

# **Kinetics of Multivalent Binding Processes**

Franziska Erlekm

**Dissertation**

zur Erlangung des akademischen Grades eines Doktors der  
Naturwissenschaften am Fachbereich Mathematik und Informatik  
der Freien Universität Berlin

Berlin, November 2022

---

Betreuer und Erstgutachter: PD Dr. Marcus Weber

Zweitgutachterin: Prof. Dr. Bettina G. Keller

Tag der Disputation: 11. Juli 2023

## Acknowledgements

First of all I thank my supervisor Marcus Weber for giving me the opportunity to write a PhD thesis in his molecular dynamics group at Zuse Institute Berlin (ZIB). Thanks to my second supervisor Beate Paulus for helping me to focus on what is essential. Thanks to my old and new ZIB colleagues for critical discussions and occasional distractions. Saba Nojoui and Margarita Kostré for proofreading parts of my thesis. Christiano Gillano for layout improvements. Thanks to all my family members and friends for support and motivation during the ups and downs of PhD life. Thanks to my son Theodor Erlekm for changing my life for the better.

My work at ZIB from 2016 - 2020 has been funded by Deutsche Forschungsgemeinschaft (DFG) through the Cooperative Research Centre grant 765 "Multivalency as chemical organisation and action principle", C02 (Non-equilibrium kinetics of multivalent binding processes). The knowledge on multivalency gained during this time set the foundation and enabled me to write this thesis.

This PhD thesis is dedicated to my grandmother Margot Weber.  
Danke.

# Abstract

Multivalent bindings are a combination of several weaker bonds between two molecules such as a receptor and a ligand. Because the combined bonds are stronger than the individual ones, the molecules may interact longer. Thus, this concept is very interesting for targeted drug design. Theoretically, these bindings and unbindings are rare stochastic events. Literature on multivalency is vast, especially on specific experimental setups. However, there exists little theoretical studies for general  $n$ -valent settings. This thesis aims to fill that gap by modelling the binding process as kinetic rate matrices and applying the clustering algorithm PCCA+. While the binding and unbinding of the single ligand-receptor pairs happens on a fast time scale i.e. in the micro-perspective, the association or dissociation of a complex is acting on a slow timescale, thus in the macro-perspective. The existing kinITC method does not capture the switch between these time scales. Thus the method proposed in this thesis describes an alteration to kinITC, called kinITC+.

The key findings of the thesis are:

- It is possible to gain kinetic information from thermodynamic data.
- The macroscopic binding rate  $k_{on}$  is not constant, but ligand concentration dependent.
- There is at least one counterexample to the assumption that the thermodynamic contribution of rebinding increases with valency.

---

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Multivalency</b>	<b>9</b>
2.1	Introduction . . . . .	9
2.1.1	Ligand Structure . . . . .	10
2.2	Examples of Multivalency . . . . .	11
2.3	Notation and Quantification . . . . .	13
2.3.1	Notation of Monovalent and Multivalent Bindings . . . . .	13
2.3.2	Fundamental Equations and Quantification Formalisms . . . . .	15
2.4	Thermodynamic and Kinetic Studies of Multivalency . . . . .	21
2.4.1	Isothermal Titration Calorimetry . . . . .	26
2.4.2	Kinetic ITC . . . . .	29
<b>3</b>	<b>Clustering Multivalent Binding Data</b>	<b>31</b>
3.1	Mathematical Preliminaries . . . . .	31
3.2	The Transfer Operator . . . . .	34
3.2.1	Discrete Transfer Operators . . . . .	36
3.2.2	Defining the Rate Matrix . . . . .	39
3.3	Clustering with PCCA+ . . . . .	45
<b>4</b>	<b>KinITC+</b>	<b>53</b>
4.1	Application of PCCA+ to Kinetic Binding Data . . . . .	53
4.1.1	Bivalent Ligand Example . . . . .	56
4.1.2	Trivalent Ligand Example . . . . .	58
4.1.3	Pentavalent Ligand Example . . . . .	59

