have to be complex conjugates). Moreover, Eqs. (4.31,4.32) then imply that $g_A N_1 = g_B N_1$ and thus $g_A = g_B$ are real valued. This implies that $N_A = N_B \in \mathbb{R}$ and thus $g_2 \in \mathbb{R}^+$. This gives a contradiction, since –with all those factors positive– Eqs. (4.28, 4.29) can not be fulfilled. □

Remark 56. *The proof of Proposition 55 instructs us how to calculate a solution for the algebraic system. A generalization of this procedure might be adequate to prove the DSR without fixing n_1 and n_2.*

4.5.2 Decoupling a molecule with four and two binding sites for different ligands

Here, we illustrate the decoupling of a hypothetical molecule with four binding sites for ligand L_1 and two binding sites for ligand L_2. Let the sites for L_1 be denoted by $1, ..., 4$ and the binding sites for L_2 be called A, B. Even though the hypothetical molecule has only six binding sites, decoupling is challenging. To find a decoupled molecule for a system with four and two binding sites, we have to solve system (4.33) consisting of 14 polynomial equations (one equation per coefficient) with 14 variables given by the binding constants and the interaction constants. To facilitate identifying the structure of the system, which is required to understand how we find solutions, we use the following substitutions:

$$
\begin{aligned}
\xi_i &:= && g_i w_{i,A} w_{i,B} \\
a^j_{4,1} &:= && g_1 g_2 g_3 g_4 g_j w_{1,j} w_{2,j} w_{3,j} w_{4,j} \\
a^j_{3,1} &:= && g_1 g_2 g_3 g_j w_{1,j} w_{2,j} w_{3,j} + g_1 g_2 g_4 g_j w_{1,j} w_{2,j} w_{4,j} + \\
& && + g_1 g_3 g_4 g_j w_{1,j} w_{3,j} w_{4,j} + g_2 g_3 g_4 g_j w_{2,j} w_{3,j} w_{4,j} \\
a^j_{2,1} &:= && g_1 g_2 g_j w_{1,j} w_{2,j} + g_1 g_3 g_j w_{1,j} w_{3,j} + g_1 g_4 g_j w_{1,j} w_{4,j} + \\
& && + g_2 g_3 g_j w_{2,j} w_{3,j} + g_2 g_4 g_j w_{2,j} w_{4,j} + g_3 g_4 g_j w_{3,j} w_{4,j} \\
a^j_{1,1} &= && g_1 g_j w_{1,j} + g_2 g_j w_{2,j} + g_3 g_j w_{3,j} + g_4 g_j w_{4,j}
\end{aligned}
$$

with $i \in \{1, 2, 3, 4\}$ and $j \in \{A, B\}$. A look at system (4.33) reveals that it consists of three systems of the type described in the proof of Proposition 23 and a system which is a sum of two systems of the same structure ($\{a_{i,1}\}_{i=1,2,3,4}$).

$$
\begin{aligned}
\tfrac{a_{4,2}}{g_A g_B} &= && \xi_1\xi_2\xi_3\xi_4 \\
\tfrac{a_{3,2}}{g_A g_B} &= && \xi_1\xi_2\xi_3 + \xi_1\xi_2\xi_4 + \xi_1\xi_3\xi_4 + \xi_2\xi_3\xi_4 \\
\tfrac{a_{2,2}}{g_A g_B} &= && \xi_1\xi_2 + \xi_1\xi_3 + \xi_1\xi_4 + \xi_2\xi_3 + \xi_2\xi_4 + \xi_3\xi_4 \\
\tfrac{a_{1,2}}{g_A g_B} &= && \xi_1 + \xi_2 + \xi_3 + \xi_4 \\
a_{4,1} &= && a^A_{4,1} + a^B_{4,1} \\
a_{3,1} &= && a^A_{3,1} + a^B_{3,1} \\
a_{2,1} &= && a^A_{2,1} + a^B_{2,1} \\
a_{1,1} &= && a^A_{1,1} + a^B_{1,1} \\
a_{0,2} &= && g_A g_B
\end{aligned}
\tag{4.33}
$$

$$
\begin{aligned}
a_{0,1} &= g_A + g_B \\
a_{4,0} &= g_1g_2g_3g_4 \\
a_{3,0} &= g_1g_2g_3 + g_1g_2g_4 + g_1g_3g_4 + g_2g_3g_4 \\
a_{2,0} &= g_1g_2 + g_1g_3 + g_1g_4 + g_2g_3 + g_2g_4 + g_3g_4 \\
a_{1,0} &= g_1 + g_2 + g_3 + g_4
\end{aligned}
$$

To find a decoupled system for Example 58, we tried to use the standard command "algsys" of the computer algebra system Maxima to solve the system. However, it was too complicated to be solved directly by this general approach. Instead we used the special structure of the system to deduce an iterative procedure, similar to Proposition 35, with decoupled systems as fixed point: Regarding the system of polynomial equations we see that the equations given by $a_{4,0}, a_{3,0}, a_{2,0}, a_{1,0}$ do neither contain any interaction constant as variable, nor the binding constants g_A, g_B. Thus, we have four equations with the four variables g_1, g_2, g_3, g_4. This subsystem can be solved using the well known procedure: $g_i = -\frac{1}{z_i}$ where z_i denote the roots of the polynomial

$$P_1(x) = a_{4,0}x^4 + a_{3,0}x^3 + a_{2,0}x^2 + a_{1,0}x + 1.$$

Analogously, coefficients $a_{0,2}, a_{0,1}$ immediately give a solution for (g_A, g_B). Hence, the binding energies are unique, except for permutations. We chose any permutation which means we fix the binding constants. The products $\xi_i = g_i w_{i,A} w_{i,B}$ can be calculated using equations $a_{4,2}, a_{3,2}, a_{2,2}, a_{1,2}$: Again, these products are the negative inverses of the roots of the polynomial

$$P_2(x) = \frac{a_{4,2}}{g_A g_B}x^4 + \frac{a_{3,2}}{g_A g_B}x^3 + \frac{a_{2,2}}{g_A g_B}x^2 + \frac{a_{1,2}}{g_A g_B}x + 1.$$

Note, that this is the major step which distinguishes between the different decoupled molecules: We have fixed an order of the binding constants previously, and have to relate the roots of P_2 and the products $g_i w_{i,A} w_{i,B}$. In general, we will receive different decoupled molecules for different permutations of the roots of P_2 (if they do not coincide due to identical binding constants, etc.). As we know the binding constants g_i, these solutions give conditions on $w_{i,A}w_{i,B}$. Regarding equations $a_{4,1}, a_{3,1}, a_{2,1}, a_{1,1}$ we see that the system is the sum of two "ordinary" systems which could be solved by the well known procedure previously described, if we knew $\{a^A_{i,1}\}_{i=1,2,3,4}$. As this is not the case we use the iterative approach described in Proposition 57.

Proposition 57. *Let the algebraic system (4.33) be given. Let $(g_i)_{i=1,\dots,4}$ be a solution to the subsystem $\{a_{4,0}, a_{3,0}, a_{2,0}, a_{1,0}\}$, (g_A, g_B) be a solution to $\{a_{0,2}, a_{0,1}\}$. Furthermore, let the products $(w_{i,A}w_{i,B})_{i=1,\dots,4}$ be a solution to $\{a_{4,2}, a_{3,2}, a_{2,2}, a_{1,2}\}$ for the given binding constants $(g_i)_{i=1,2,3,4,A,B}$ and let σ^n be a sequence of permutations of $\{1, ..., 4\}$. We consider the following algorithm:*

1) $a^{A,0}_{i,1} := a_{i,1}, \qquad i \in \{1,2,3,4\}$

2) $P_n(x) := \frac{a^{A,n}_{4,1}}{g_A}x^4 + \frac{a^{A,n}_{3,1}}{g_A}x^3 + \frac{a^{A,n}_{2,1}}{g_A}x^2 + \frac{a^{A,n}_{1,1}}{g_A}x + 1$

3) $(z_1^n, z_2^n, z_3^n, z_4^n) :=$ *the roots of* P_n

4) $w_{i,A}^n := -\frac{1}{z_{\sigma^n(i)}^n g_i}$

5) $w_{i,B}^n := \frac{w_{i,A} w_{i,B}}{w_{i,A}^n}$

6) Calculate $a_{i,1}^{B,n}$ *using equations* $\{a_{i,1}\}_{i=1,2,3,4}$ *and*

7) restart with $a_{i,1}^{A,n+1} := a_{i,1} - a_{i,1}^{B,n}$

Then, $x = (w_{1,A}, w_{1,B}, w_{2,A}, w_{2,B}, w_{3,A}, w_{3,B}, w_{4,A}, w_{4,B})$ *is a solution to the subsystem* $\{a_{i,n}\}_{i,n\neq 0}$ *which satisfies the conditions on the products* $(w_{i,A}w_{i,B})_{i=1,\dots,4}$ *if and only if a permutation* $\sigma \in \mathcal{S}_4$ *exists such that* x *is a fixed point of the algorithm with* $\sigma^n \equiv \sigma$ *(the constant sequence).*

Proof. Let x be a solution to the subsystem, satisfying the conditions on the products $(w_{i,A}w_{i,B})_{i=1,\dots,4}$. Then

$$w_{i,A}^n = w_{i,A} \Rightarrow w_{i,B}^n = \frac{w_{i,A}w_{i,B}}{w_{i,A}} = w_{i,B} \Rightarrow a_{i,1}^{B,n} = a_{i,1}^B \Rightarrow a_{i,1}^{A,n+1} = a_{i,1}^A$$

which gives

$$w_{i,A}^{n+1} = w_{i,A} = w_{i,A}^n$$

if the correct permutation of the roots is used.
Conversely, let x be a fixed point and, without loss of generality, let σ be identity. Then $w_{i,A}^n = w_{i,A}^{n+1} = w_{i,A}$. This means that the roots of P_n coincide with the roots of P_{n+1}. As both polynomials have the same roots and the same constant term, this shows that all coefficients are equal, which implies $a_{i,1}^{A,n} + a_{i,1}^{B,n} = a_{i,1}$, and that equations $a_{i,1}$ are satisfied. Consequently, x solves the system. □

Example 58. *Let the molecule be described by* $M = (G, W)$ *with*

$$G = (g_1, g_2, g_3, g_4, g_A, g_B) = (4 \cdot 10^3, 10, 2 \cdot 10^3, 500, 10^3, 10)$$

$$W = (w_{i,j})_{i,j=1,2,3,4,A,B} = \begin{pmatrix} 1 & 0.1 & 0.1 & 0.05 & 10 & 1 \\ 0.1 & 1 & 0.5 & 0.5 & 10^3 & 2 \cdot 10^3 \\ 0.1 & 0.5 & 1 & 0.05 & 10^2 & 10 \\ 0.05 & 0.5 & 0.05 & 1 & 10^2 & 20 \\ 10 & 10^3 & 10^2 & 10^2 & 1 & 0.1 \\ 1 & 2 \cdot 10^3 & 10 & 20 & 0.1 & 1 \end{pmatrix}$$

For the sake of a clear view, we use a matrix notation for W *which repeats information but underlines which pairwise interaction is described. The binding polynomial of the molecule is given by*

$$\Phi(M) =$$
$$= 10^{22}\lambda^4\kappa^2 + 2.5001 \cdot 10^{16}\lambda^4\kappa + 250000\lambda^4 +$$
$$+5.1002 \cdot 10^{20}\lambda^3\kappa^2 + 1.800602 \cdot 10^{15}\lambda^3\kappa + 1575000\lambda^3 +$$
$$+3.0190 \cdot 10^{16}\lambda^2\kappa^2 + 2.69328 \cdot 10^{12}\lambda^2\kappa + 966500\lambda^2 +$$
$$+2.3040 \cdot 10^{10}\lambda\kappa^2 + 3.0054 \cdot 10^{8}\lambda\kappa + 6510\lambda +$$

$$1000\kappa^2 + 1010\kappa + 1$$

The binding to the individual sites is illustrated in Fig. 4.4. We used the iterative approach described in Proposition 57 to calculate 24 different decoupled molecules, which correspond to the different permutations of the roots of P_n in step 3) of Proposition 57. The titration curves of the individual binding sites of two different decoupled molecules N and K are illustrated in Fig. 4.4. The binding constants of the decoupled molecules are always coinciding and given by (rounded):

$$G^{De} = (6358.026, 150.328, 1.468, 0.178, 1009.009, 0.991).$$

The interaction constants of two chosen decoupled molecules N, K are (rounded)

$$W_N = \begin{pmatrix} 1 & 1 & 1 & 1 & 45.323 & 75.119 \\ 1 & 1 & 1 & 1 & 56.358 & 162.473 \\ 1 & 1 & 1 & 1 & 487.352 & 23.900 \\ 1 & 1 & 1 & 1 & 79.618 & 1.384 \\ 45.323 & 56.358 & 487.352 & 79.618 & 1 & 1 \\ 75.119 & 162.473 & 23.900 & 1.384 & 1 & 1 \end{pmatrix}$$

$$W_K = \begin{pmatrix} 1 & 1 & 1 & 1 & 45.336 & 6.810 \cdot 10^{-05} \\ 1 & 1 & 1 & 1 & 56.464 & 2550.214 \\ 1 & 1 & 1 & 1 & 486.320 & 23.950 \\ 1 & 1 & 1 & 1 & 79.613 & 97028.384 \\ 45.336 & 56.464 & 486.320 & 79.613 & 1 & 1 \\ 6.810 \cdot 10^{-5} & 2550.214 & 23.950 & 97028.384 & 1 & 1 \end{pmatrix}$$

Remark 59. *a) The fixed point algorithm described in Proposition 57 can easily be generalized to a situation of $(n, 2)$ binding sites. Only the degree of the polynomial whose roots have to be calculated increases.*

b) We note that it is not clear whether this procedure will always be attracted by its fixed point. However, our numerical test suggest that it converges quickly.

c) The algorithm described in Proposition 57 can also be used with site B as reference site ($a_{i,1}^{B,0} := a_{i,1}, ...$). In all examples we calculated, this altered procedure led to another molecule with different titration curves of sites A and B. However, the titration curves of sites $1, ..., 4$ only depend on the chosen permutation of the products $w_{i,A}w_{i,B}$, and not on the choice of the reference site. In particular, this shows that our iterative approach has at least two (which equals $n_2!$) fixed points for any permutation of the products $(w_{i,A}w_{i,B})_{i=1,..,4}$. Switching the reference site and calculating the decoupled molecules with our iterative procedure and all possible permutations of the roots of P_n gives additional 24 decoupled molecules sharing other titration curves of the binding sites for the second ligand. Moreover, not only the reference site, but also the starting point can be changed (e.g. $a_{i,1}^{A,0} := \frac{1}{2}a_{i,1}$). Yet, we do not know how the choice of the starting point and reference site determines which fixed point will be reached. It might be the case that the choice of the reference site implies that a certain fixed point is attractive and the other one repulsive.

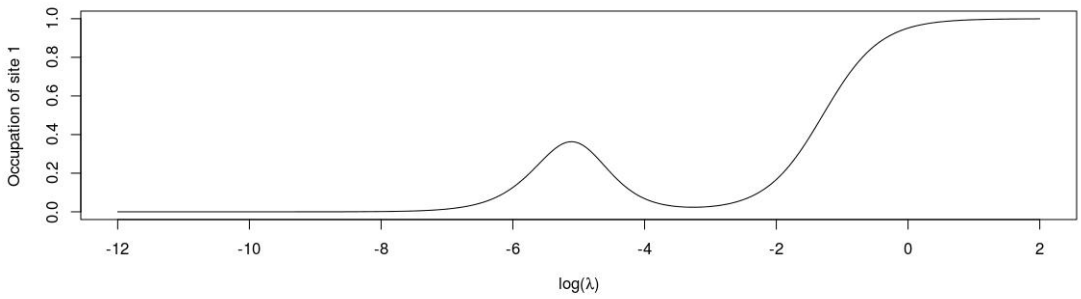

Figure 4.3: Ligand binding to site 1 of the decoupled molecule K of Example 58, dependent on $\log(\lambda)$ for fixed $\log(\kappa) = -3$.

d) Our implementation is based on the R function "polyroot" (R Core Team, 2012) and we regarded the permutation which was returned by this function as $id \in \mathcal{S}_4$.

Regarding the titration curve of site 1 of the decoupled molecule K for fixed activity $\log(\kappa) = -3$ (Fig. 4.3), we can see an extreme form of secondary interaction, which we have already described, previously in this chapter: Even though none of the binding sites for ligand L_1 interact directly, their one-dimensional titration curves are not of classical Henderson-Hasselbalch shape when the activity κ of the second ligand is fixed. Secondary interaction between the binding sites for ligand L_1 is a result of the interaction with the second ligand: As site 1 of the decoupled molecules has a great binding constant (compared to the other binding sites), it will be occupied at a comparatively low activity. With an increase of activity λ, more ligands will bind to the other sites which will enhance the binding of the second ligand, in particular to site B. However, this decreases the affinity of ligand L_1 to site 1, due to the small interaction constant $w_{1,B}$ of molecule K.
Analogously to our observation in the first part of this chapter, we see here once more that in the decoupled molecules, the area of transition from 0.1 to 0.9 occupation probability is small compared to the original molecule (Fig. 4.4).

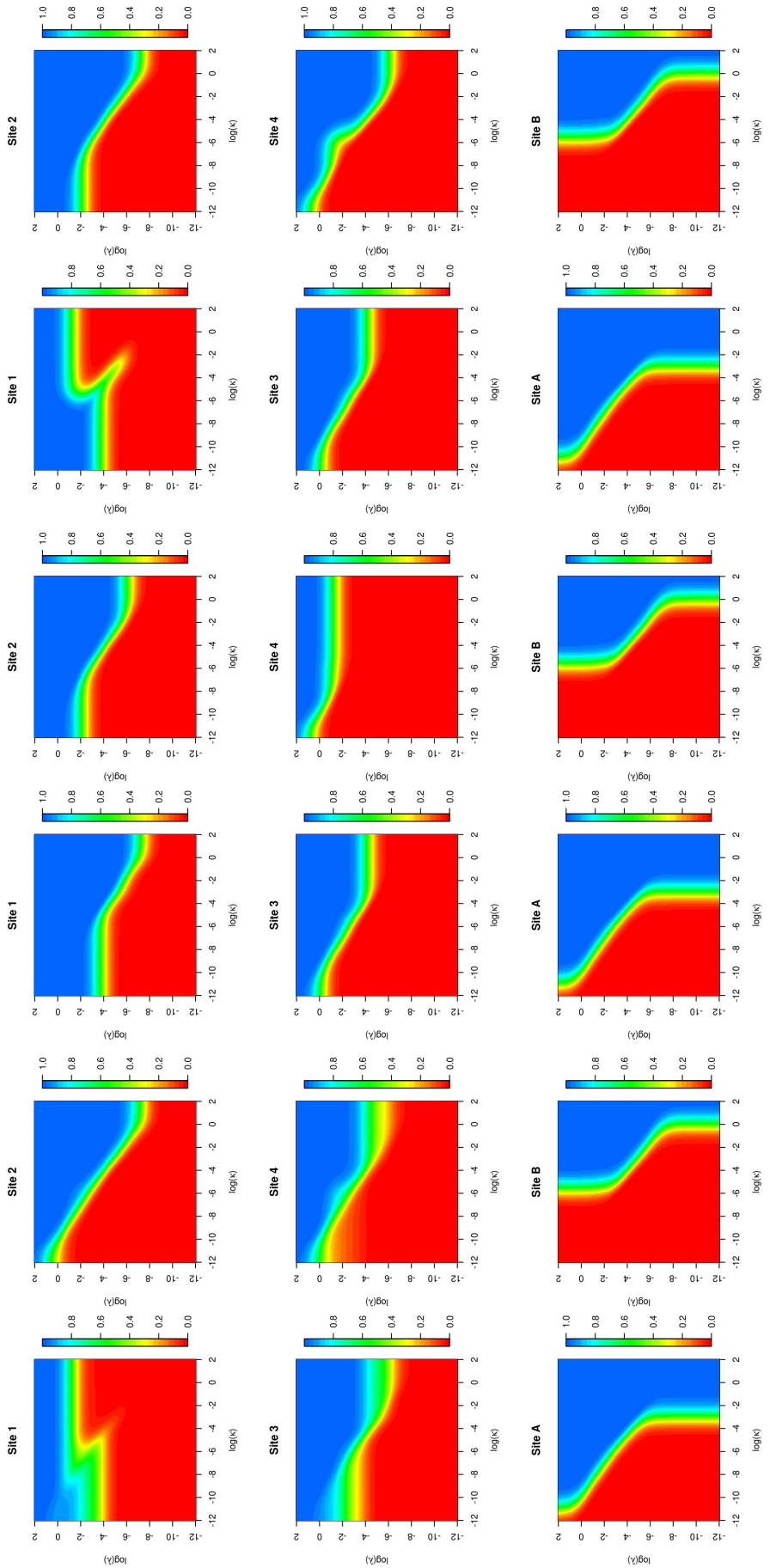

Figure 4.4: Activity dependent ligand binding to each site in the original molecule M (left pair of columns) and the decoupled molecules N (middle pair of columns) and K (right pair of columns) of Example 58. A logarithmic scale of the ligand activities λ and κ is used. The probability of occupation is encoded by the colors, according to the colorbars.