## CONTENTS

---

4.2	Parameter Interdependency . . . . .	59
4.2.1	Shapes of $k_{on}$ Curves . . . . .	60
4.2.2	General Relationship of the Single and Overall Binding Rates	62
4.2.3	Alternative Enhancement Factor . . . . .	69
4.3	Limitations of the Model . . . . .	73
<b>5</b>	<b>Wiseman Fitting</b>	<b>75</b>
5.1	The classical Wiseman Fitting . . . . .	75
5.2	The $Q_c$ Fitting . . . . .	77
5.3	Extracting experimental ITC Data . . . . .	78
5.4	Application of $Q_c$ and extended Wiseman Fitting . . . . .	78
5.4.1	Bivalent Binding Example . . . . .	79
5.4.2	Trivalent Binding Example . . . . .	81
5.5	Limitations of the Model . . . . .	82
<b>6</b>	<b>Rebinding</b>	<b>87</b>
6.1	Biochemical and Mathematical Perspectives on Rebinding . . . . .	88
6.2	Quantification of Rebinding . . . . .	89
6.3	Numerical Examples . . . . .	94
6.3.1	Bivalent Ligand Example . . . . .	94
6.3.2	Trivalent Ligand Example . . . . .	95
6.3.3	Pentavalent Ligand Example . . . . .	96
6.4	Conclusion . . . . .	99
<b>7</b>	<b>Conclusion</b>	<b>101</b>
	<b>Zusammenfassung</b>	<b>107</b>
	<b>Appendix</b>	<b>120</b>
	<b>References</b>	<b>127</b>



# Chapter 1

## Introduction

With the increase of computation power of modern supercomputers, the boundary of the curse of dimensionality has been pushed further and further. One of the many research fields that benefited substantially is computational drug design. With the help of supercomputing, not every pharmacological experiment has to be carried out in the laboratory, but the molecule can be first systematically altered virtually. Effective medication not only fits to the specific pain receptor, but binds as long as possible with as many ligands as possible.

One use case for the virtual drug development was the wide ranging opioid use and abuse especially in the USA in the last 20 years with severe physical and lethal side effects. According to the US Centers for Disease Control and Prevention (CDC) almost "500,000 people died from an overdose involving any opioid, including prescription and illicit opioids" from 1999-2019 [1]. In the US the fentanyl induced death rates spiked in the years 2013-2014 [2]. Because of the addictive side effect, the black market was flourishing and fentanyl submissions did not match subscription counts. The goal of effective drug development was to design an opioid without these lethal and addictive side effects. A successful example for virtual drug development in this case is the prototype of fluorinated fentanyl "which activates the  $\mu$ -opioid receptors in the inflamed tissue only" (pH up to 5.5) instead of healthy tissue (pH of 7.4) and thus reduces noxious and/or lethal side effects [3].

## 1. INTRODUCTION

---

Another use case for the research of multivalent virus interaction is to tackle respiratory diseases. With the outbreak of the worldwide coronavirus (SARS-CoV-2) in late 2019, the topic of multivalent virus research got worldwide attention. The immense health and economic related lagged effects of COVID-19 have already cost over 6.6 millions of lives according to the World Health Organization (WHO) [4] and brought public life as well as economy to a standstill. 2020-2021 and probably the subsequent years will go down in history as the years of the corona virus pandemic. Tabish and Hamblin describe the multivalent nature of SARS-CoV-2 and how multivalent nanomedicine has the potential to "target multiple viral factors involved in infections at cellular levels" [5]. What is evident is that a multivalent vaccine needed to be developed urgently to treat this pulmonary disease. Hardly any topic has had such a broad media coverage as COVID-19 vaccination development, vaccinations, and infection incidences in 2021. With the ongoing mutation of the virus and yet increasing incidence numbers and death tolls, the search for multivalent vaccines is still ongoing.

These uses cases show how existing drugs can be altered virtually in order to a) minimise side effects and b) to develop new drugs faster than in traditional laboratory research in order to quickly react to a pandemic.

Multivalent ligand-receptor interaction studies were trending in the last 20 years, thus biochemistic literature is vast. However, it is either rather conceptual [6; 7] or for very specific experimental setups [8; 9]. A mathematical approach, treating multivalent bindings as stochastic events, is still lacking. This thesis aims to fill that gap.