4.6 Unique features shared by all decoupled molecules

Fig. 4.4 creates the suspicion that the binding curves of the individual sites for ligand one of the decoupled system N in Example 58 are similar to those of the decoupled molecule K in a certain way: The titration curve of site 1 of molecule K seems to share the "right part" with site 4 of molecule N and its "left part" seems to be identical to the "left part" of site 1 of molecule N. We have already described this observation for the case of $(n_1, 1)$ binding sites. Analogously, to the first part of this chapter, we want to identify unique features all decoupled molecules share. We will have a look on the microstate constants of the different molecules first. Table 1 lists the non-trivial microstate constants of the molecules M, N and K of Example 58. Microstates constants which are not listed are identical for the molecules N and K (when the permutation of the binding constants is fixed). We see here, that the microstate constants of macrostate $(1, 1)$ of molecules N and K are not permutations of each other. However, for all macrostates in which the binding sites for one type of ligand are fully occupied, the corresponding microstate constants are permutations. We can prove this statement in general.

Proposition 60. *Let M be a molecule with n_1 binding sites for ligand L_1 and n_2 binding sites for ligand L_2. Moreover, let the order of the sites in the decoupled molecules be fixed to the same permutations. Then the following statements hold:*

a) For any microstate k with only one type of ligand bound, all decoupled molecules share the same microstate constant $g(k)$.

b) For every macrostate (i, n_2), $\binom{n_1}{i}$ numbers exist such that for any decoupled molecule the tuple of its constants of microstates belonging to this macrostate is a permutation of these numbers. Analogously, this statements holds for macrostates (n_1, j).

c) The permutation of microstate constants of macrostate $(1, n_2)$ fixes the permutations of the microstate constants of all other macrostates (i, n_2). Analogously, for macrostate $(n_1, 1)$ and (n_1, j).

Proof. a) As the permutation of the binding sites is fixed, and since the constants $g(k)$ are the product of the binding constants they are identical for all decoupled molecules.

b) Let k be a microstate of macrostate $(1, n_2)$. Its constant is given by

$$g(k) = g_i g_{A_1} g_{A_2} \cdots g_{A_{n_2}} w_{i,A_1} \cdots w_{i,A_n}.$$

The coefficients $a_{n_1,n_2}, a_{n_1-1,n_2}, ..., a_{1,n_2}$ are enough to calculate these constants which are the negative inverses of the roots of a polynomial. Thus, for any decoupled molecule, the constants correspond to a permutation of these roots as the decoupled molecule has to fulfill the equations given by $a_{n_1,n_2}, a_{n_1-1,n_2}, ..., a_{1,n_2}$ in particular. Let k_1 be a microstate of macrostate (i, n_2). Then its constant is the product of i microstate constants belonging to macrostate $(1, n_2)$ divided by

$$\left(\prod_{j=1}^{n_2} g_{A_j} \right)^{i-1}.$$

This proves *b*) and *c*).

□

Table 4.2: Microstate constants of molecules M and the two corresponding decoupled molecules N and K of Example 58. The constants of microstates which are not listed are identical for molecules N and K.

Macrostate	Microstate	M	N	K
$(1,1)$	$(1,0,0,0,1,0)$	$4 \cdot 10^7$	290757860	290757860
	$(0,1,0,0,1,0)$	10^7	8548419	8564541
	$(0,0,1,0,1,0)$	$2 \cdot 10^8$	721821.7	720293.5
	$(0,0,0,1,1,0)$	$5 \cdot 10^7$	14315.07	14314.21
	$(1,0,0,0,0,1)$	$4 \cdot 10^4$	473343.6	0.4291216
	$(0,1,0,0,0,1)$	$2 \cdot 10^5$	24206.05	379944.48
	$(0,0,1,0,0,1)$	$2 \cdot 10^5$	34.76863	34.84239
	$(0,0,0,1,0,1)$	10^5	0.2443494	17135.27
$(2,1)$	$(1,1,0,0,1,0)$	$4 \cdot 10^{10}$	$2.463328 \cdot 10^{12}$	$2.468703 \cdot 10^{12}$
	$(1,0,1,0,1,0)$	$8 \cdot 10^{11}$	208001453616	207622407458
	$(1,0,0,1,1,0)$	10^{11}	4125056936	4126026921
	$(0,1,1,0,1,0)$	10^{12}	$6.115341 \cdot 10^9$	$6.113904 \cdot 10^9$
	$(0,1,0,1,1,0)$	$2.5 \cdot 10^{11}$	121278630	121500042
	$(0,0,1,1,1,0)$	$5 \cdot 10^{11}$	10240671	10218376
	$(1,1,0,0,0,1)$	$8 \cdot 10^7$	$1.156100 \cdot 10^{10}$	$1.645112 \cdot 10^5$
	$(1,0,1,0,0,1)$	$8 \cdot 10^7$	$1.660577 \cdot 10^7$	15.08632
	$(1,0,0,1,0,1)$	$2 \cdot 10^7$	116703.229	7419.357
	$(0,1,1,0,0,1)$	$2 \cdot 10^9$	849193.2	13357436.4
	$(0,1,0,1,0,1)$	10^9	$5.968020 \cdot 10^3$	$6.569103 \cdot 10^9$
	$(0,0,1,1,0,1)$	10^8	8.572232	$6.024124 \cdot 10^5$
$(3,1)$	$(1,1,1,0,1,0)$	$4 \cdot 10^{14}$	$1.762208 \cdot 10^{15}$	$1.762314 \cdot 10^{15}$
	$(1,1,0,1,1,0)$	$5 \cdot 10^{13}$	$3.494787 \cdot 10^{13}$	$3.502202 \cdot 10^{13}$
	$(1,0,1,1,1,0)$	10^{14}	$2.950970 \cdot 10^{12}$	$2.945415 \cdot 10^{12}$
	$(0,1,1,1,1,0)$	$1.25 \cdot 10^{15}$	$8.675993 \cdot 10^{10}$	$8.673431 \cdot 10^{10}$
	$(1,1,1,0,0,1)$	$8 \cdot 10^{10}$	405581389998	5783603
	$(1,1,0,1,0,1)$	$2 \cdot 10^{10}$	2850373606	2844339211
	$(1,0,1,1,0,1)$	$2 \cdot 10^9$	4094166	260837
	$(0,1,1,1,0,1)$	$5 \cdot 10^{11}$	$2.093692 \cdot 10^5$	$2.309452 \cdot 10^{11}$
$(4,1)$	$(1,1,1,1,1,0)$	$2.5 \cdot 10^{16}$	$2.50009 \cdot 10^{16}$	$2.50009 \cdot 10^{16}$
	$(1,1,1,1,0,1)$	10^{12}	$9.99964 \cdot 10^{10}$	$9.99964 \cdot 10^{10}$
$(1,2)$	$(1,0,0,0,1,1)$	$4 \cdot 10^7$	$2.164640 \cdot 10^{10}$	$1.962988 \cdot 10^4$
	$(0,1,0,0,1,1)$	$2 \cdot 10^{10}$	1376483919	21646399232
	$(0,0,1,0,1,1)$	$2 \cdot 10^9$	17097219	17097219
	$(0,0,0,1,1,1)$	10^9	$1.962988 \cdot 10^4$	$1.376484 \cdot 10^9$
$(2,2)$	$(1,1,0,0,1,1)$	$8 \cdot 10^{13}$	$2.979592 \cdot 10^{16}$	$4.249163 \cdot 10^{11}$
	$(1,0,1,0,1,1)$	$8 \cdot 10^{12}$	$3.700932 \cdot 10^{14}$	$3.356164 \cdot 10^8$
	$(1,0,0,1,1,1)$	$2 \cdot 10^{12}$	$4.249163 \cdot 10^{11}$	$2.702022 \cdot 10^{10}$
	$(0,1,1,0,1,1)$	$2 \cdot 10^{16}$	$2.353405 \cdot 10^{13}$	$3.700932 \cdot 10^{14}$
	$(0,1,0,1,1,1)$	10^{16}	$2.702022 \cdot 10^{10}$	$2.979592 \cdot 10^{16}$
	$(0,0,1,1,1,1)$	10^{14}	$3.356164 \cdot 10^8$	$2.353405 \cdot 10^{13}$
$(3,2)$	$(1,1,1,0,1,1)$	$8 \cdot 10^{18}$	$5.094274 \cdot 10^{20}$	$7.264887 \cdot 10^{15}$
	$(1,1,0,1,1,1)$	$2 \cdot 10^{18}$	$5.848904 \cdot 10^{17}$	$5.848904 \cdot 10^{17}$
	$(1,0,1,1,1,1)$	$2 \cdot 10^{16}$	$7.264887 \cdot 10^{15}$	$4.619706 \cdot 10^{14}$
	$(0,1,1,1,1,1)$	$5 \cdot 10^{20}$	$4.619706 \cdot 10^{14}$	$5.094274 \cdot 10^{20}$

4.7 Decoupling a molecule with three to three binding sites

Finally, we show how the algorithm of Proposition 57 can be extended to more than two binding sites for both ligands by presenting an iterative procedure for the case of $(3,3)$. Let

$$\begin{aligned} P_M(\lambda,\kappa) = a_{3,3}\lambda^3\kappa^3 + a_{3,2}\lambda^3\kappa^2 + a_{3,1}\lambda^3\kappa + a_{3,0}\lambda^3 + \\ a_{2,3}\lambda^2\kappa^3 + a_{2,2}\lambda^2\kappa^2 + a_{2,1}\lambda^2\kappa + a_{2,0}\lambda^2 + \\ a_{1,3}\lambda\kappa^3 + a_{1,2}\lambda\kappa^2 + a_{1,1}\lambda\kappa + a_{1,0}\lambda + \\ a_{0,3}\kappa^3 + a_{0,2}\kappa^2 + a_{0,1}\kappa + 1 \end{aligned}$$

be a binding polynomial. We look for a corresponding decoupled molecule N. Let the sites for ligand L_1 be denoted by $1,2,3$ and for ligand L_2 by A,B,C. The coefficients $(a_{i,0})_{i=1,2,3}$ and $(a_{0,j})_{j=A,B,C}$ give the binding constants. Let a permutation be chosen, that is, the order of the sites be fixed. Then the roots of the polynomial

$$P_1(x) = \frac{a_{3,3}}{g_A g_B g_C}x^3 + \frac{a_{2,3}}{g_A g_B g_C}x^2 + \frac{a_{1,3}}{g_A g_B g_C}x + 1$$

give the products $(g_i w_{i,A} w_{i,B} w_{i,C})_{i=1,2,3}$. Analogously to the case of $(4,2)$ binding sites the choice of the permutation is an important step to distinguish between different solutions. Having solved this subsystem, system (4.35) is left to be solved. We use analogous substitutions to the case of $(4,2)$ binding sites to facilitate understanding the structure of the system ($a_{1,2}^{AB}$ denotes the part of coefficient $a_{1,2}$ derived from microstate with sites A and B occupied):

$$\begin{aligned} \xi_i^{jk} &:= && g_i w_{i,j} w_{i,k} \\ a_{3,2}^{jk} &:= && g_j g_k \xi_1^{jk}\xi_2^{jk}\xi_3^{jk} \\ a_{2,2}^{jk} &:= && g_j g_k(\xi_1^{jk}\xi_2^{jk} + \xi_1^{jk}\xi_3^{jk} + \xi_2^{jk}\xi_3^{jk}) \\ a_{1,2}^{jk} &:= && g_j g_k(\xi_1^{jk} + \xi_2^{jk} + \xi_3^{jk}) \\ a_{3,1}^{j} &:= && g_j g_1 w_{1,j} g_2 w_{2,j} g_3 w_{3,j} \\ a_{2,1}^{j} &:= && g_j g_1 w_{1,j} g_2 w_{2,j} + g_j g_1 w_{1,j} g_3 w_{3,j} + g_j g_2 w_{2,j} g_3 w_{3,j} \\ a_{1,1}^{j} &:= && g_j g_1 w_{1,j} + g_j g_2 w_{2,j} + g_j g_3 w_{3,j} \\ & && \text{with } i \in \{1,2,3\} \text{ and } j,k \in \{A,B,C\},\ j \neq k \end{aligned} \tag{4.34}$$

Thus, the systems consisting of equations $a_{3,2}^{jk}, a_{2,2}^{jk}, a_{1,2}^{jk}$ and $a_{3,1}^{j}, a_{2,1}^{j}, a_{1,1}^{j}$ are of well known form and we see that the remaining equations given by the bp are the sum of

the three systems:

$$\begin{aligned} a_{3,2} &= a_{3,2}^{AB} + a_{3,2}^{AC} + a_{3,2}^{BC} \\ a_{2,2} &= a_{2,2}^{AB} + a_{2,2}^{AC} + a_{2,2}^{BC} \\ a_{1,2} &= a_{1,2}^{AB} + a_{1,2}^{AC} + a_{1,2}^{BC} \\ a_{3,1} &= a_{3,1}^{A} + a_{3,1}^{B} + a_{3,1}^{C} \\ a_{2,1} &= a_{2,1}^{A} + a_{2,1}^{B} + a_{2,1}^{C} \\ a_{1,1} &= a_{1,1}^{A} + a_{1,1}^{B} + a_{1,1}^{C} \end{aligned} \tag{4.35}$$

To solve this system of equations we used the iterative procedure described in Proposition 61 which is an extension of the algorithm of Proposition 57.

Proposition 61. *Let the algebraic system (4.35) be given. Moreover, let* $(g_i)_{i=1,\dots,3,A,..,C}$ *and* $(w_{i,A}w_{i,B}w_{i,C})_{i=1,\dots,3}$ *be known (fixed permutations are chosen) and let* σ^n *be a sequence of permutations of* $\{1,2,3\}$*. We consider the following algorithm:*

1) $a_{i,2}^{AB,0} := a_{i,2}, \qquad i \in \{1,2,3\}$

2) $P_n(x) := \frac{a_{3,2}^{AB,n}}{g_A g_B}x^3 + \frac{a_{2,2}^{AB,n}}{g_A g_B}x^2 + \frac{a_{1,2}^{AB,n}}{g_A g_B}x + 1$

3) $(z_1^n, z_2^n, z_3^n) :=$ *the roots of* P_n

4) $w_{i,A}^n w_{i,B}^n := -\frac{1}{z_{\sigma^n(i)}^n g_i}$

5) $w_{i,C}^n := \frac{w_{i,A}w_{i,B}w_{i,C}}{w_{i,A}^n w_{i,B}^n}$

6) *Calculate* $a_{i,1}^{C,n}$ *using equations* $\{a_{i,1}^C\}_{i=1,2,3}$ *of (4.34) and* $w_{i,C}^n$*.*

7) *Use the procedure of Proposition 57 with a sequence of permutations* σ_2^n *and the condition on the products* $w_{i,A}^n w_{i,B}^n$ *to calculate* $w_{i,A}^n$ *and* $w_{i,B}^n$ *from* $a_{i,1} - a_{i,1}^{C,n} = a_{i,1}^{A,n} + a_{i,1}^{B,n}$*.*

8) *Use* $w_{i,A}^n, w_{i,B}^n, w_{i,C}^n$ *to calculate* $a_{i,2}^{AC,n}$ *and* $a_{i,2}^{BC,n}$*.*

9) *Restart with* $a_{i,2}^{AB,n+1} := a_{i,2} - a_{i,2}^{AC,n} - a_{i,2}^{BC,n}$*.*

Then $x = (w_{1,A}, w_{1,B}, w_{1,C}, w_{2,A}, w_{2,B}, w_{2,C}, w_{3,A}, w_{3,B}, w_{3,C})$ *is a solution to system (4.35) which satisfies the conditions on the products* $(w_{i,A}w_{i,B}w_{i,C})_{i=1,\dots,3}$ *if and only if two permutations* $\sigma_1, \sigma_2 \in \mathcal{S}_3$ *exist such that* x *is a fixed point of the algorithm with* $\sigma^n \equiv \sigma_1$ *and* $\sigma_2^n \equiv \sigma_2$ *(constant sequences).*

Proof. Let $x = (w_{1,A}, w_{1,B}, w_{1,C}, w_{2,A}, w_{2,B}, w_{2,C}, w_{3,A}, w_{3,B}, w_{3,C})$ be a solution to system (4.35) which satisfies the conditions on the products $(w_{i,A}w_{i,B}w_{i,C})_{i=1,\dots,3}$.
Let $w_{i,A}^n = w_{i,A}$, $w_{i,B}^n = w_{i,B}$ and $w_{i,C}^n = w_{i,C}$. As x solves the system, $a_{i,2}^{AC,n} = a_{i,2}^{AC}$ and $a_{i,2}^{BC,n} = a_{i,2}^{BC}$ and consequently $a_{i,2}^{AB,n+1} = a_{i,2}^{AB}$. The roots of the polynomial give

exact solutions $w_{i,A}^{n+1}w_{i,B}^{n+1}$ and thus exact solutions $w_{i,C}^{n+1} = w_{i,C}$, if the appropriate permutation σ_1 is used. This means $a_{i,1}^{C,n+1} = a_{i,1}^{C}$. As $w_{i,A}^{n+1}w_{i,B}^{n+1} = w_{i,A}w_{i,B}$, if an appropriate permutation σ_2 is used in the procedure of Proposition 57, $w_{i,A}$ and $w_{i,B}$ will be fixed. Consequently,

$$a_{i,2}^{AB,n+1} = a_{i,2}^{AB,n} = a_{i,2}^{AB} \text{ and } P_n = P_{n+1}.$$

Conversely, let $x = (w_{1,A}, w_{1,B}, w_{1,C}, w_{2,A}, w_{2,B}, w_{2,C}, w_{3,A}, w_{3,B}, w_{3,C})$ be a fixed point and σ_1, σ_2 be identity (without loss of generality). Then:
$w_{i,A}^{n}w_{i,B}^{n} = w_{i,A}^{n+1}w_{i,B}^{n+1} \Rightarrow P_n = P_{n+1} \Rightarrow a_{i,2}^{AB,n} = a_{i,2}^{AB,n+1}$. This means x satisfies all equations given by $a_{i,2}$. Since x is a fixed point $(w_{i,A}, w_{i,B})_{i=1,2,3}$ has to be a fixed point of the iterative procedure described in Proposition 57. Since $a_{i,1}^{C,n} = a_{i,1}^{C,n+1}$, this means $(w_{i,A}, w_{i,B})_{i=1,2,3}$ also solve $a_{i,1}^{A,n} + a_{i,1}^{B,n} + a_{i,1}^{C,n} = a_{i,1}$ which shows that x solves the system. □

We implemented the iterative procedure described in Proposition 61 to give an example with $(3, 3)$ binding sites.