Physically, a molecular system has  $N$  atoms that move on some energy landscape. Each atom has three spatial coordinates  $q_i$  and three momenta  $p_i$  with  $i = 1, 2, 3$ . Thus, the phase state space  $\Gamma = \Omega \times \mathbb{R}^{3N}$ . In the micro perspective, the molecules have a probability to be in a certain micro state according to the Boltzmann distribution [10]

$$\pi(q, p) = \frac{1}{Z} \exp(-\beta H(q, p))$$

with the inverse temperature  $\beta = 1/(k_B T)$  divided by the Boltzmann constant

---

$k_B$  and  $Z = \int_{\Gamma} \exp(-\beta H(q, p)) d(q, p)$ .  $H$  is a Hamiltonian function given by the sum of kinetic energy  $K$  and potential energy  $V$ :  $H(q, p) = K(p) + V(q)$ . This is the usual procedure of classical molecular dynamics. A trajectory is started at a certain point in the energy landscape and is simulated over the longest possible time span. One is interested which of the 'valleys' and 'hills' of the energy landscape are visited and how often. However, if one is interested in the macro perspective, this approach is not suitable. Here, the molecular kinetics are obtained as a projection by clustering many micro states into very few macro states [11]. Another important aspect of the differentiation between the macro and the micro perspective are the timescales. The single formation and loss of bonds between ligand and receptor happen on a fast timescale. The clustered macro binding and unbinding take place on a slow timescale. These bindings with their respective timescales are depicted in Figure 1.1.

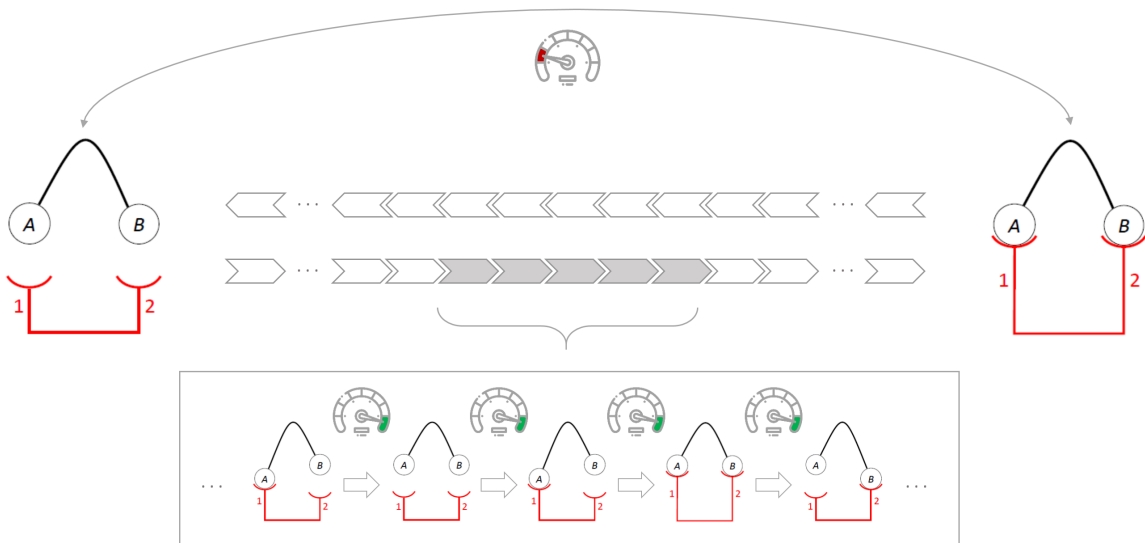


Figure 1.1: This schema shows a simplification of a multivalent binding with exemplary two binding sites A and B at the ligand and two sites 1 and 2 at the receptor. The upper part of the figure shows the slow timescale of binding and unbinding the whole entities. The lower scheme shows an excerpt of the many microstates.

In this work we are particularly interested into the two macro states 'bound'

## 1. INTRODUCTION

---

and 'unbound'. For monovalent binding interactions this is straight forward, but when can a multivalent receptor-ligand complex be considered as bound? Is it only bound if all of the binding sites are occupied or more than half of them? What if ligand and receptor do not have the same number of binding sites, how can one tell if the system is bound or not? This question cannot be answered in a black and white manner. There are many shades in between that will be captured by membership functions. Mathematically, these are bases having values between 0 and 1, capturing the situations of being gradually bound.

Once the first bond is made, both ligand and receptor backbones are spatially close. That means, that every subsequent bond is more likely. Just like binding, they can also unbind again. After unbinding, the ligand and receptors are still in a spatial proximity, such that another binding is probable. If this rebinding phenomenon can be quantified, it could be applied in drug design in such a way that drug dosing may be decreased and with it the drug's side effects. With rebinding taking place, the ligand has a high probability to bind again and has more time to develop its effect on the receptor.