Example 62. *Let the molecule be described by $M = (G, W)$ with*

$$G = (g_1, g_2, g_3, g_A, g_B, g_C) = (4 \cdot 10^3, 10, 2 \cdot 10^3, 500, 10^3, 10)$$

$$W = (w_{i,j})_{i,j=1,2,3,A,B,C} = \begin{pmatrix} 1 & 0.001 & 0.01 & 10 & 10 & 1000 \\ 0.001 & 1 & 0.05 & 100 & 1000 & 2000 \\ 0.01 & 0.05 & 1 & 100 & 100 & 1000 \\ 10 & 100 & 100 & 1 & 0.001 & 0.01 \\ 10 & 1000 & 100 & 0.001 & 1 & 0.05 \\ 1000 & 2000 & 1000 & 0.01 & 0.05 & 1 \end{pmatrix}$$

Then two decoupled molecules are given by $N = (G_{de}, W_N)$ and $K = (G_{de}, W_K)$ with

$$G_{de} = (g_1, g_2, g_3, g_A, g_B, g_C) = (5996.485, 13.51409, 4.936015 \cdot 10^4, 1509.304, 0.6932945, 2.389161 \cdot 10^{-3})$$

$$W_N = \begin{pmatrix} 1 & 1 & 1 & 46.94515 & 514.7757 & 153.93242 \\ 1 & 1 & 1 & 180.15859 & 395.1010 & 96.96765 \\ 1 & 1 & 1 & 1645.26622 & 341.0639 & 13.88127 \\ 46.94515 & 180.15859 & 1645.26622 & 1 & 1 & 1 \\ 514.7757 & 395.1010 & 341.0639 & 1 & 1 & 1 \\ 153.93242 & 96.9676 & 13.88127 & 1 & 1 & 1 \end{pmatrix}$$

$$W_K = \begin{pmatrix} 1 & 1 & 1 & 46.94539 & 514.6135 & 2.654036 \cdot 10^{-5} \\ 1 & 1 & 1 & 180.15843 & 395.1741 & 23184.93 \\ 1 & 1 & 1 & 1645.25945 & 341.1083 & 336723.6 \\ 46.94539 & 180.15843 & 1645.25945 & 1 & 1 & 1 \\ 514.6135 & 395.1741 & 341.1083 & 1 & 1 & 1 \\ 2.654036 \cdot 10^{-5} & 23184.93 & 336723.6 & 1 & 1 & 1 \end{pmatrix}$$

The titration curves of all individual sites of molecules M, N and K are illustrated in Fig. 4.5.

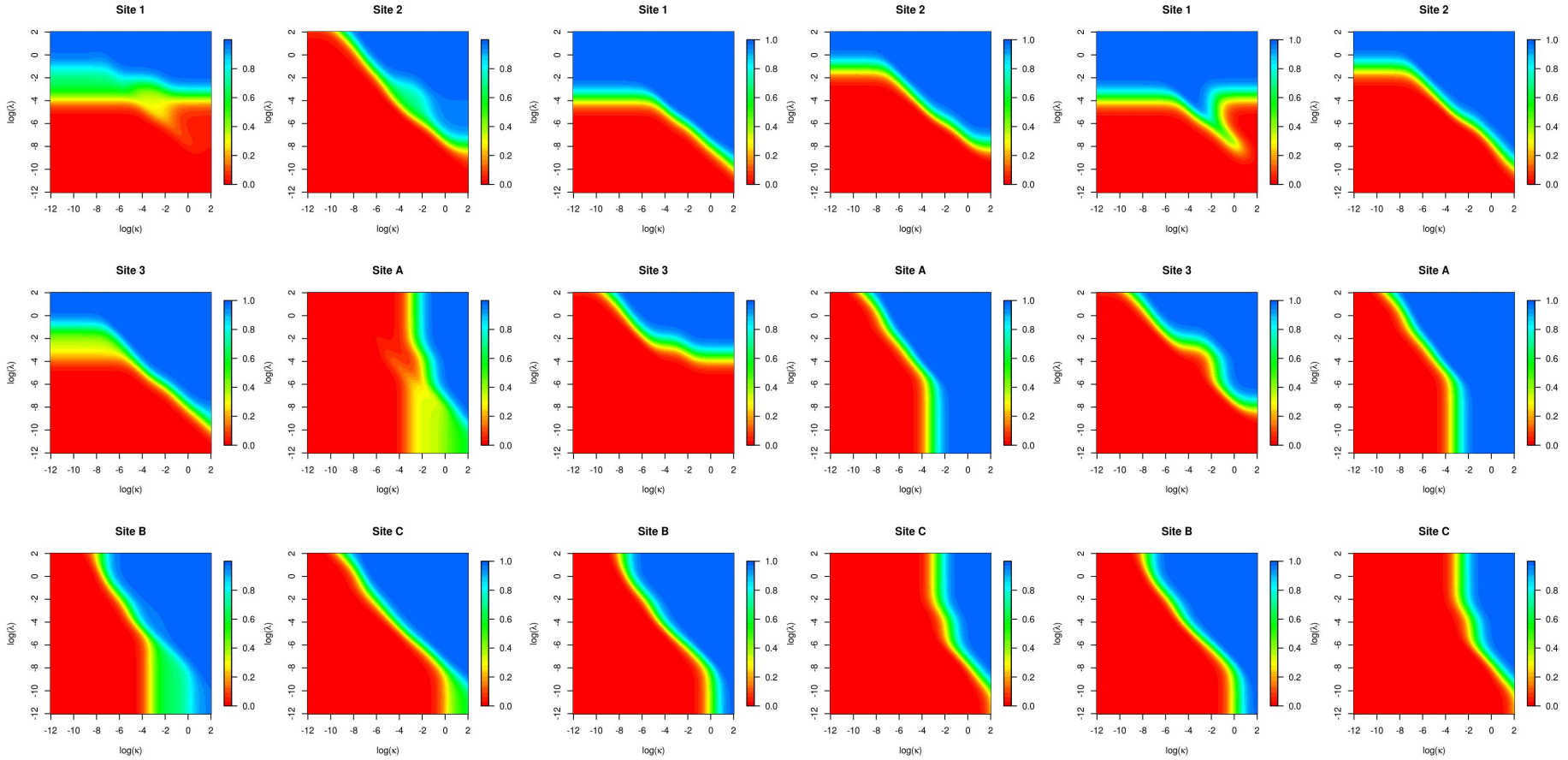

Figure 4.5: Activity dependent ligand binding to each site in the original molecule M (left pair of columns) and the decoupled molecules N (middle pair of columns) and K (right pair of columns) of Example 62. A logarithmic scale of the ligand activities λ and κ is used. The probability of occupation is encoded by the colors, according to the colorbars.

5 The Meaning of the Decoupled Sites Representation in Terms of Statistical Mechanics and Stochastics

The following chapter presents results of the paper Martini et al. (2013c). The text is in main parts adopted verbatim except for notational adaption to the previous chapters including altered mathematical symbols at several positions.

5.1 Motivation

In Chapter 3 and Chapter 4, the Decouples Sites Representation was discussed for one and two types of ligands. All results were based on an algebraic view on the theory. However, as presented in the introduction and in Chapter 2, the algebraic model is a result of the probabilistic setup of statistical mechanics and its concept of the Grand Canonical Partition Function. In this chapter, the DSR is reconsidered from a stochastic point of view. It closes the circle from the initial stochastic model, to an algebraic description in which the DSR was developed and analyzed, back to its meaning in statistical mechanics and stochastics. In this regard, we translate results in the periphery of the DSR, which were derived within the algebraic model, into stochastics. The shifted point of view facilitates some proofs and physical interpretations and provides the basis for future work which might investigate how certain phenomena of the algebraic concept can be interpreted stochastically.

5.2 One type of ligand

In the following, we investigate the stochastic features of the Decoupled Sites Representation for one type of ligands.

Remark 63. *In the previously described setup of Chapter 3 and Chapter 4, we allowed the binding and interaction constants to be complex valued. This made the use of the fundamental theorem of algebra possible and facilitated theory. However, a complex valued binding constant of a certain binding site translates into a "complex probability measure". This phenomenon and its possible physical interpretations will partly be discussed in Chapter 6. In this chapter, we assume that all molecules which we talk about have real, positive binding and interaction constants. Moreover, instead of regarding equivalence classes we will focus on a map of tuples on measures. This is, in the following section, of advantage as otherwise the equivalence relation has also to be transfered to the image space when certain maps are defined. This transfer would lead to a notation which is more complicated than necessary. However, all results can be directly transfered to the case with equivalence classes of tuples and of measures.*

As already has been described in Chapter 1, the basis for this work is the map

$$P : \mathbb{R}^{+m} \longrightarrow \mathcal{L}(K)(\lambda) \tag{5.1}$$

$$P(M)(\{k\},\lambda) = \frac{g(k)\lambda^{S(k)}}{\Phi(M)} \tag{5.2}$$

which maps a tuple to a family of distributions on the set of microstates K which is parameterized by $\lambda \in [0,\infty)$. $\mathcal{L}(K)(\lambda)$ denotes the set of functions $f : [0,\infty) \longrightarrow \mathcal{L}(K)$ mapping the activity λ onto a distribution on K and $S(k)$ the number of the occupied sites in state k ($S(k)$ is used instead of $|k|$ since, the function will be split in $S_1(k)$ and $S_2(k)$ in the case of two ligands).

Lemma 64. *The map P given by Eq. (5.2) mapping a tuple onto a family of measures is injective.*

Proof. Let M and N be two tuples and $P(M) = P(N)\ \forall k \in K, \forall\lambda$. Then

$$P(M)(\{0\}^n,\lambda) = \frac{1}{\Phi(M)} = \frac{1}{\Phi(N)} = P(N)(\{0\}^n,\lambda)$$

for every λ. This implies that $\Phi(M) = \Phi(N)$. Since the measures of a state k with only one site occupied shall be identical for every value of λ, we receive $g_i^M = g_i^N$. Consequently, with the same argument for states with two sites occupied, this gives identical interaction constants of M and N (We can continue this procedure for interaction constants of higher order, for the extended model). □

Since the map P is injective, we can use the letter M also for its image $P(M)$ and $M(\{k\})$ for $P(M)(\{k\})$ describing the probability of microstate k depending on λ. We will investigate which properties, the family of distributions of a tuple has. The random variables X_i will always denote the Bernoulli variables indicating whether site i is occupied ($X_i = 1$) or not ($X_i = 0$).

Proposition 65. *Let $M = (g_1,...,g_n,1,...,1)$ be a decoupled system, and let M_λ be the corresponding family of measures on the power set of K. Let $\{X_i\}_{1\le i\le n}$ be the Bernoulli variables which describe the occupation state of the individual sites. Then $\forall\lambda \in [0,\infty)$, $\{X_i\}_{1\le i\le n}$ are stochastically independent.*

Proof. Let $m_1+m_2 = n$ and let k_0 denote the corresponding microstate of the following event. Then

$$M_\lambda(X_{i_1} = ... = X_{i_{m_1}} = 0, X_{j_1} = ... = X_{j_{m_2}} = 1) = M_\lambda(\{k_0\}) =$$

$$\overset{\text{Eqs. (1.2,3.6)}}{=} \frac{\left(\prod_{l=1}^{m_2} g_{j_l}\right)\lambda^{m_2}}{\Phi(M)} = \frac{\prod_{l=1}^{m_2}(g_{j_l}\lambda)\prod_{l=1}^{m_1}1}{\Phi(M)} \overset{\text{Prop. 23}}{=} \frac{\prod_{l=1}^{m_2}(g_{j_l}\lambda)\prod_{l=1}^{m_1}1}{\prod_{i=1}^{n}(g_i\lambda+1)} =$$

$$\overset{\text{Prop. 26}}{=} \prod_{l=1}^{m_1} M_\lambda(X_{i_l}=0)\prod_{l=1}^{m_2} M_\lambda(X_{j_l}=1).$$

The Law of Total Probability shows that the probability can be factored also in the case of $m_1 + m_2 < n$. □

Proposition 65 shows that the natural intuition, saying that decoupled sites in the algebraic system correspond to stochastic independence of the Bernoulli variables in the stochastic setup, is correct. This gives the following view on the Decoupled Sites Representation for one type of ligand: A molecule M with n binding sites for the ligand is given. It corresponds to a family of measures on $\{0,1\}^n$ which is parametrized by the ligand activity λ. We look for a family of product measures N on $\{0,1\}^n$ such that $S(M) = S(N)$ for any choice of λ. $S(M)$ denotes the distribution of the function S with respect to the measure M on the domain.
Since we showed in Chapter 4 that not all interaction constants can be trivial and that the one-dimensional titration curves do not have to be of classical HH shape (Martini et al., 2013b), the question arises how a decoupled molecule with two different types of ligands can be interpreted from a stochastic point of view. We will investigate this phenomenon in the next section.

5.3 Two types of ligands

In Chapter 4, we showed that in the case of two different types of ligands and the constraint that both overall titration curves of a molecule M shall be preserved, there is not in general a molecule N, in which all interaction constants are trivial and which possesses the same overall titration curves (Example 40). Thus, we call a molecule with different types of ligands decoupled if all interaction constants of binding sites for the same type of ligand are equal to one (Definition 42). Moreover, it was shown that even though in a decoupled molecule, the binding sites for the same type of ligands do not interact directly, the one-dimensional titration curves, when the activity of the second ligand is fixed, are not in general of HH shape (Figures 4.2 and 4.3). In the following we will investigate how this non-HH shape can be interpreted stochastically and where HH titration curves are hidden.

In this chapter, the notation

$$k = (k_1, k_2) = (x_1^k, ..., x_{n_1}^k, x_{A_1}^k, ..., x_{A_{n_2}}^k)$$

will be used for the microstate of a molecule with n_1 binding sites for ligand L_1 and n_2 binding sites for ligand L_2. The lowercase letter x_i is used to indicate that it describes a value of the random variable X_i and k_1 and k_2 are used to subdivide k into microstates for the binding sites for the different ligands ($k_1 = (x_1^k, ..., x_{n_1}^k)$ and $k_2 = (x_{A_1}^k, ..., x_{A_{n_2}}^k)$).

All tuples in this section are assumed to have real valued, positive interaction and binding constants. Analogously to Subsection 5.2, we consider the map

$$P : \mathbb{R}^{+m} \longrightarrow \mathcal{L}(K)(\lambda, \kappa) \tag{5.3}$$

$$P(M)(\{k\}, \lambda, \kappa) = \frac{g(k)\lambda^{S_1(k)}\kappa^{S_2(k)}}{\Phi(M)}, \tag{5.4}$$

with $S_i(k)$ the number of bound ligands of type i in state k. We will again use the letter M as well for its image $P(M)$ to facilitate notation. Moreover, note that we defined the map P again on the set of tuples, without using the equivalence relation, to avoid transferring it to the set of measures. Using the equivalence relation would only lead

to a more complicated notation. Before we prove some general statements, we will give an illustrating example.

Example 66. *We choose a hypothetical decoupled molecule*

$$M = (g_1, g_2, g_A, w_{1,2}, w_{1,A}, w_{2,A}) = (2, \frac{3}{2}, 2, 1, \frac{3}{2}, \frac{8}{3})$$

with two binding sites for ligand L_1 and one binding site for ligand L_2. These binding and interaction energies were chosen as example in which the polynomial and the probabilities of the states are easy to calculate. For a temperature of $300°$ Kelvin these constants translate to the binding and interaction energies of $-(1.7, 1.0, 1.7, 0, 1, 2.5)$ in kJ/mol, since $g_i = exp(-\beta G_i)$ with $\beta = \frac{1}{RT}$ and T the absolute temperature in Kelvin and R the Boltzmann constant. We will not present the translation of constants into energies anymore, in the following examples. Thus, let X_1, X_2 be the Bernoulli variables describing the binding state of the sites 1 and 2. The binding polynomial of M is given by

$$\Phi(M) = 24\lambda^2\kappa + 3\lambda^2 + 14\lambda\kappa + 3.5\lambda + 2\kappa + 1.$$

Moreover,

$$M_{\lambda,\kappa}(X_1 = 1) = \frac{24\lambda^2\kappa + 3\lambda^2 + 6\lambda\kappa + 2\lambda}{\Phi(M)} \tag{5.5}$$

$$M_{\lambda,\kappa}(X_2 = 1) = \frac{24\lambda^2\kappa + 3\lambda^2 + 8\lambda\kappa + 1.5\lambda}{\Phi(M)} \tag{5.6}$$

$$M_{\lambda,\kappa}(X_1 = 1, X_2 = 1) = \frac{24\lambda^2\kappa + 3\lambda^2}{\Phi(M)} \tag{5.7}$$

For the choice $(\lambda, \kappa) = (1, 1)$ we receive

$$M_{1,1}(X_1 = 1) \cdot M_{1,1}(X_2 = 1) = \frac{35}{47.5} \cdot \frac{36.5}{47.5} \neq \frac{27}{47.5} = M_{1,1}(X_1 = 1, X_2 = 1) \tag{5.8}$$

which shows that the random variables X_1 and X_2 are not stochastically independent for all choices of (λ, κ). The situation changes if we consider the conditional distribution on microstates $\{k \in K | k_2 = 0\}$ (distribution on the microstates in which site A is unoccupied) or on $\{k \in K | k_2 = 1\}$ (distribution on the microstates in which site A is occupied):
Let $M_{\lambda,\kappa}(\cdot | k_2 = i)$ denote the conditional distribution on $\{k \in K | k_2 = i\}$. A conditional binding polynomial of M is given by

$$\Phi(M)_{|k_2=0} = 3\lambda^2 + 3.5\lambda + 1.$$

and thus

$$M_{\lambda,\kappa}(X_1 = 1 | k_2 = 0) = \frac{3\lambda^2 + 2\lambda}{\Phi(M)_{|k_2=0}} \tag{5.9}$$

$$M_{\lambda,\kappa}(X_2 = 1 | k_2 = 0) = \frac{3\lambda^2 + 1.5\lambda}{\Phi(M)_{|k_2=0}} \tag{5.10}$$

$$M_{\lambda,\kappa}(X_1 = 1, X_2 = 1 | k_2 = 0) = \frac{3\lambda^2}{\Phi(M)_{|k_2=0}} \tag{5.11}$$

which demonstrates independence of X_1 and X_2 with respect to the family of conditional distributions for any choice of (λ, κ):

$$M_{\lambda,\kappa}(X_1 = 1|k_2 = 0) \cdot M_{\lambda,\kappa}(X_2 = 1|k_2 = 0) = \frac{(3\lambda^2 + 2\lambda)(3\lambda^2 + 1.5\lambda)}{\Phi(M)^2_{|k_2=0}} =$$

$$= \frac{3\lambda^2\Phi(M)_{|k_2=0}}{\Phi(M)^2_{|k_2=0}} = \frac{3\lambda^2}{\Phi(M)_{|k_2=0}} = M_{\lambda,\kappa}(X_1 = 1, X_2 = 1|k_2 = 0). \tag{5.12}$$

This result might be obvious as in a decoupled system the one-dimensional titration curves of an individual site is of HH shape if the activity κ of the second ligand equals zero. However, conditional stochastic independence is also given if the condition is changed to $k_2 = 1$:

$$\Phi(M)_{|k_2=1} = 24\lambda^2\kappa + 14\lambda\kappa + 2\kappa.$$

$$M_{\lambda,\kappa}(X_1 = 1|k_2 = 1) = \frac{24\lambda^2\kappa + 6\lambda\kappa}{\Phi(M)_{|k_2=1}} \tag{5.13}$$

$$M_{\lambda,\kappa}(X_2 = 1|k_2 = 1) = \frac{24\lambda^2\kappa + 8\lambda\kappa}{\Phi(M)_{|k_2=1}} \tag{5.14}$$

$$M_{\lambda,\kappa}(X_1 = 1, X_2 = 1|k_2 = 1) = \frac{24\lambda^2\kappa}{\Phi(M)_{|k_2=1}} \tag{5.15}$$

which gives

$$M_{\lambda,\kappa}(X_1 = 1|k_2 = 1) \cdot M_{\lambda,\kappa}(X_2 = 1|k_2 = 1) =$$
$$= M_{\lambda,\kappa}(X_1 = 1, X_2 = 1|k_2 = 1) \ \ \forall(\lambda, \kappa) \in [0, \infty)^2.$$

We will formulate the observation of Example 66 generally in Proposition 67.

Proposition 67. *Let M be a decoupled molecule with n_1 binding sites for ligand L_1 and n_2 binding sites for ligand L_2. Then the random variables $\{X_i\}_{i=1}^{n_1}$ are conditionally stochastically independent for every condition $k_2 = c$ with $c \in \{0, 1\}^{n_2}$. Moreover, $\forall\,(\lambda, \kappa) \in [0, \infty)^2$ and $1 \leq i \leq n_1$ a $g'_{c,i} \in \mathbb{R}^+$ exists such that*

$$M_{\lambda,\kappa}(X_i = 1|k_2 = c) = \frac{g'_{c,i}\lambda}{g'_{c,i}\lambda + 1}. \tag{5.16}$$

Proof. Let $k_2 = c \in \{0, 1\}^{n_2}$ describe the state

$$X_{A_{\sigma(1)}} = ... = X_{A_{\sigma(l)}} = 1 \text{ and } X_{A_{\sigma(l+1)}} = ... = X_{A_{\sigma(n_2)}} = 0$$

with σ a permutation of $\{1, ..., n_2\}$ and $l \leq n_2$. Let $K_{|k_2=c} := \{k \in K|k_2 = c\}$ and let $k \in K_{|k_2=c}$. Then

$$g(k) = \prod_{i=1}^{l} g_{A_{\sigma(i)}} \cdot \prod_{j=1}^{n_1} \left(g_j \prod_{i=1}^{l} w_{j,A_{\sigma(i)}} \right)^{x_j^k} \tag{5.17}$$

and, due to the definition of conditional probability:

$$M_{\lambda,\kappa}(X_i = 1|k_2 = c) = \frac{\sum\limits_{k \in K_{|k_2=c}, x_i^k = 1} g(k)\lambda^{S_1(k)}\kappa^l}{\sum\limits_{k \in K_{|k_2=c}} g(k)\lambda^{S_1(k)}\kappa^l} =$$

$$\overset{\text{Eq. (5.17)}}{=} \frac{\sum_{k\in K_{|k_2=c}, x_i^k=1} \left(\prod_{j=1}^{n_1} \left(g_j \prod_{i=1}^{l} w_{j,A_{\sigma(i)}} \right)^{x_j^k} \lambda^{S_1(k)} \right)}{\sum_{k\in K_{|k_2=c}} \left(\prod_{j=1}^{n_1} \left(g_j \prod_{i=1}^{l} w_{j,A_{\sigma(i)}} \right)^{x_j^k} \lambda^{S_1(k)} \right)}$$

The last term equals the description of the titration curve of site i in a decoupled molecule with only n_1 binding sites for one type of ligand and binding constants

$$g'_{c,j} := g_j \prod_{i=1}^{l} w_{j,A_{\sigma(i)}}. \tag{5.18}$$

Eq. (5.18) proves the statements (and allows to calculate $g'_{c,i}$). □

Remark 68. *In other words, Proposition 67 means that if we use the condition that the second ligand occupied its binding sites according to a fixed microstate, the complex of the decoupled molecule and the bound molecules of the second ligand can be regarded as a new molecule with different binding constants, but with independent sites. Note here that the condition necessarily has to be that strict (microstate of the second ligand). A relaxation is not possible, if independence of the sites for ligand L_1 shall be guaranteed.*

Corollary 69. *The two-dimensional titration curve of a certain site of a decoupled molecule is a parameterized convex combination of one-dimensional HH curves.*

Proof.

$$M_{\lambda,\kappa}(X_i = 1) = \sum_{c\in\{0,1\}^{n_2}} M_{\lambda,\kappa}(X_i = 1|k_2 = c) M_{\lambda,\kappa}(k_2 = c),$$

where $M_{\lambda,\kappa}(X_i = 1|k_2 = c)$ is a HH curve, according to Proposition 67, and

$$\sum_{c\in\{0,1\}^{n_2}} M_{\lambda,\kappa}(k_2 = c) = 1.$$

□

We will illustrate the statement of Corollary 69 with an example.

Example 70. *Let $M = (g_1, g_2, g_A, w_{1,2}, w_{1,A}, w_{2,A}) = (2, \frac{3}{2}, 2, 1, \frac{3}{2}, \frac{8}{3})$ be the decoupled molecule of Example 66. Then*

$$M_{\lambda,\kappa}(X_1 = 1) =$$

$$= M_{\lambda,\kappa}(X_1 = 1|k_2 = 0) M_{\lambda,\kappa}(k_2 = 0) + M_{\lambda,\kappa}(X_1 = 1|k_2 = 1) M_{\lambda,\kappa}(k_2 = 1). \tag{5.19}$$

Note that, since in this example, there is only one binding site for ligand L_2, $k_2 \in \{0, 1\}$ and thus $M_{\lambda,\kappa}(k_2 = 1) = M_{\lambda,\kappa}(X_A = 1)$ is the titration curve of site A and $M_{\lambda,\kappa}(k_2 = 0) = 1 - M_{\lambda,\kappa}(k_2 = 1)$. Thus, the titration curve of site 1 is a convex combination of two HH curves weighted by the curve of site A. To calculate (5.19) we need to know the distribution of X_A:

$$M_{\lambda,\kappa}(X_A = 0) = \frac{\overbrace{3\lambda^2 + 3.5\lambda + 1}^{=\Phi(M)_{|k_2=0}}}{\Phi(M)} \qquad M_{\lambda,\kappa}(X_A = 1) = \frac{\overbrace{24\lambda^2\kappa + 14\lambda\kappa + 2\kappa}^{=\Phi(M)_{|k_2=1}}}{\Phi(M)} \tag{5.20}$$

and consequently

$$M_{\lambda,\kappa}(X_1 = 1) = \frac{3\lambda^2 + 2\lambda}{\Phi(M)_{|k_2=0}} \frac{\Phi(M)_{|k_2=0}}{\Phi(M)} + \frac{24\lambda^2\kappa + 6\lambda\kappa}{\Phi(M)_{|k_2=1}} \frac{\Phi(M)_{|k_2=1}}{\Phi(M)} =$$

$$= \frac{24\lambda^2\kappa + 3\lambda^2 + 6\lambda\kappa + 2\lambda}{\Phi(M)} = (5.5).$$

The curves $M_{\lambda,\kappa}(X_1 = 1|k_2 = 0)$, $M_{\lambda,\kappa}(X_1 = 1|k_2 = 1)$, $M_{\lambda,\kappa}(k_2 = 1)$ as well as the convex combination $M_{\lambda,\kappa}(X_1 = 1)$ are illustrated in Fig 5.1.

The previous results draw the following picture: The algebraic decoupling of molecules with two different types of ligands corresponds to finding a conditionally stochastically independent system with the same overall titration curves. In detail, this means that, for a given family of measures M on $\{0,1\}^{n_1+n_2}$, we look for a family of measures N such that all conditional measures $N_{|k_2}$ are product measures on $\{0,1\}^{n_1}$ for every $k_2 \in K_2$, $N_{|k_1}$ are product measures for every $k_1 \in K_1$ and $S_1(M) = S_1(N), S_2(M) = S_2(N)$. Compared to the setup with only one type of ligand we have an additional constraint since the function S was split into two parts. This constraint makes it impossible to find a family of product measures on $\{0,1\}^{n_1+n_2}$ for every given M. Consequently, the constraint of being a product measure is relaxed. However, the weakening of the constraints leads to the existence of several different solutions. A naturally arising question is which features the different distributions of different decoupled molecules share. We will compare the decoupled molecule of Examples 66, 70 with the other decoupled molecule sharing the same overall titration curves.

Example 71. *Let $M = (g_1, g_2, g_A, w_{1,2}, w_{1,A}, w_{2,A}) = (2, \frac{3}{2}, 2, 1, \frac{3}{2}, \frac{8}{3})$ be the molecule of Example 66 and let $N = (2, \frac{3}{2}, 2, 1, 2, 2)$ be the second decoupled molecule with the same bp (the maximal number of decoupled systems is 2, according to Corollary 47). Then*

$$N_{\lambda,\kappa}(X_i = 1|k_2 = 0) = M_{\lambda,\kappa}(X_i = 1|k_2 = 0),\ i = 1, 2$$

$$N_{\lambda,\kappa}(X_1 = 1|k_2 = 1) = M_{\lambda,\kappa}(X_2 = 1|k_2 = 1)$$

and

$$N_{\lambda,\kappa}(X_2 = 1|k_2 = 1) = M_{\lambda,\kappa}(X_1 = 1|k_2 = 1)$$

These equations mean that the decoupled molecules M and N share the same conditional HH titration curves. However they are permuted in the case of $k_2 = 1$.

We can formulate this observation as proposition.

Proposition 72. *Let M and N be two different decoupled molecules with n_1 binding sites for ligand L_1 and n_2 binding sites for ligand L_2 sharing the same binding polynomial. Moreover, let the order of the binding sites be equal, that is $g_i^M = g_i^N\ \forall i \in \{1, .., ., A_{n_2}\}$. Then the following statements hold:*

a) $N_{\lambda,\kappa}(X_i = 1|k_2 = \{0\}^{n_2}) = M_{\lambda,\kappa}(X_i = 1|k_2 = \{0\}^{n_2}) \quad \forall \lambda, \kappa \in [0, \infty)$

b) A permutation σ of $\{1, ..., n_1\}$ exists such that

$$N_{\lambda,\kappa}(X_i = 1|k_2 = \{1\}^{n_2}) = M_{\lambda,\kappa}(X_{\sigma(i)} = 1|k_2 = \{1\}^{n_2})$$

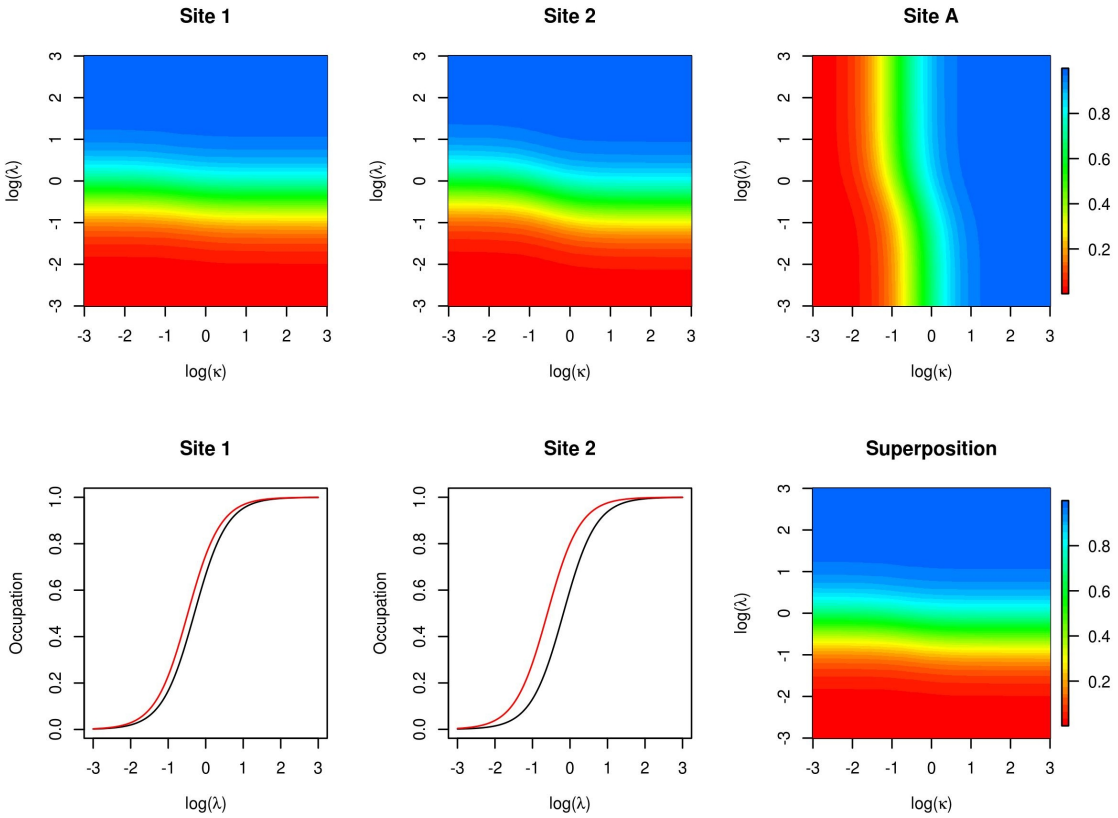

Figure 5.1: First row: Activity dependent ligand binding to each site of the tuple M of Example 70. Logarithmic scales of the activities of the ligands are used. The probability of a site being occupied is encoded by colors, according to the color bar on the right side of the figure. Second row: Conditional binding curves of site 1 and 2 for the conditions "site A is unoccupied" (black line) and "site A is occupied" (red line). The superposition represents the titration curve of site 1, however was calculated using the conditional HH curves and the titration curve of site A: $M(X_1 = 1) = M(X_1 = 1|k_2 = 0)M(k_2 = 0) + M(X_1 = 1|k_2 = 1)M(k_2 = 1)$.

Proof. a) According to Eq. (5.18) $g'_{i,c} = g_i$, and $g_i^M = g_i^N$ since the order of the sites is assumed to be fixed.

b) Again, Eq. (5.18) proves the statement, since the products

$$g_j \prod_{i=1}^{n_2} w_{j,A_{\sigma(i)}}$$

have to solve a subsystem of equations given by the coefficients a_{i,n_2} of the bp and consequently are permutations of each other, since they correspond to the roots of a polynomial (Proposition 23).

□

To complete our illustrations, we give a final example of a system with two binding sites for each ligand. This example shows, how to deal with the weights of the superposition in the case of more than one binding site for both ligands.

Example 73. *Let the decoupled tuple* $M = (g_1, g_2, g_A, g_B, w_{1,2}, w_{1,A}, w_{1,B}, w_{2,A}, w_{2,B}, w_{A,B})$ $= (2, 16, 4, 8, 1, 8, 4, \frac{1}{16}, \frac{1}{32}, 1)$ *be given. Its binding polynomial is*

$$\Phi(M) = 64\lambda^2\kappa^2 + 96\lambda^2\kappa + 32\lambda^2 + 2049\lambda\kappa^2 + 136\lambda\kappa + 18\lambda + 32\kappa^2 + 12\kappa + 1$$

We calculate the conditional HH curves of site 1, exemplarily.

$$M(X_1 = 1|k_2 = (0,0)) = \frac{32\lambda^2 + 2\lambda}{32\lambda^2 + 18\lambda + 1} = \frac{2\lambda}{2\lambda + 1} \tag{5.21}$$

$$M(X_1 = 1|k_2 = (0,1)) = \frac{32\lambda^2\kappa + 64\lambda\kappa}{32\lambda^2\kappa + 68\lambda\kappa +} = \frac{4\lambda^2 + 8\lambda}{4\lambda^2 + 8.5\lambda + 1} = \frac{8\lambda}{8\lambda + 1} \tag{5.22}$$

$$M(X_1 = 1|k_2 = (1,0)) = \frac{64\lambda^2\kappa + 64\lambda\kappa}{64\lambda^2\kappa + 68\lambda\kappa + 4\kappa} = \frac{16\lambda^2 + 16\lambda}{16\lambda^2 + 17\lambda + 1} = \frac{16\lambda}{16\lambda + 1} \tag{5.23}$$

$$M(X_1 = 1|k_2 = (1,1)) = \frac{64\lambda^2\kappa^2 + 2048\lambda\kappa^2}{64\lambda^2\kappa^2 + 2049\lambda\kappa^2 + 32\kappa^2} = \frac{64\lambda}{64\lambda + 1} \tag{5.24}$$

The corresponding weights for the superposition are given by:

$$M(k_2 = (0,0)) = \frac{32\lambda^2 + 18\lambda + 1}{\Phi(M)} \tag{5.25}$$

$$M(k_2 = (0,1)) = \frac{32\lambda^2\kappa + 68\lambda\kappa + 8\kappa}{\Phi(M)} \tag{5.26}$$

$$M(k_2 = (1,0)) = \frac{64\lambda^2\kappa + 68\lambda\kappa + 4\kappa}{\Phi(M)} \tag{5.27}$$

$$M(k_2 = (1,1)) = \frac{64\lambda^2\kappa^2 + 2049\lambda\kappa^2 + 32\kappa^2}{\Phi(M)} \tag{5.28}$$

Thus, we receive the following representation of $M(X_1 = 1)$ *with Eqs.* $(5.21 - 5.28)$*:*

$$M(X_1 = 1) = (5.21) \cdot (5.25) + (5.22) \cdot (5.26) + (5.23) \cdot (5.27) + (5.24) \cdot (5.28) \tag{5.29}$$

The binding curves of the individual sites as well as the HH curves and the corresponding weights are illustrated in Fig. 5.2.

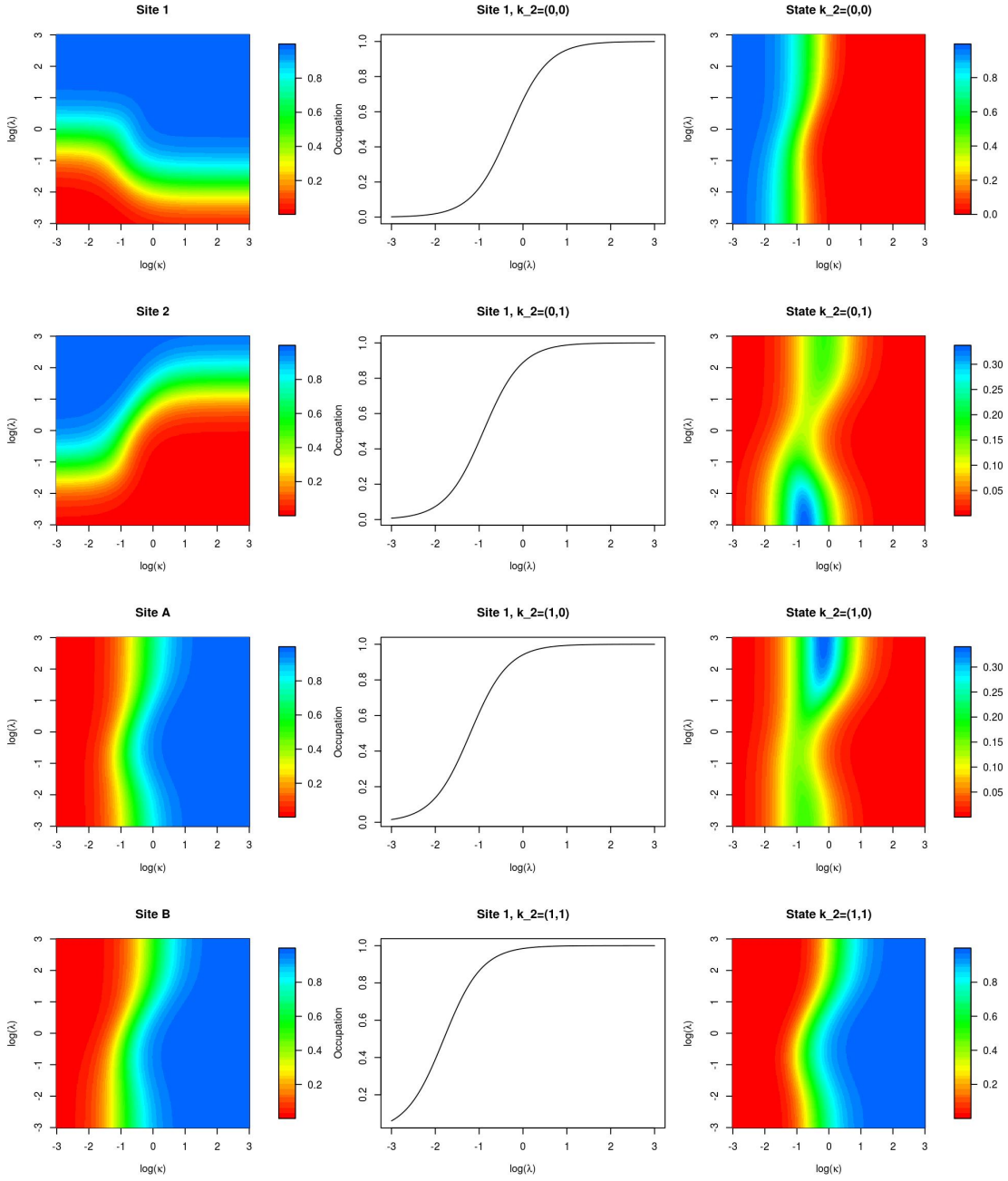

Figure 5.2: First column: Activity dependent ligand binding to each site of the tuple M of Example 73 . Logarithmic scales of the activities of the ligands are used. The probability of a site being occupied is encoded by colors, according to the color bar on the right side of each image. Second column: Conditional binding curves of site 1 for the different microstates of ligand L_2 ($k_2 \in \{(0,0),(0,1),(1,0),(1,1)\}$). Third column: probabilities of the different microstates of the binding sites for the second ligand.

Remark 74. *In a decoupled molecule M with only one type of ligand, the following calculation rule for the binding constants of a microstate $k = (x_1^k, ..., x_n^k)$ is a direct result of stochastic independence of the binding sites:*

$$M_\lambda(k) = M_\lambda(x_1^k, ..., x_n^k) = \prod_{i=1}^{n} \left(M_\lambda(X_i = 1)^{x_i^k} \cdot M_\lambda(X_i = 0)^{1-x_i^k} \right). \tag{5.30}$$

The sum of all probabilities of microstates belonging to the same macrostate gives the probability of the macrostate

$$M_\lambda(S(k) = i) = \sum_{\{k|S(k)=i\}} M_\lambda(k) = \frac{a_i \lambda^i}{\Phi(M)} \tag{5.31}$$

which the decoupled molecule obviously shares with every molecule with the same binding polynomial, since the coefficient a_i of the polynomial as well as the polynomial $\Phi(M)$ are identical (the coefficient a_i is given by the sum of all constants of the microstates with macrostate i). These equations can be transfered to the case of a decoupled molecule N binding two ligands. For a microstate $k = (k_1, k_2) = (x_1^k, ..., x_{n_1}^k, ..., x_{A_{n_2}}^k)$, Eqs. (5.30-5.31) translate to:

$$N_{\lambda,\kappa}(k) = N_{\lambda,\kappa}(x_1^k, ..., x_{A_{n_2}}^k) =$$

$$= N_{\lambda,\kappa}(k_2) \prod_{i=1}^{n_1} \left(N_{\lambda,\kappa}(X_i = 1|k_2)^{x_i^k} \cdot N_{\lambda,\kappa}(X_i = 0|k_2)^{1-x_i^k} \right) \tag{5.32}$$

and

$$N_{\lambda,\kappa}(S_1(k) = i, S_2(k) = j) = \sum_{\{k|S(k)=(i,j)\}} N_{\lambda,\kappa}(k) = \frac{a_{i,j} \lambda^i \kappa^j}{\Phi(M)} \tag{5.33}$$

Due to the symmetric role of the two ligands Eq. (5.32) can also be rewritten with k_1 instead of k_2 (and $i \in \{A_1, ..., A_{n_2}\}$). The coefficient $a_{i,j}$ is given by the sum of all constants of microstates belonging to macrostate (i, j).

6 Complex Measures and Cooperativity

Definition 75, Example 79, Lemma 86, Corollary 87 and Example 90 have already been presented in Martini and Ullmann (2013) and have been adopted verbatim.

6.1 Motivation

In Chapter 3, it was shown that –to use the DSR as general tool– we need to extend the domain of the energies to a stripe within the complex numbers. In this chapter, some ideas about how complex binding constants in a decoupled system can be interpreted are presented. In particular, it is discussed what complex binding constants can mean for the original system. The main idea presented in this chapter is to relate complex roots of the binding polynomial to positive cooperative binding of the ligand to the original molecule. This idea is not totally new, and has already been used in Onufriev and Ullmann (2004), Ullmann and Ullmann (2011) and Martini and Ullmann (2013). The intention of this chapter is to investigate the ideas more rigorously, as far as possible. Since the presentation and discussion of this phenomenon also includes the question whether a polynomial has complex roots, it will not be possible to treat this topic satisfactory. However, this knowledge about the difficulty of the nature of the problem shall not prevent us from trying to solve it. We will forthwith see that the problem starts with the definition of cooperativity.

6.2 Cooperativity

Cooperativity is a term which is used to describe the fact that different binding sites of a molecule can interact. The most stressed example is the target molecule hemoglobine with O_2 as a ligand: Positive cooperativity between the four binding sites is what makes this molecule an effective transporter from a region with a higher O_2 partial pressure to regions where it is consumed. Cooperativity was the main topic in many publications (e.g. Ben-Naim (2001); Hunter and Anderson (2009); Ullmann and Ullmann (2011)) and has recently become of importance in gene regulation and cell biology (e.g. Gutierrez et al. (2012); Sugiyama et al. (2013)). However, in most publications it is hardly mentioned that there are two different definitions which are not equivalent but which are often presented together, without distinguishing between them explicitly (e.g. see Stefan and Le Novère, 2013). The first definition is based on the change of affinity of a binding site to its ligand if another site becomes occupied. The second definition, based on the Hill coefficient is often described as a "measure" of cooperativity, but should be seen as another definition. In the following, we will consider both definitions but with different vocabulary, and investigate their relations. All binding and interaction constants describing a molecule are regarded as a positive real number. Complex numbers, with imaginary part nonzero are only considered, if a binding polynomial is factored.

Definition 75. *Let $M = (g_1, ..., g_n, w_{1,2}, ..., w_{n-1,n})$ be a molecule. The ligand is said to bind*

- *positive-cooperatively to sites i and j if and only if $w_{i,j} > 1$.*
- *non-cooperatively to sites i and j if and only if $w_{i,j} = 1$.*
- *negative-cooperatively to sites i and j if and only if $w_{i,j} < 1$.*

Thus, positive-cooperative binding to site i and j means that binding of the ligand to site i increases the affinity of the ligand to site j and vice versa. This definition is commonly used (Berg et al. (2007); Ben-Naim (2001); Ullmann and Ullmann (2011)) and is adequate to characterize biochemical regulatory mechanisms by pointing out whether a ligand enhances or represses the binding to other sites. The second definition requires some preliminary: The following definition is based on Hill (1910, 1913).

Definition 76 (Hill Plot and Hill Coefficient)**.** *Let Ψ be an overall titration curve of a molecule M with n binding sites. Then the function*

$$H(\log(\lambda)) := \log\left(\frac{\Psi}{n - \Psi}\right) \tag{6.1}$$

as a function of $\log(\lambda)$ is called **Hill Plot** *of the overall titration curve. Its slope at the point of $n/2$ saturation is called the* **Hill Coefficient** *η_H of this curve.*

Remark 77. *In literature, Eq. (6.1) is frequently also written as*

$$H(\log(\lambda)) := \log\left(\frac{\tilde{\Psi}}{1 - \tilde{\Psi}}\right).$$

This coincides with the presented version, since $\tilde{\Psi}$ is usually used as normed version of Ψ as it is used in this work ($\tilde{\Psi} = \Psi/n$).

Definition 76 gives the second definition of cooperativity:

Definition 78. *Let $M = (g_1, ..., g_n, w_{1,2}, ..., w_{n-1,n})$ be a molecule. The ligand is said to bind*

- *Hill-positive-cooperatively to the molecule if $\eta_H > 1$.*
- *Hill-non-cooperatively to the molecule if $\eta_H = 1$.*
- *Hill-negative-cooperatively to the molecule if $\eta_H < 1$.*

Regarding Definition 76 and based on the previous chapters, it is not obvious, why the Hill Coefficient should be related to cooperativity in any way. Nevertheless, we will see that it is a useful tool: Comparing different Henderson-Hasselbalch titration curves, we will see, if we plot them as functions of $\log(\lambda)$ that their shapes are identical. Except for the pK_a-value, there is nothing to distinguish between them. In particular, the slope at half saturation is always the same. Having the example of hemoglobin in mind, we think of a molecule to have a certain type of cooperativity if the difference in the chemical activity $\Delta\lambda$, which is required to transit from a low value of $\mathbb{E}_\lambda S$ to a high expectation $\mathbb{E}_{\lambda+\Delta\lambda} S$ is low. Descriptively, this should be the case, since as soon

as a ligand binds, the binding to the other sites will be enhanced which induces the binding at the other sites. Thus, in a relatively small interval of the activity scale, a mixed population of microstates will be present. At most values of activity, nearly the whole the population of hemoglobin molecules will be fully unoccupied or, at a higher activity, fully occupied. Or as Hunter and Anderson (2009) describes cooperativity in the first headline: "It's All or Nothing". This concept of cooperativity creates the idea that we can measure cooperativity as a slope of $\Psi(\log(\lambda)$ and we compare it to the slope of a Henderson-Hasselbalch titration curve. If the slope of an overall curve is greater to that of a Henderson-Hasselbalch curve, this should be caused by an interaction of that kind, that the binding of the ligand to a site enhances the binding to the other sites. Consequently, we need a characteristic of the slope of a titration curve which we compare to that of a Henderson-Hasselbalch titration curve. If possible, the charateristic should be normed: $\eta_{HH} = 1$. This is what the Hill Coefficient satisfies: The nonnegative function $\Psi(\log(\lambda))$ as a function of $\log(\lambda)$ increases if and only if $n - \Psi(\log(\lambda))$ decreases and consequently if and only if $\frac{\Psi}{n-\Psi}$ increases. This is still true if we apply the monotone function log to the equation. Moreover, for a Henderson-Hasselbalch curve, we receive:

$$\log\left(\frac{\Psi}{1-\Psi}\right) = \log\left(g10^{\log(\lambda)}\right) = \log(g) + \log(\lambda). \tag{6.2}$$

As a function of $\log(\lambda)$, Eq. (6.2) is linear with slope 1, which means in particular that the Hill Coefficient of a Henderson-Hasselbalch curve μ_{HH} equals 1. Thinking about a molecule with several binding sites, which all are energetically identical $g_i = g_j$ and without interaction, the overall titration curve will be of Henderson-Hasselbalch shape. If all $g_i = g_j$ and the binding sites cooperate positively, the Hill Coefficient should be greater than 1. For the case of two binding sites, this result is presented in Proposition 89.

However, a problem with Definition 78 is obvious: It does not distinguish between the individual binding sites. Even though the Hill Coefficient is used as a "measure" for cooperativity it is impossible to represent the interaction between the binding sites if they are not homogeneous. A discrimination between the sites in its initial application for hemoglobin was not necessary but it is of importance if other target molecules are investigated. An idea to extract more information from the overall titration curve could be the extension of the concept of the Hill Coefficient by investigating the slope at the points of $(1, 2, ..., n-1)$ to describe the interaction of each site with the rest of the system. However, even this extension does not solve the main problem of the Hill Coefficient: Even in the case of only two binding sites, a measure which is based on the overall titration curve can not distinguish between the different types of cooperativity of Definition 75, which is illustrated by Example 79:

Example 79. *The following molecules are examples of positive-, negative-, or non-cooperative ligand binding, but share the same overall titration curve:*

$$(2, 1, 0.5), \quad (1.5 + \sqrt{1.25}, 1.5 - \sqrt{1.25}, 1), \quad \left(0.1, 2.9, \frac{100}{29}\right).$$

Since we want to relate the different types of cooperativity with complex roots of the binding polynomial, we give a third definition

Definition 80. *Let M be molecule. M is said to bind its ligand* **DSR-cooperatively** *if $\Phi(M)$ has a complex root with imaginary part nonzero.*

In the following, we will present and discuss some properties of derivatives of the binding polynomial and the overall titration curve. Moreover, we will rewrite the Hill Coefficient in a different form and investigate the relation between the different types of cooperativity in the special case of $n = 2$, afterwards. We use the notation

$$f' := \frac{\mathrm{d}}{\mathrm{d}\log(\lambda)} f.$$

All derivatives are understood with respect to $\log(\lambda)$.
Calculating the derivative shows that the Hill coefficient is a linear function of the derivative of the titration curve Ψ at $\Psi = \frac{n}{2}$.

Lemma 81.

$$\left(\log\left(\frac{\Psi}{n-\Psi}\right)\right)' = \frac{1}{\ln(10)} \frac{n\Psi'}{-\Psi^2 + n\Psi}. \tag{6.3}$$

Thus, the Hill Coefficient can be rewritten as

$$\eta_H = \left(\log\left(\frac{\Psi}{n-\Psi}\right)\right)'_{|\Psi=\frac{n}{2}} = \frac{4}{n\ln(10)} \Psi'_{|\Psi=\frac{n}{2}}. \tag{6.4}$$

Moreover, the overall titration curve also writes

Lemma 82.

$$\Psi = \frac{\Phi'}{\ln(10)\Phi}. \tag{6.5}$$

Proof. The derivative of a monomial $a_i\lambda^i$ is:

$$(a_i\lambda^i)' = \ln(10) i a_i \lambda^i.$$

The relation of the binding polynomial Φ and the titration curve (Proposition 19) proves the statement. □

For $\Psi = \frac{n}{2}$, Ψ' can be expressed as

Lemma 83.

$$\Psi'_{|\Psi=\frac{n}{2}} = \left(\frac{\Phi''}{\ln(10)\Phi}\right)_{|\Psi=\frac{n}{2}} - \frac{n^2\ln(10)}{4} \tag{6.6}$$

Proof. Eq. (6.5) gives

$$\Psi' = \frac{\Phi''\Phi - \Phi'\Phi'}{\ln(10)\Phi^2} = \frac{\Phi''}{\ln(10)\Phi} - \frac{\Phi'\Phi'}{\ln(10)\Phi^2}, \tag{6.7}$$

which implies the statement, since $\Psi = \frac{\Phi'}{\ln(10)\Phi} = \frac{n}{2}$ is the point, where the function is evaluated. □

The following simple statement will turn out to be useful:

Lemma 84. *Let $\Phi = a_n\lambda^n + a_{n-1}\lambda^{n-1} + a_{n-2}\lambda^{n-2} + ... + 1$ be a binding polynomial. Let a family of polynomials be defined by*

$$\Phi_t(\lambda) := \Phi(t\lambda) \text{ with } t \in (0, \infty).$$

Then Φ has only real roots if and only if $\Phi_t(\lambda)$ has only real roots for every $t \in (0, \infty)$.

Proof. r is a root of Φ if and only if $\frac{r}{t}$ is a root of Φ_t. □

Remark 85. *Lemma 84 is not only a useful tool (which will be used in Proposition 89), but has also a biophysical interpretation of an altered reference solution with a different chemical activity.*

6.3 Two binding sites

Lemma 86. *A molecule described by $(g_1, g_2, w_{1,2}) \in \mathbb{R}^{+3}$ requires the use of complex binding constants with a nonzero imaginary part to be presented as decoupled system if and only if*

$$(g_1 + g_2)^2 < 4g_1g_2w_{1,2}. \tag{6.8}$$

Proof.

$$\Phi(g_1, g_2, w_{1,2}) = g_1g_2w_{1,2}\lambda^2 + (g_1 + g_2)\lambda + 1$$

A system $(d, e, 1)$ without interaction has to solve the equations:

$$de = g_1g_2w_{1,2} \quad \text{and} \quad d + e = g_1 + g_2.$$

Thus, d has to solve

$$d^2 - (g_1 + g_2)d + g_1g_2w_{1,2} = 0$$

which shows that $d \notin \mathbb{R}$ if and only if $(g_1 + g_2)^2 < 4g_1g_2w_{1,2}$. □

The following corollary illustrates that complex binding energies with nonzero imaginary part are an indicator for negative interaction energies ($w_{1,2} > 1$) in the original molecule in the case of two binding sites.

Corollary 87 (Repulsion)**.** *Let $(g_1, g_2, w_{1,2}) \in \mathbb{R}^{+3}$ be a molecule with two ligand binding sites and interaction constant $w_{1,2} \leq 1$. Then the binding constants of the corresponding decoupled system are real.*

Proof.

$$w_{1,2} \leq 1 \Longrightarrow g_1^2 + (2 - 4w_{1,2})g_1g_2 + g_2^2 \geq (g_1 - g_2)^2 \geq 0.$$

This implies

$$(g_1 + g_2)^2 \geq 4g_1g_2w_{1,2}$$

which proves the statement due to Lemma 86. □

Corollary 87 states, that repulsion ($w_{1,2} \leq 1$) guarantees real roots in the decoupled system, in case of a molecule with two binding sites.

Lemma 88 (Hill Coefficient for a molecule with two binding sites)**.** *For a molecule $M = (g_1, g_2, w_{1,2})$ with two binding sites, the Hill Coefficient is given by*

$$\mu_H = 4 \left(\frac{g_1 g_2 w_{1,2} \lambda^2}{\Phi} \right)_{|\Psi=1}. \tag{6.9}$$

Proof. Using Eqs. (6.4) and (6.6) gives (with $n = 2$)

$$\eta_H = \left(\log \left(\frac{\Psi}{2 - \Psi} \right) \right)'_{|\Psi=1} = \frac{4}{2\ln(10)} \left(\left(\frac{\Phi''}{\ln(10)\Phi} \right)_{|\Psi=1} - \frac{4\ln(10)}{4} \right) =$$

$$= \frac{2}{\ln(10)} \left(\left(\frac{\ln(10)^2 (4 g_1 g_2 w_{1,2} \lambda^2 + (g_1 + g_2)\lambda)}{\ln(10)\Phi} \right)_{|\Psi=1} - \ln(10) \right) =$$

$$= \frac{2}{\ln(10)} \left(\ln(10) \left(1 + \frac{2 g_1 g_2 w_{1,2} \lambda^2)}{\Phi} \right)_{|\Psi=1} - \ln(10) \right) = \left(\frac{4 g_1 g_2 w_{1,2} \lambda^2}{\Phi} \right)_{|\Psi=1}$$

□

For the relation of the different types of cooperativity, we receive the following implications for the case of a molecule with two binding sites:

Proposition 89. *Let $M = (g_1, g_2, w_{1,2})$ be a molecule. Then the following implications are valid:*

i) If M is DSR-cooperative then it binds its ligand positive-cooperatively.

ii) If $g_1 = g_2$ then M binds its ligand positive-cooperatively if and only if it binds its ligand DSR-cooperatively.

iii) M binds its ligand DSR-cooperatively if and only if it binds its ligand Hill-positive-cooperatively.

Proof. i) Corollary 87.

ii) Lemma 86 with $g_1 = g_2$.

iii) To show iii) we use Lemma 88 and show that

$$\left(\frac{4 g_1 g_2 w_{1,2} \lambda^2}{\Phi} \right)_{|\Psi=1} > 1 \Leftrightarrow \Phi \text{ has complex roots with nonzero imaginary part.}$$

For this, let t denote the activity at which $\Psi = \mathbb{E}_t S = 1$. Let us regard the polynomial

$$\Phi_t(\lambda) = g_1 g_2 w_{1,2} t^2 \lambda^2 + (g_1 + g_2) t\lambda + 1$$

as described in Lemma 84. We will show that

$$\left(\frac{4 g_1 g_2 w_{1,2} \lambda^2}{\Phi} \right)_{|\Psi=1} > 1 \Leftrightarrow \Phi_t \text{ has complex roots with nonzero imaginary part.}$$

The condition $\Psi = 1$ means, that the corresponding measure (since we translated the measure to t) at $\lambda = 1$ has expectation 1:

$$2\underbrace{\frac{g_1 g_2 w_{1,2} t^2}{\Phi_t(1)}}_{=:P_2} + \underbrace{\frac{(g_1+g_2)t}{\Phi_t(1)}}_{=:P_1} = 1,$$

with P_i denoting the probability of i sites being occupied, in the measure defined at $\Phi_t(1)$. Since $\Phi_t(1)$ is a constant, $\Phi_t(\lambda)$ shares its roots with the probability generating function of the measure of $\Phi_t(1)$

$$\frac{\Phi_t(\lambda)}{\Phi_t(1)} = P_2\lambda^2 + (1-2P_2)\lambda + \underbrace{(1-P_2-1+2P_2)}_{=P_2},$$

which has non-real roots if and only if

$$(1-2P_2)^2 < 4P_2^2 \Leftrightarrow 1 < 4P_2 = 4\left(\frac{g_1 g_2 w_{1,2}\lambda^2}{\Phi(\lambda)}\right)_{|\Psi=1}.$$

Here, the very right equation represents the Hill coefficient, according to Lemma 88. □

6.4 More binding sites

For the case of two binding sites, Section 6.3 showed that DSR-cooperativity is equivalent to Hill-positive-cooperativity. Moreover, both types of cooperativity imply the positive-cooperativity defined by $w_{1,2} > 1$. An interesting question is, which statements on the relations of the different types of cooperativity can be transfered to molecules with more than two binding site. Example 90 shows that Corollary 87 can not simply be transfered to the case of five binding sites.

Example 90. *Let us regard the molecule*

$$\Phi_1(M) = (g_1, g_2, g_3, g_4, g_5, w_{1,2}, w_{1,3}, w_{1,4}, w_{1,5}, w_{2,3}, w_{2,4}, w_{2,5}, w_{3,4}, w_{3,5}, w_{4,5}) =$$

$$= \left(2, 2, 2, 2, 2, \frac{1023}{1024}, \frac{1023}{1024}, \frac{1}{1024}, \frac{1023}{1024}, \frac{1023}{1024}, \frac{1}{1024}, \frac{1}{1024}, \frac{1}{1024}, \frac{1}{1024}, \frac{1}{1024}\right)$$

with positive interaction energies (repulsion). However, its bp has two complex roots with nonzero imaginary part which can be shown easily by an investigation of its extremes. This example demonstrates that even if repulsion is assumed, decoupling can require the use of complex binding energies.

As explained at the beginning of this section, the Hill coefficient is related to the idea that positive-cooperative binding leads to a family of distributions in which a small change in ligand activity is sufficient to transit from a distribution in which the main part of the molecule population is fully unoccupied to a distribution with the main part of the molecules fully occupied. Thus, we can also investigate directly the slope of Ψ as a function of $\log(\lambda)$: Instead of Eq. (6.1), we investigate the slope of

$\Psi(\log(\lambda))$. This can be used to calculate the Hill coefficient with Eq. (6.4) and has the advantage of being linear due to the linearity of the differential operator:

$$\Psi' = \left(\sum_{i=1}^{n} \Psi_i\right)' = \sum_{i=1}^{n} \Psi_i' \tag{6.10}$$

To see to which value of $\Psi'_{|\Psi=\frac{n}{2}}$ a Hill coefficient of 1 corresponds to, we write down the following lemma.

Lemma 91. *Let Ψ be a HH curve. Then $\Psi'_{\Psi=\frac{1}{2}} = \frac{\ln(10)}{4}$.*

Proof. Either use that $\eta_{HH} = 1$ and Eq. (6.4) or calculate the derivative of $\frac{g\lambda}{g\lambda+1}$ and use $\frac{g\lambda}{g\lambda+1} = \frac{1}{2} \Leftrightarrow g\lambda = 1$. □

Lemma 92. *Let Ψ be a HH curve with bp $g\lambda + 1$ and let $\mathrm{Var}_\lambda(X)$ denote the activity dependent variance of the Bernoulli variable X, describing whether site i is occupied or not. Then $\Psi' = \ln(10)\mathrm{Var}_\lambda X$. Moreover, $\mathrm{Var}_\lambda(X) = \frac{g\lambda}{(g\lambda+1)^2}$.*

Proof. $\Psi' = \ln(10)\frac{g\lambda(g\lambda+1)-g^2\lambda^2}{(g\lambda+1)^2} = \ln(10) \underbrace{\frac{g\lambda}{g\lambda+1}}_{=\mathbb{E}_\lambda(X)=\mathbb{E}_\lambda(X^2)} - \ln(10) \underbrace{\frac{(g\lambda)^2}{(g\lambda+1)^2}}_{=\mathbb{E}(X)^2}$. □

Proposition 93. *Let Ψ be an overall titration curve and let S denote the random variable describing the number of occupied sites. Then $\Psi' = \ln(10)\mathrm{Var}_\lambda(S)$.*

Proof. Eq. (6.7) gives the statement, since this equation is the variance of S (except for the factor $\ln(10)$). □

Remark 94. *We already have mentioned that the parameterization of the family of distributions that we deal with are very special and that from the expected values we can calculate the whole family of distributions of S. Here, another remarkable property was shown: The variance of a fixed distribution for a certain activity λ is related to the change of the expected value of the family if the activity is changed.*

Corollary 95. *Let Φ be a binding polynomial of degree n with only real roots, let $\{X_i\}_{i=1,\dots,n}$ be the corresponding decoupled system and let Ψ be the corresponding overall titration curve. Then*

$$\Psi' \leq \ln(10)\frac{n}{4}, \tag{6.11}$$

which is equivalent to $\mathrm{Var}_\lambda(S) \leq \frac{n}{4}$. If all binding constants of the decoupled system are identical, we get $\Psi'_{|\Psi=\frac{n}{2}} = \ln(10)\frac{n}{4}$, which means the Hill coefficient is 1.

Proof. $\mathrm{Var}(X_i) = p - p^2 \leq \frac{1}{4}\ \forall p \in [0,1]$. The maximum of $\mathrm{Var}(X_i)$ is $\frac{1}{4}$. If all binding constants are identical, the curves will have their maximal value at the activity at which $\Psi = \frac{n}{2}$. □

Corollary 95 gives a criterion to decide whether a polynomial has complex roots.

Example 96. *Let us regard the polynomial* $\Phi = \lambda^2 + \lambda + 1$. *The variance of the family of distributions is given by*

$$\mathrm{Var}_\lambda(S) = \frac{4\lambda^2+\lambda}{\lambda^2+\lambda+1} - \left(\frac{2\lambda^2+\lambda}{\lambda^2+\lambda+1}\right)^2 =$$

$$= \frac{\lambda^3+4\lambda^2+\lambda}{(\lambda^2+\lambda+1)^2}.$$

For $\lambda = 1$ *this gives* $\frac{2}{3} \geq \frac{1}{2}$. *Consequently* Φ *has to have complex valued roots with nonzero imaginary part, according to Corollary 95.*
Without having checked which calculation rules can be transfered from probability measures to complex valued Bernoulli measures, we will treat these measures as if they were "usual" probability measures. We will see that the formula for the variance of the sum of independent variables is true for this example. Calculating the roots of the binding polynomial shows that S *can be interpreted as the sum of the independent Bernoulli variables* X_1, X_2 *with*

$$P(X_1 = 1) = \frac{(0.5 + \frac{\sqrt{3}}{2}i)\lambda}{(0.5 + \frac{\sqrt{3}}{2}i)\lambda + 1}$$

and

$$P(X_2 = 1) = \frac{(0.5 - \frac{\sqrt{3}}{2}i)\lambda}{(0.5 - \frac{\sqrt{3}}{2}i)\lambda + 1}.$$

The variance of X_1 *is given by*

$$\mathrm{Var}_\lambda(X_1) = \frac{(0.5 + \frac{\sqrt{3}}{2}i)\lambda}{(0.5 + \frac{\sqrt{3}}{2}i)\lambda + 1} - \left(\frac{(0.5 + \frac{\sqrt{3}}{2}i)\lambda}{(0.5 + \frac{\sqrt{3}}{2}i)\lambda + 1}\right)^2 =$$

$$= \frac{(0.5 + \frac{\sqrt{3}}{2}i)\lambda}{((0.5 + \frac{\sqrt{3}}{2}i)\lambda + 1)^2}.$$

Calculating $\mathrm{Var}_\lambda(X_2)$, *shows that* $\mathrm{Var}_\lambda(S) = \mathrm{Var}_\lambda(X_1) + \mathrm{Var}_\lambda(X_2)$ *for these complex distributions.*

The next proposition shall demonstrate that we can use the well-known calculation rules for variances of independent variables, as well for the variances of complex normed measures for the Bernoulli variables of the decoupled system.

Proposition 97. *Let* Φ *be a binding polynomial of degree* n *and let* $\{X_i\}_{i=1,\dots,n}$ *be the corresponding decoupled system which may have complex Bernoulli distributions. Moreover, let* Ψ *be the corresponding overall titration curve. Then*

$$\mathrm{Var}(S) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Proof.

$$\ln(10)\mathrm{Var}(S) = \Psi' = \left(\sum_{i=1}^{n} \Psi_i\right)' = \sum_{i=1}^{n} \Psi_i' \overset{\text{HH, Lemma 92}}{=} \sum_{i=1}^{n} \ln(10)\mathrm{Var}(X_i).$$

□

6.5 Complex normed measures in probability theory

The emergence of complex-valued normed measures in the Decoupled Sites Representation is not surprising, since the Bernoulli measures correspond to the roots of a polynomial. Already if we only consider single discrete distributions on $\{0, ..., n\}$ and not parameterized families it is known that the roots of the probability generating function give independent Bernoulli variables whose sum has the original distribution. However, discrete measures give not the only motivation leading to questions on **extended probabilities** or **quasi probabilities**. The topic has been discussed in several publications in the context of quantum mechanics and splitting certain processes into independent parts (e.g. Burgin and Meissner, 2012; Cox, 1955; Feynman, 1987; Hartle, 2008; Hofmann, 2009; Mückenheim et al., 1986; Parisi, 1983; Zak, 1998). Thus, extended probabilities seem to provide a way of dealing with certain kinds of dependencies of variables in a formal stochastic independent setup. We can consider this concept of extended probabilities in two different ways in probability theory:

- As a quantity telling us something about the original system, without investigating complex-valued normed measures as an own separate object in probability theory.
- As an own topic in probability theory in which dependencies of variables can be encoded in complex-valued normed measures with formal independent variables.

In this chapter, extended probabilities were only described as an indicator for cooperativity in the original system, but were neither investigated deeply from a measure theoretic or probabilistic point of view, nor was a possible philosophic meaning discussed. From a measure theoretic point of view, extended probabilities should not pose a lot of problems. Obviously, Kolmogorov's first axiom of the measure being real-valued and positive will be violated, but his second and third axioms

$$P(\Omega) = 1 \qquad \text{and} \qquad P(A_1 \cup A_2 \cup ...) = \sum_{i=1}^{\infty} P(A_i) \tag{6.12}$$

for disjunct events $(A_i)_{i\in\mathbb{N}}$ are valid for extended probabilities, as well. Moreover, Kolmogorov's axioms are already presented in an abstract form and the interpretation of the "meaning" of probability defined by Kolmogorov's axioms is more or less given by the Law of Large numbers. Thus, an extension to complex-valued normed measures should not be a fundamental problem. However, some of the well-known consequences of the axioms will not be valid, since the first axiom is violated. Thus, one of the first consequences of the axioms –the monotonicity– cannot be derived if complex-valued measures are considered. A detailed analysis which laws of probability theory can be transfered to extended probabilities is required. This analysis has not yet been done in this work, due to the lack of a motivating example from theoretical probability theory. Building up theory on extended probabilities including rewriting the main results of probability theory is not satisfying if no benefit for the theory can be realized. Thus, a first task might be to identify applications of a formal independence of variables to certain theoretical questions and to relate complex-valued probabilities to certain kinds of stochastic dependencies. If there are examples in which the use of extended probabilities is of advantage, rewriting the basics of probability theory with extended probability measures might be valuable.

7 Modeling Regulatory And Sensing Processes With Titration Curves

In this chapter, two examples will be presented of how titration curves can be used to model biological phenomena. The first example is an approach of modeling a two-component gene regulatory system of bacteria. Modeling gene expression regulatory systems mathematically is a topic of current research and several different mathematical approaches have been used to model different gene regulatory switches or networks. Among the different approaches deterministic models based on titration curves or on differential equations (e.g. Brown, 2010; Buchler and Louis, 2008; Grima, 2010; Karlsson et al., 2007; Kuttler and Hense, 2008) can be found as well as stochastic approaches which were in particular used to investigate the propagation of noise in signals in biological information processing systems theoretically (e.g. Kærn et al., 2005; Monteoliva et al., 2013; Thattai and van Oudenaarden, 2002; Walczak et al., 2009). Here, a model for a two-component gene regulatory system shall be presented which is based on titration curves. It has the advantage of simple equations, but the limitations of a model whose basic assumptions (large number of the molecules involved in signal transduction, stochastic independence of certain processes) might be violated in a cell. Moreover, another disadvantage of this model is that it only focuses on the stationary equilibrium states and that it does not consider the dynamics of the system in nonequilibrium. Nevertheless, the model should be appropriate to investigate some properties of the respective modeled gene regulatory systems.
The second example is a model for the olfactory sensing of insects. A difference to the gene regulation model is that only one binding step is considered (the binding of an odorant to a receptor), however with different ligands competing for one and the same binding site. This situation is also given in several other biologically or medically relevant situations involving receptors and several competing drugs.

7.1 Modeling a two-component system

7.1.1 Motivation

In a fluctuating environment, the survival of organisms strongly depends on their ability to adapt quickly to new conditions. To respond to changes in an appropriate way, environmental conditions have to be sensed and the information has to be transmitted to a gene regulatory level to express the genes for required gene products such as special enzymes, accordingly. Different signal transduction and gene regulatory systems with different signal processing properties have evolved for different situations. A special type of gene regulatory system is the so-called two component system of eubacteria. It consists of a sensing enzyme, usually a **sensor-kinase**, which is connected to a receptor on the surface of the cell. If the receptor binds a signaling-molecule, the sensor-kinase becomes active and phosphorylates the **response regulator**, which is the second

component of the system. The response regulator can only bind to a certain DNA sequence to start transcription if it is phosphorylated. Therefore, the phosphorylated form of the response regulator shall be called the active form. Thus, the overall signal transfer consists of three steps: The ligand binding of the signal to the receptor, the activation of the response regulator and the ligand binding of the activated response regulator to the DNA binding site. We will model the first and the third step with ligand binding curves.

7.1.2 The model

To model the response of cells to a signal transmitted by a two component system we split the regulatory system into three parts, which we model separately and which we combine afterwards:

- The number of activated sensor-kinase molecules
- The number of activated response regulator molecules
- The transcriptional activation of target genes

In the following we will present the assumptions which we make to model each part of the system.

The number of activated sensor-kinase molecules

We model the activation of the kinase molecules in the following way: We assume that each kinase molecule has its own receptor and that the kinase is active at a certain time point if and only if a ligand is bound to the corresponding receptor at the same time. Moreover, we assume that the activity of the kinase is not coregulated by a feedback mechanism, which might reduce its activity after a period of high gene expression level, but only by its receptor. Thus, this way of activating the kinase equals usual ligand binding theory, when a ligand binds reversibly to the target molecule. Assuming that the receptor has only one binding site, the probability of a receptor being occupied in thermodynamical equilibrium is described by the classical Henderson-Hasselbalch titration curve

$$\frac{g_i \lambda}{g_i \lambda + 1} \tag{7.1}$$

with λ denoting the chemical activity of the ligand and g_i a transformation of the free energy of the binding reaction which depends on temperature. Since the number of cells in a population and consequently the number of receptors is high, the number of active kinase molecules in the population should be close to

$$k \frac{g_i \lambda}{g_i \lambda + 1} \tag{7.2}$$

with k denoting the total number of kinase molecules in all cells (Law of Large Numbers, assumption of stochastic independent binding).

The number of activated response regulator molecules

We assume that if the number of activated sensor kinases increases, the number of activated response regulator will increase until a steady state will be reached, in which the number of deactivations equals the number of activations per unit time. We assume that both reactions – activation and deactivation– are catalyzed enzymatically by the sensor kinase and an antagonistic phosphatase. The substrate for the respective reaction is the deactivated or the activated form of the response regulator. Since we do not know anything about an allosteric regulation of the kinase, we assume a Michaelis-Menten kinetic for the activation as well as for the deactivation. Thus, these assumptions lead to the equation in which the difference between two Michaelis-Menten kinetics is zero (steady state):

$$0 = \frac{V_k(r_t - r_a)c}{K_k + (r_t - r_a)c} - \frac{V_p(r_a)c}{K_p + (r_a)c}\,, \tag{7.3}$$

with V_k, V_p denoting the maximum velocity of the enzyme catalyzed reaction (k=kinase, p=phosphatase), K_i the respective Michaelis constant, r the number of response regulator molecules (in **t**otal (r_t) or the number of **a**ctive molecules (r_a)) and c a factor to convert the number of molecules to concentration. Note that V_k is proportional to the number of active kinase molecules given by Eq. (7.2). Rewriting Eq. (7.3) shows, that the number of active response regulator molecules in steady state r_a is a root of the polynomial

$$P(r_a) = a_2 r_a^2 + a_1 r_a + a_0 \tag{7.4}$$

with

$$a_2 = (V_p - V_k)$$
$$a_1 = (r_t V_k - r_t V_p - V_k \frac{K_p}{c} - V_p \frac{K_k}{c})$$
$$a_0 = V_k r_t \frac{K_p}{c}.$$

Even though $P(r_a)$ is a polynomial of degree two, its roots are not obvious, since the coefficients depend on the other variables. We summarize the effects of a change of certain parameters on its roots in Proposition 98.

Proposition 98. *a) For any choice of $r_t, V_k, V_p, K_p, K_k, c > 0$ exactly one root of $P(r_a)$ is in the interval $(0, r_t)$.*

b) Let $V_{k,1} < V_{k,2}$. Let $P_1(r_a)$ and $P_2(r_a)$ denote the corresponding polynomials with identical $r_t, V_p, K_p, K_k, c > 0$ but with different maximal kinase activities $V_{k,1}, V_{k,2}$, respectively. Moreover, let $r_{a,i}$ denote the root of P_i in the interval $(0, r_t)$. Then,

$$r_{a,1} < r_{a,2}.$$

c) For $V_{k,n} \to \infty$ and all other parameters fixed, $r_{a,n} \to r_t$.

d) Let $r_{t,1} < r_{t,2}$. Let $P_1(r_a)$ and $P_2(r_a)$ denote the corresponding polynomials with identical $V_k, V_p, K_p, K_k, c > 0$ but with total number of response regulator $r_{t,1}, r_{t,2}$, respectively. Moreover, let $r_{a,i}$ denote the root of P_i in the interval $(0, r_{t,2})$. Then,

$$r_{a,1} < r_{a,2}.$$

Proof. a) Note that $P(0) = a_0 > 0$ and $P(r_t) = -V_p\frac{K_k}{c}r_t < 0$. Consequently, the Intermediate Value Theorem states that there is at least one root in the interval $[0, r_t]$. Since P is a polynomial of degree two, it has another real root. If $(V_p - V_k) < 0$, then P tends to $-\infty$ if $|r_a| \to \infty$ which implies the other root to be smaller than 0, again according to the Intermediate Value Theorem. If $(V_p - V_k) > 0$ the same arguments show, that the second root is greater than r_t. If $V_k = V_p$ only one root exists.

b) We regard the polynomial

$$P_{2-1} := P_2 - P_1 = (V_{k,2} - V_{k,1})(b_2 r_a^2 + b_1 r_a + b_0),$$

with

$$b_2 = -1,\ b_1 = (r_t - \frac{K_p}{c}),\ b_0 = r_t\frac{K_p}{c}$$

which gives the difference between the values of P_2 and P_1. Recall that $P_2(0) > P_1(0) > 0$. Thus, it is sufficient to show that $P_{2-1} > 0$ on the interval $[0, r_t)$, since this implies $r_{a,1} < r_{a,2}$. The roots of P_{2-1} are given by $r_{a,0} = -\frac{K_p}{c}$ and $\tilde{r}_{a,0} = r_t$. In the case of $V_{k,1} < V_{k,2}$, P_{2-1} tends to $-\infty$ if $|r_a| \to \infty$ which shows, that it is positive on $[0, r_t)$.

c) Let $V_{k,n} \to \infty$ be a sequence of positive numbers. Without loss of generality, we can assume the sequence to be monotone increasing. Let P_n denote the corresponding sequence of polynomials. Part b) implies that the sequence of roots $r_{a,n}$ ($\subset [0, r_t]$) is monotone increasing and bounded. Consequently, it will converge. What remains to be shown is that it will really converge to the upper bound r_t. We show that for every $\epsilon > 0$, an index n_ϵ exists such that $r_{a,n} \in (r_t - \epsilon, r_t]\ \forall n > n_\epsilon$. Let $\epsilon > 0$ be chosen arbitrarily. Then

$$\begin{aligned} P_n(r_t - \epsilon) &= a_2(r_t - \epsilon)^2 + a_1(r_t - \epsilon) + a_0 = \\ &= P_n(r_t) + (-2r_t\epsilon + \epsilon^2)a_2 - \epsilon a_1. \end{aligned} \tag{7.5}$$

We know that, $P_n(r_t) = -V_p\frac{K_k}{c}r_t < 0$. Moreover,

$$\begin{aligned} &(-2r_t\epsilon + \epsilon^2)a_2 - \epsilon a_1 = \\ &= \epsilon\left((-2r_t + \epsilon)(V_p - V_k) - (r_tV_k - r_tV_p - V_k\frac{K_p}{c} - V_p\frac{K_k}{c})\right) = \\ &= \epsilon\left((r_t - \epsilon + \frac{K_p}{c})V_k + (-r_t + \epsilon + \frac{K_k}{c})V_p\right). \end{aligned} \tag{7.6}$$

Since, for $\epsilon' < r_t + \frac{K_p}{c}$, the factor that V_k is multiplied with is positive, for V_k large enough Eq. (7.6) and Eq. (7.5) will be positive. This means, that from this value of V_k on, the roots will be in the interval $(r_t - \epsilon', r_t] \subset (r_t - \epsilon, r_t]$, due to the Intermediate Value Theorem ($P_n(r_t - \epsilon) > 0 > P_n(r_t)$).

d) As described, P_i denotes the polynomial with total number of response regulator $r_{t,i}$. All other parameters are identical for P_1 and P_2. Then $P_2(0) > P_1(0) > 0$ and

$$P_2 - P_1 = ((r_{t,2} - r_{t,1})V_k - (r_{t,2} - r_{t,1})V_p)r_a + (r_{t,2} - r_{t,1})V_k\frac{K_p}{c}.$$

If $V_k = V_p$, the graphs of the polynomials will not intersect, which means $P_2 > P_1$ on $\mathbb{R}$ and $r_{a,2} > r_{a,1}$. In the case of $V_k \neq V_p$, $P_2 - P_1$ has a single root at

$$r_{a,0} = -\frac{V_k \frac{K_p}{c}}{V_k - V_p}.$$

In the case of $V_k > V_p$, the root is negative and consequently $P_2 > P_1$ on $\mathbb{R}^+$, which gives the statement.
If $V_p > V_k$, then

$$P_2(r_{a,0}) = \frac{V_p V_k K_p K_k}{c^2(V_k - V_p)} < 0,$$

which implies that the roots of P_2 and P_1 are in a part of the domain on which $P_2 > P_1$. This gives the statement.

□

Proposition 98 shows that the polynomial resulting from Eq. (7.3) has the properties that a reasonable model should provide. Firstly, part a) shows, that the number of active response regulator, if all other variables are fixed, is unique, positive and smaller than the total number of response regulators. Secondly, part b) shows, that if all variables, except for the number of active kinases, are fixed, the number of active response regulator molecules will increase, if the number of active kinases is increased and vice versa. Thirdly, part c) illustrates that for a fixed number of response regulator molecules r_t and increasing kinase activity, r_a can be close to r_t. The last statement shows, that an increase of r_t leads to a higher number of active response regulator molecules r_a. We illustrate the effect of an altered V_k on the number of active response regulator r_a in Example 99.

Example 99. *As an example we set $V_k = aV_p$ for $a \in \mathbb{R}^+$, $V_p = 5$, $\frac{K_p}{c} = \frac{K_k}{c} = 10^5$ and $r_t = 10^6$. The number of active response regulator molecules r_a, dependent on a, is illustrated in Fig. 7.1.*

Example 99 shows how the number of active response regulator r_a depends on V_k, which is the target point of the signal from outside: the concentration of the signal molecule in the environment changes the number of active kinases and consequently V_k.

The transcriptional activation of target genes

For modeling the last part of the regulatory system we assume that a gene is active and transcribed with a certain rate if an active response regulator is bound to the DNA. In one cell, with only one binding site for the active response regulator at the DNA, this is a stochastic variable, but it can be described in a deterministic way using ligand binding curves of shape of Eq. (7.1), when the whole population is described (Law of Large Numbers, regarding cells as independent identically distributed gene expression random variables, fixed time point). Another possibility to justify the deterministic description in the case of regarding only one cell over time, is the assumption of ergodicity and consequently the convergence of the mean over time to the expected value of the variable for fixed time. Analogously to Eq. (7.2) we model the gene expression level by

$$G = G_{\max,\text{th}} \frac{g_2[r_a]}{g_2[r_a] + 1} \tag{7.7}$$

with G denoting the level of gene expression and $G_{\max,\text{th}}$ the theoretical maximal level of gene expression, if $\frac{g_2[r_a]}{g_2[r_a]+1} = 1$. Moreover, g_2 is a constant characterizing the binding of the active response regulator to the DNA and $[r_a]$ the concentration (or activity) of active response regulator. Note, that the practical maximal level of gene expression is lower, since it is also bounded by the condition $r_a < r_t$, with r_t fixed. The practical maximal level of gene expression within a cell is given by

$$G_{\max,\text{pr}} = G_{\max,\text{th}} \frac{g_2[r_t]}{g_2[r_t]+1} \tag{7.8}$$

Example 100. *To give an example of how the gene expression level may depend on the signal from outside, we extend Example 99. For this, we choose r_t to equal a concentration of 10^{-9} (the total number of 10^6 molecules translated into concentration in volume/volume). Moreover, we set $g_1 = 10^6, g_2 = 10^7$ and assume the maximum of V_k to be 20. This choice for the variables leads to the response curves presented in Fig. 7.1. Note that the gene expression level is not a ligand binding curve for a molecule with one binding site. In this example, the absolute slope of the gene response curve, dependent on the activity λ is relatively high. In this regard, this curve resembles an overall titration curve with more than one binding site and cooperative binding. This fact shows that the response regulator mediated signal can deviate qualitatively from Henderson-Hasselbalch curves.*

We calculated other examples, in which other quantities were changed. For instance, it is also possible to set the parameters in such a way that the concentration at which half maximal gene expression is reached differs significantly from the concentration at which half maximal number of kinase molecules is activated.

7.1.3 One enzyme for phosphorylation and dephosphorylation of the response regulator

In some two-component systems, the enzyme which is responsible for activation and for deactivation can be the same. Of course, an enzyme can always catalyze both directions, but here the situation will be considered that an activated kinase transfers a phosphoryl group from GTP to the response regulator and an deactivated sensor kinase removes the phosphoryl group from a response regulator producing inactivated response regulator and free phosphate. This situation might be true, for example, for the Gac/Rsm system (e.g. Brown, 2010). In this situation, we can rewrite Eq. (7.3) substituting

$$V_k = k\frac{g_i\lambda}{g_i\lambda+1} \text{ and } V_p = k\frac{1}{g_i\lambda+1}$$

and $K_k = K_p$, which gives

$$0 = \frac{k\frac{g_i\lambda}{g_i\lambda+1}(r_t - r_a)c}{K_k + (r_t - r_a)c} - \frac{k\frac{1}{g_i\lambda+1}(r_a)c}{K_p + (r_a)c}. \tag{7.9}$$

Eq. (7.9) directly shows that in this special case of the model, the concentration of the active response regulator is constant for varying total number of sensor kinase molecules, since $k > 0$ cancels out. Of course, in a model incorporating the dynamics, a system with reduced number of sensor kinase molecules will react more slowly. Assuming that

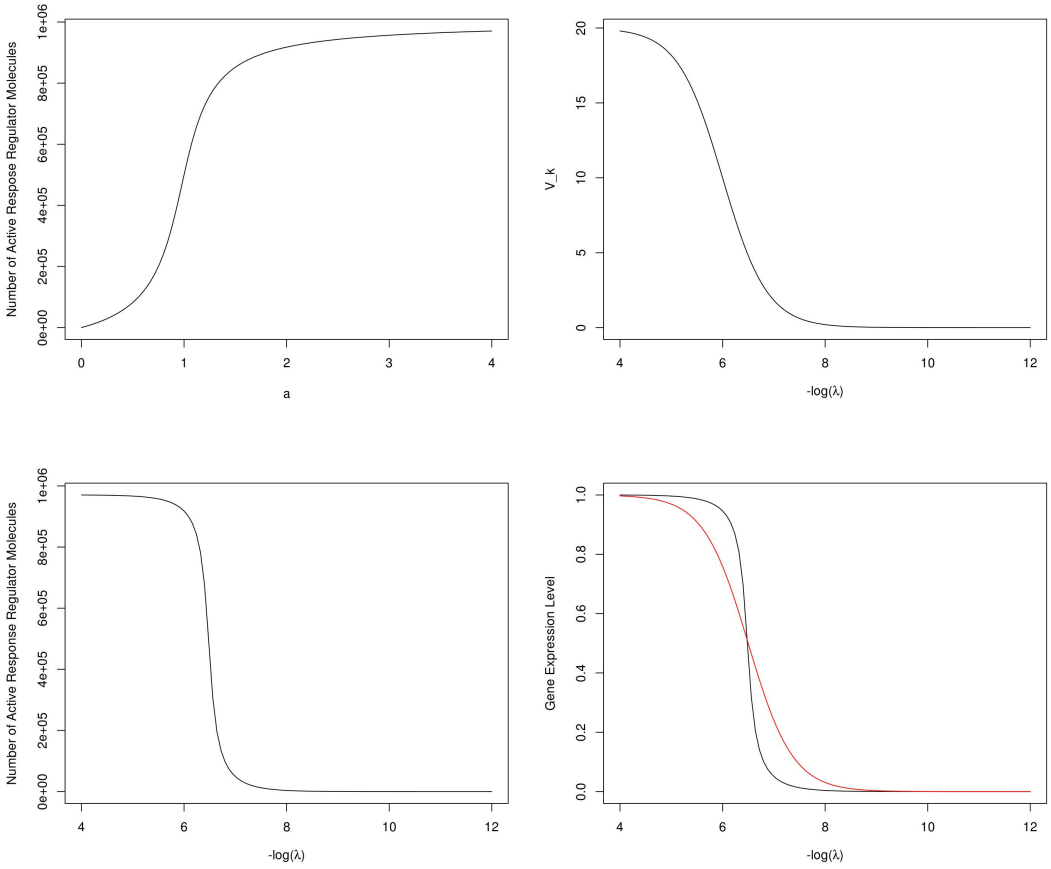

Figure 7.1: The number of active response regulator molecules r_a in steady state, as a function of the factor a describing the maximal velocity of the kinase compared to the deactivating phosphatase ($V_k = aV_p$). All other variables are fixed (as described in Example 99, $r_t = 10^6$). The other pictures illustrate the number of active kinase (V_k), the number of active response regulator r_a and G, relatively to $G_{\max,\mathrm{pr}}$, as a function of the activity of the signal molecule (negative logarithmic scale). Red line: A Henderson-Hasselbalch curve with identical position of the half maximum.

the Michaelis-constants for both directions are identical and abandoning the factor c by using concentrations $[r_i]$ instead of r_i, we receive the polynomial

$$\left(-\frac{g_i\lambda}{g_i\lambda+1}+\frac{1}{g_i\lambda+1}\right)[r_a]^2+\left(-K_k+\frac{g_i\lambda}{g_i\lambda+1}[r_t]-\frac{1}{g_i\lambda+1}[r_t]\right)r_a+\frac{g_i\lambda}{g_i\lambda+1}[r_t]K_k.$$

A signal transduction system in which the same enzyme catalyzes the phosphorylation as well as the dephosphorylation reaction (dependent on the respective binding state) should possess a higher sensitivity to a signal from outside than a signal transduction system with two separate enzymes for these reactions, since the signal simultaneously suppresses the phosphatase activity and enhances the kinase reaction.

7.2 Modeling olfactory sensing

7.2.1 Motivation

Since in olfactory sensing, the binding of an odor to a receptor plays the key role, we would like to model the electroantennographic response of beetles under the use of titration curves. In particular, we are interested in signals, caused by compositions of different odorants.

7.2.2 The model

We assume that the electric electroantennographic response is caused by ion channels which open if and only if a corresponding receptor is occupied by a signal molecule. The intensity of the signal corresponds to the number of occupied receptors. We will not incorporate feedback loops, which for example might close the ion channels after some time. Thus, the intensity of the signal can directly be translated to a number of occupied receptors and consequently can be described by a ligand binding curve. In this basic model, we will assume that each receptor has only one binding site.

7.2.3 One type of receptor with two different types of ligands

If olfactory sensing shall be modeled, of special interest is the situation in which a receptor can be stimulated by different substances. Thus, the ligand binding model has to be adapted. A receptor with one common binding site for two types of ligands can exist in three different states $0, A, B$, where 0 denotes the unoccupied state, A denotes the state with a bound molecule of L_1 and B denotes the state of being occupied by ligand L_2. In this first model, only one binding site exists which can be occupied exclusively by one of the ligands. As described in Chapter 2 and Chapter 3, the probabilities of being in a certain state depends on the relative energy levels of the state. Analogously to the previous chapters, λ, κ denote the activity of the ligands in the environment and g_i the binding constants. The probability of a receptor being occupied by any ligand is given by

$$\Psi(\lambda,\kappa) := \frac{g_1\lambda+g_2\kappa}{g_1\lambda+g_2\kappa+1}. \tag{7.10}$$

The fraction of receptors occupied by ligand L_1 is

$$\Psi_1(\lambda,\kappa) := \frac{g_1\lambda}{g_1\lambda+g_2\kappa+1}. \tag{7.11}$$

Analogously, for the fraction of receptors occupied by the second ligand L_2:

$$\Psi_2(\lambda, \kappa) := \frac{g_2\kappa}{g_1\lambda + g_2\kappa + 1}. \tag{7.12}$$

Obviously $\Psi = \Psi_1 + \Psi_2$. We will compare the signal which is generated by L_1 at activity λ_1 in the absence of L_2 to the signal increase caused by the addition of L_1 when L_2 is already present at activity κ. The signal increase is given by:

$$\Delta_1(\lambda, \kappa) := \Psi(\lambda, \kappa) - \Psi_2(0, \kappa) =$$

$$= \frac{g_1\lambda + g_2\kappa}{g_1\lambda + g_2\kappa + 1} - \frac{g_2\kappa}{g_2\kappa + 1} = \frac{g_1\lambda}{(g_1\lambda + g_2\kappa + 1)(g_2\kappa + 1)} \tag{7.13}$$

Proposition 101 (Signal attenuation by other ligands)**.** *In this model, the signal increase generated by L_1 at activity λ in the presence of a second ligand L_2 at activity κ, is smaller than the signal generated by L_1 at activity λ and L_2 at activity κ, which is smaller than the signal generated by L_1 at activity λ if a second ligand is absent:*

$$\Delta_1(\lambda, \kappa) < \Psi_1(\lambda, \kappa) < \Psi_1(\lambda, 0). \tag{7.14}$$

Proof. For $\kappa > 0$ and $g_2 > 0$, Eq. (7.14) states

$$\frac{g_1\lambda}{(g_1\lambda + g_2\kappa + 1)(g_2\kappa + 1)} < \frac{g_1\lambda}{g_1\lambda + g_2\kappa + 1} < \frac{g_1\lambda}{g_1\lambda + 1}$$

which is obviously true. □

We will give an example.

Example 102. *Eq. (7.10)-(7.13) are evaluated for the values $g_1 = 10^6, g_2 = 0.5 \cdot 10^6, \kappa = 10^{-5.5}$. The curves are illustrated in Fig. 7.2.*

7.2.4 Phenomenological description

Kosanke-Schütz et al. (2011) investigated the signal in nerves of the Colorado potato beetle when certain flavors were used as stimulus sequentially. The basic model presented in this section is not appropriate to describe this situation fully, since nerves have a mixed population of receptors and a receptor might have different, maybe overlapping binding sites. Thus, it is not surprising that some of the effects observed by Kosanke-Schütz et al. (2011) are impossible in the basic model. Additionally, one has to take into account that slightly different results of experiments can be regarded as being equal in practice. For instance, Type I interaction (Definition 103) cannot be observed in this model, even if we extend the model to a mixed population of different kinds of receptors, which will be discussed later. However, it can be observed in practice, if a certain difference of two amplitudes is not significant. In the following, some aspects of interacting odors will be discussed, based on the basic model, which will be extended afterwards. Kosanke-Schütz et al. (2011) divided interactions between different ligands into different types, which will be explained in the following. For this, they used 12 different substances, each in a fixed concentration as stimulus or background. Distinguishing between stimulus and background is important, since asymmetries in the interactions were observed if the role was commuted. Interaction always refers to an interaction between two ligands L_1, L_2 at certain fixed activities (λ_0, κ_0) with L_2 as background.

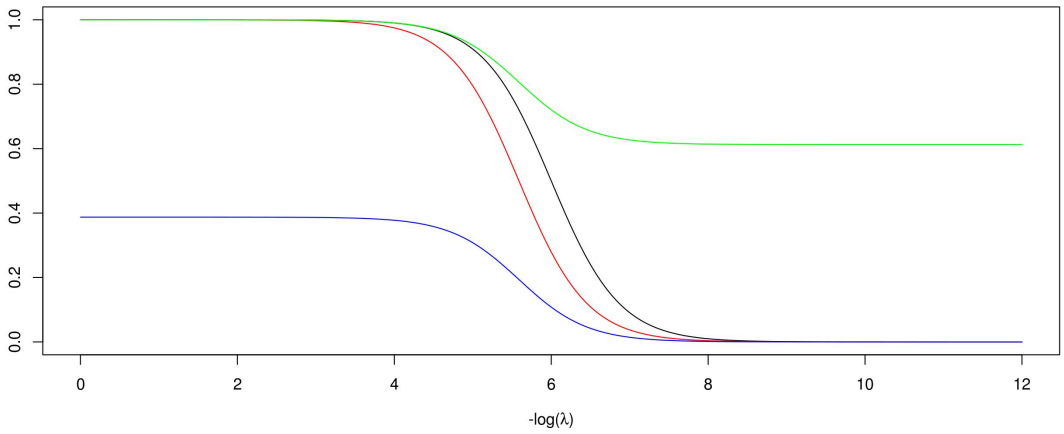

Figure 7.2: The signal caused by ligand L_1 as a function of λ in the absence of a second ligand ($\Psi_1(\lambda, 0)$, black curve) and for $\kappa = 10^{-5.5}$ ($\Psi_1(\lambda, 10^{-5.5})$, red curve). The common signal of both ligands is illustrated by the green curve ($\Psi(\lambda, 10^{-5.5})$) and the signal difference ($\Delta_1(\lambda, 10^{-5.5})$) is illustrated by the blue curve. The values $g_1 = 10^6, g_2 = 0.5 \cdot 10^6, \kappa = 10^{-5.5}$ were used.

Definition 103 (Type 1 Interaction)**.** *An interaction between two stimuli is called a Type 1 Interaction between stimulus L_1 and background L_2 at (λ_0, κ_0) if*

$$\Delta_1(\lambda_0, \kappa_0) = \Psi_1(\lambda_0, 0). \tag{7.15}$$

Obviously, Proposition 101 states that a Type 1 Interaction is impossible (also in mixed populations of receptors). However, the difference might not be detectable in an experiment if in a mixed population of receptors, for those receptors to which L_1 binds to g_2 is close to 0, which means that L_2 does not bind well to the receptors for L_1.

As expected from this model, in all other cases of interactions described by Kosanke-Schütz et al. (2011), the observed amplitude of the reaction to the test stimulus was reduced by the presence of a background. To distinguish between these interactions further, the interaction of a stimulus with itself as background was compared to the situation with the other substance as background. For this reason, we will use the notation $\Delta_1(\lambda_0, \lambda_0)$ for $\Psi_1(2\lambda_0, 0) - \Psi_1(\lambda_0, 0)$ to indicate that the same substance in the same concentration was used as a background.

Definition 104 (Type 3 Interaction)**.** *An interaction between two stimuli is called a Type 3 Interaction between stimulus L_1 and background L_2 at (λ_0, κ_0) if*

$$\Delta_1(\lambda_0, \kappa_0) = \Delta_1(\lambda_0, \lambda_0) \tag{7.16}$$

Lemma 105. *Eq. (7.16) holds if and only if*

$$g_2\kappa_0 = g_1\lambda_0.$$

Proof. Eq. (7.16) can be translated to

$$\frac{g_1\lambda_0}{(g_1\lambda_0 + g_2\kappa_0 + 1)(g_2\kappa_0 + 1)} = \frac{g_1\lambda_0}{(g_1\lambda_0 + g_1\lambda_0 + 1)(g_1\lambda_0 + 1)}$$

$$\begin{aligned}
&\iff (g_1\lambda_0 + g_2\kappa_0 + 1)(g_2\kappa_0 + 1) - (2g_1\lambda_0 + 1)(g_1\lambda_0 + 1) = 0 \\
&\iff (g_2\kappa_0)^2 + (g_1\lambda_0 + 2)g_2\kappa_0 + (g_1\lambda_0 + 1)(-2g_1\lambda_0) = 0 \\
&\iff g_2\kappa_0 = 0.5\left(-2 - g_1\lambda_0 \pm \sqrt{4 + 12g_1\lambda_0 + 9g_1^2\lambda_0^2}\right) \\
&\iff g_2\kappa_0 = 0.5\left(-2 - g_1\lambda_0 \pm (2 + 3g_1\lambda_0)\right) \\
&\overset{g_2\kappa_0 \geq 0}{\iff} g_2\kappa_0 = 0.5\left(-2 - g_1\lambda_0 + (2 + 3g_1\lambda_0)\right) \iff g_2\kappa_0 = a_1\lambda_0
\end{aligned}$$

□

Definition 106 (Type 2 Interaction)**.** *An interaction between two stimuli is called a Type 2 Interaction between stimulus L_1 and background L_2 at (λ_0, κ_0) if*

$$\Delta_1(\lambda_0, \kappa_0) > \Delta_1(\lambda_0, \lambda_0) \tag{7.17}$$

Lemma 107. *Eq. (7.17) holds if and only if*

$$g_2\kappa_0 < g_1\lambda_0.$$

Proof. Analogously to the proof of Lemma 105. □

Definition 108 (Type 4 Interaction)**.** *An interaction between two stimuli is called a Type 4 Interaction between stimulus L_1 and background L_2 at (λ_0, κ_0) if*

$$\Delta_1(\lambda_0, \kappa_0) < \Delta_1(\lambda_0, \lambda_0) \tag{7.18}$$

Lemma 109. *Eq. (7.18) holds if and only if*

$$g_2\kappa_0 > g_1\lambda_0.$$

Proof. Analogously to the proof of Lemma 105. □

What is conspicuous in this classification is, that it depends on the activities (λ_0, κ_0) how the interaction of L_1 and L_2 is classified. This means that a pair (L_1, L_2) which might show Type 2 interaction can also show Type 3 or 4 interaction at different concentrations.

7.2.5 Different types of receptors

Assuming that the overall signal measured by the electroantennographic recordings is a weighted sum of the signals $\Psi^{(1)}, ..., \Psi^{(n)}$ of different types of receptors, we receive

$$\Psi = \sum_{i=1}^{n} \omega_i \Psi^{(i)} \quad \text{with weights } \omega_i \text{ with } \sum_{i=1}^{n} \omega_i = 1. \tag{7.19}$$

Proposition 101 can be transfered directly to this extended situation since it is true for every summand. This also implies that absolute Type 1 interaction does neither exist in this extended model. However, as already mentioned, observed Type 1 interaction is an indicator for the spectra of receptors to which the respective ligand can bind, being more or less disjunct. The question arises how this concept can be transfered to

other measurable quantities. We think that properties of the (maybe further modified) function

$$\frac{\Psi(\lambda,0)-\Delta_1(\lambda,\kappa)}{\Psi(0,\kappa)}, \tag{7.20}$$

which describes how "overlapping" the signals produced by the different ligands are at certain activities might offer appropriate quantities. Its numerator also describes how "far" the function Ψ is from a certain linearity

$$\Psi(\lambda,0)-\Delta_1(\lambda,\kappa)=\Psi(\lambda,0)-\Psi(\lambda,\kappa)+\Psi(0,\kappa) \tag{7.21}$$

which will be close to zero for approximate Type 1 interaction.

Bibliography

Ackers, G. K., M. A. Shea, and F. R. Smith (1983). Free energy coupling within macromolecules: the chemical work of ligand binding at the individual sites in cooperative systems. *Journal of Molecular Biology 170*, 223–242.

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter (2008). *Molecular Biology of the Cell* (5th ed.). Garland Science, New York.

Bashford, D. and M. Karplus (1991). Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. *Journal of Physical Chemistry 95*(23), 9556–9561.

Becker, T., R. T. Ullmann, and G. M. Ullmann (2007). Simulation of the electron transfer between the tetraheme subunit and the special pair of the photosynthetic reaction center using a microstate description. *Journal of Physical Chemistry B 111*(11), 2957–2968.

Ben-Naim, A. (2001). *Cooperativity and Regulation in Biochemical Processes*. Kluwer Academic/Plenum Publishers, Dordrecht.

Berg, J., J. Tymoczko, and L. Stryer (2007). *Biochemistry*. W. H. Freeman, New York.

Bombarda, E. and G. M. Ullmann (2010). pH-dependent pKa values in proteins - a theoretical analysis of protonation energies with practical consequences for enzymatic reactions. *Journal of Physical Chemistry B 114*(5), 1994–2003.

Brennan, D. J., D. P. O'Connor, H. Laursen, S. F. McGee, S. McCarthy, R. Zagozdzon, E. Rexhepaj, A. C. Culhane, F. M. Martin, M. J. Duffy, G. Landberg, L. Rydén, S. M. Hewitt, M. J. Kuhar, R. Bernards, R. C. Millikan, J. P. Crown, K. Jirström, and W. M. Gallagher (2012). The cocaine-and amphetamine-regulated transcript mediates ligand-independent activation of ER alpha, and is an independent prognostic factor in node-negative breast cancer. *Oncogene 31*(30), 3483–3494.

Brown, D. (2010). A mathematical model of the Gac/Rsm quorum sensing network in *Pseudomonas fluorescens*. *Biosystems 101*(3), 200–212.

Buchler, N. E. and M. Louis (2008). Molecular titration and ultrasensitivity in regulatory networks. *Journal of Molecular Biology 384*(5), 1106–1119.

Burgin, M. and G. Meissner (2012). Negative probabilities in financial modeling. *Wilmott 2012*(58), 60–65.

Cantor, C. R. and P. R. Schimmel (1980). *Biophysical Chemistry. Part III. The Behavior of Biological Macromolecules* (1st ed.). W. H. Freeman, New York.

Cohen, E., T. Cvitas, J. Frey, B. Holmstroem, K. Kuchitsu, A. R. Marquardt, I. Mills, F. Pavese, M. Quack, J. Stohner, H. Strauss, M. Takami, and A. Thor (2008). *Quantities, Units and Symbols in Physical Chemistry* (3rd ed.). IUPAC and RSC Publishing, Cambridge.

Cox, D. A., J. Little, and D. O'Shea (2005). *Using Algebraic Geometry* (2nd ed.). Springer, New York.

Cox, D. A., J. Little, and D. O'Shea (2008). *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra.* (3rd ed.). Springer, New York.

Cox, D. R. (1955). A use of complex probabilities in the theory of stochastic processes. In *Proceedings of the Cambridge Philosophical Society*, Volume 51, pp. 313–319. Cambridge University Press, Cambridge.

Feynman, R. P. (1987). Negative probability. *Quantum implications: essays in honour of David Bohm*, 235–248.

Garcia-Moreno, B. E. (1995). Probing structural and physical basis of protein energetics linked to protons and salt. *Methods in Enzymology 259*, 512–538.

Gnacadja, G. (2011). A method to calculate binding equilibrium concentrations in the allosteric ternary complex model that supports ligand depletion. *Mathematical Biosciences 232*(2), 135–141.

Grima, R. (2010). An effective rate equation approach to reaction kinetics in small volumes: theory and application to biochemical reactions in nonequilibrium steady-state conditions. *Journal of Chemical Physics 133*(3), 035101–035101.

Gutierrez, P. S., D. Monteoliva, and L. Diambra (2012, 09). Cooperative binding of transcription factors promotes bimodal gene expression response. *PLoS ONE 7*, e44812.

Hameed, A. G., N. D. Arnold, J. Chamberlain, J. A. Pickworth, C. Paiva, S. Dawson, S. Cross, L. Long, L. Zhao, N. W. Morrell, et al. (2013). Tumour necrosis factor-related apoptosis-inducing ligand is a novel therapeutic target in pulmonary arterial hypertension. *The Lancet 381*, S47.

Hartle, J. B. (2008). Quantum mechanics with extended probabilities. *Physical Review A 78*(1), 012108.

Hasselbalch, K. (1916). *Die Berechnung der Wasserstoffzahl des Blutes aus der freien und gebundenen Kohlensäure desselben, und die Sauerstoffbindung des Blutes als Funktion der Wasserstoffzahl.* Julius Springer, Berlin.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*(1), 97–109.

Henderson, L. J. (1913). *The Fitness of the Environment.* Macmillan Company, New York.

Hill, A. V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *Journal of Physiology 40*, iv–vii.

Hill, A. V. (1913). The combinations of haemoglobin with oxygen and with carbon monoxide. I. *Biochemical Journal 7*(5), 471.

Hofmann, H. F. (2009). How to simulate a universal quantum computer using negative probabilities. *Journal of Physics A: Mathematical and Theoretical 42*(27), 275–304.

Horecker, B. L. (1943). The absorption spectra of hemoglobin and its derivatives in the visible and near infra-red regions. *Journal of Biological Chemistry* (148), 173–183.

Hunter, C. A. and H. L. Anderson (2009). What is cooperativity? *Angewandte Chemie International Edition 48*(41), 7488–7499.

Kærn, M., T. C. Elston, W. J. Blake, and J. J. Collins (2005). Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics 6*(6), 451–464.

Karlsson, D., S. Karlsson, E. Gustafsson, B. H. Normark, and P. Nilsson (2007). Modeling the regulation of the competence-evoking quorum sensing network in *Streptococcus pneumoniae*. *Biosystems 90*(1), 211–223.

Kelly, F. P. (2011). *Reversibility and Stochastic Networks*. Cambridge University Press, Cambridge.

Kosanke-Schütz, K., M. Gabriel, B. Weiß becker, H. Reinecke, D. Werner, U. T. Koch, and S. Schütz (2011). Sequence matters – selective adaptation in electroantennographic response to binary odour mixtures by the Colorado potato beetle. *Journal of Applied Entomology*, 372–385.

Kragh-Hansen, U. (2013). Molecular and practical aspects of the enzymatic properties of human serum albumin and of albumin-ligand complexes. *Biochimica et Biophysica Acta (BBA)-General Subjects*.

Kuttler, C. and B. A. Hense (2008). Interplay of two quorum sensing regulation systems of *Vibrio fischeri* . *Journal of Theoretical Biology 251*(1), 167–180.

Landau, L. and E. Lifschitz (1987). *Statistische Physik Teil I*. Lehrbuch der theoretischen Physik Band V. Akademie-Verlag, Berlin.

Leppänen, V.-M., D. Tvorogov, K. Kisko, A. E. Prota, M. Jeltsch, A. Anisimov, S. Markovic-Mueller, E. Stuttfeld, K. N. Goldie, K. Ballmer-Hofer, et al. (2013). Structural and mechanistic insights into VEGF receptor 3 ligand binding and activation. *Proceedings of the National Academy of Sciences 110*(32), 12960–12965.

Martini, J. W. R., M. Habeck, and M. Schlather (2014). A derivation of the grand canonical partition function for systems with a finite number of binding sites using a markov chain model for the dynamics of single molecules. *Journal of Mathematical Chemistry 52*(2), 665–674.

Martini, J. W. R., M. Schlather, and G. M. Ullmann (2013a). On the interaction of different types of ligands binding to the same molecule part II: systems with n to 2 and n to 3 binding sites. *Journal of Mathematical Chemistry 51*(2), 696–714.

Martini, J. W. R., M. Schlather, and G. M. Ullmann (2013b). On the interaction of two different types of ligands binding to the same molecule part I: basics and the transfer of the decoupled sites representation to systems with n and one binding sites. *Journal of Mathematical Chemistry 51*(2), 672–695.

Martini, J. W. R., M. Schlather, and G. M. Ullmann (2013c). The meaning of the decoupled sites representation in terms of statistical mechanics and stochastics. *MATCH Communications in Mathematical and in Computer Chemistry 70*(3), 829–850.

Martini, J. W. R. and G. M. Ullmann (2013). A mathematical view on the decoupled sites representation. *Journal of Mathematical Biology 66*(3), 477–503.

Medvedev, E. and A. Stuchebrukhov (2006). Kinetics of proton diffusion in the regimes of fast and slow exchange between the membrane surface and the bulk solution. *Journal of Mathematical Biology 52*, 209–234.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics 21*(6), 1087–1092.

Monteoliva, D., C. B. McCarthy, and L. Diambra (2013). Noise minimisation in gene expression switches. *PLoS ONE 8*(12), e84020.

Mückenheim, W., G. Ludwig, C. Dewdney, P. Holland, A. Kyprianidis, J. Vigier, N. Cufaro Petroni, M. Bartlett, and E. Jaynes (1986). A review of extended probabilities. *Physics Reports 133*(6), 337–401.

Onufriev, A., D. A. Case, and G. M. Ullmann (2001). A novel view of pH titration in biomolecules. *Biochemistry 40*(12), 3413–3419.

Onufriev, A. and G. M. Ullmann (2004). Decomposing complex cooperative ligand binding into simple components: connections between microscopic and macroscopic models. *Journal of Physical Chemistry B 108*(30), 11157–11169.

Parisi, G. (1983). On complex probabilities. *Physics Letters B 131*(4), 393–395.

R Core Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Reif, F. (1987). *Statistische Physik und Theorie der Wärme*. Walter de Gruyter, Berlin.

Schellman, J. A. (1975). Macromolecular binding. *Biopolymers* (14), 999–1018.

Seneta, E. (2006). *Non-negative Matrices and Markov Chains*. Springer.

Stefan, M. I. and N. Le Novère (2013, 06). Cooperative binding. *PLoS Computational Biology 9*(6), e1003106.

Sugiyama, T., T. Asai, Y. M. Nedachi, Y. Katanasaka, K. Shimizu, N. Maeda, and N. Oku (2013). Enhanced active targeting via cooperative binding of ligands on liposomes to target receptors. *PloS ONE 8*(6), e67550.

Tanford, C. and J. G. Kirkwood (1957). Theory of protein tiration curves I. General equations for impenetrable spheres. *Journal of the American Chemical Society 79*(20), 5333–5339.

Thattai, M. and A. van Oudenaarden (2002). Attenuation of noise in ultrasensitive signaling cascades. *Biophysical Journal 82*(6), 2943–2950.

Till, M. S., T. Essigke, T. Becker, and G. M. Ullmann (2008). Simulating the proton transfer in Gramicidin A by a sequential dynamical Monte Carlo method. *Journal of Physical Chemistry B 112*(42), 13401–13410.

Ullmann, R. T. and G. M. Ullmann (2011). Coupling of protonation, reduction and conformational change in azurin from *Pseudomonas aeruginosa* investigated with free energy measures of cooperativity. *Journal of Physical Chemistry B 115*, 10346–10359.

Walczak, A. M., A. Mugler, and C. H. Wiggins (2009). A stochastic spectral analysis of transcriptional regulatory cascades. *Proceedings of the National Academy of Sciences 106*(16), 6529–6534.

Wyman, J. and S. J. Gill (1990). *Binding and Linkage: Functional Chemistry of Biological Macromolecules*. University Science Books, Mill Valley, CA.

Zak, M. (1998). Incompatible stochastic processes and complex probabilities. *Physics Letters A 238*(1), 1–7.