## Related Work

Monovalent and bivalent binding processes are well understood and there exists extensive amounts of literature, e.g. Merchant et al. [12], Howorka et al. [13], and Corazza et al. [14]. Existing multivalent ligand-receptor interaction studies are either rather conceptual like in Krishnamurthy et al. [6], Mammen and Whitesides [7] or more experimental for very specific setups such as in Kiessling et al. [8] or Lundquist and Toone [9]. Biochemistic literature is vast, but a mathematical approach, treating multivalent bindings as stochastic events, is still lacking.

One of the first and most cited works in multivalency is the review by Mammen, Choi and Whitesides from the late 1990s [7]. Therein the authors describe many examples of multivalent interactions in biology such as *E. coli* bacteria attachment to endothelial urethral cells or the binding of antibodies to bacteria. The nomenclature and descriptions of thermodynamic features such as enthalpy and entropy introduced by these authors are also used in the present thesis.

---

Another often cited paper is the fundamental research by Cairo, Kiessling et al. [15]. In this paper the authors analyzed the influence of stoichiometry, rate of cluster formation and receptor proximity on multivalent ligand binding.

A more recent overview of multivalent interactions is given by Fasting et al. [16]. Their work ranges from theoretical aspects such as artificial models to practical aspects such as measurement techniques and scaffold architecture.

A contemporary and comprehensive overview of concepts, research and applications of multivalency is given in the book 'Multivalency - Concepts, Research and Applications' by Huskens, Prins, Haag and Ravoo [17]. Especially the chapters on models and methods by Huskens and design principles for super selectivity by Curk et al. were of particular interest for the present work.

The method of kinetic isothermal titration calorimetry (kinITC) shows how to obtain kinetic information of a thermodynamic measurement technique and can be found in the papers by Dumas et al. [18] and Egawa et al. [19]. However, this work focuses more on experimental aspects rather than mathematical implications.

Concerning the mathematical foundation of molecular dynamics, the book 'Metastability and Markov state models in molecular dynamics: modeling, analysis, algorithmic approaches' by Schütte and Sarich [10] must be named. In the present thesis the nomenclature of Markov processes, metastability and transfer operators have been adapted from their book.

Another thorough analysis of the transfer operator approach and Galerkin discretization is the dissertation of Nielsen [20].

Details of the PCCA+ clustering algorithm used throughout the thesis to analyse experimental data can be found in the works of Weber [11; 21].

A thorough overview of the rebinding effect from a biochemical perspective is given by Vauquelin [22]. The mathematical aspects of rebinding and the resulting lack of Markovianity are presented in Weber et al. [23]. In addition to that paper, the quantification of rebinding for non-reversible binding processes is further described in the thesis of Röhl [24; 25].

The core of this thesis is to show that the binding and unbinding rates are concentration dependent. This fact is not yet fully understood and not commonly

## 1. INTRODUCTION

---

accepted within the scientific community. Therefore, there exists very little literature to explain this phenomenon. However, the paper of Sing and coworkers [26] must be mentioned. Therein the authors deliver an explanation for this phenomenon for a dimeric protein by coarse-grained simulation. They emphasize that the concentration-dependent unbinding was only studied for some proteins interacting with DNA and do not generalize this phenomenon.

### Thesis Structure

The thesis is organized as follows:

In Chapter 2, multivalency is described from a biochemistry point of view. Therein the most important parameters quantifying multivalent interactions and their notations are introduced, as well as isothermal titration calorimetry (ITC), a measurement technique to quantify binding effects. Experimental ITC data will be used in later chapters for numerical applications.

The core of this work is Chapter 3, where multivalent binding kinetics are coarse grained via clustering and projection onto an invariant subspace. Before experimental data are clustered, the necessary mathematical background is introduced.

In Chapter 4 the clustering concept lined out in Chapter 3 is applied numerically for bivalent, trivalent and pentavalent ITC data. Further, it is investigated how the binding parameters interdepend. Finally, the model's limitations are discussed.

In Chapter 5 an alternative way of obtaining binding information is presented: the Wiseman fitting. It is applied to a bivalent and a trivalent ITC example. The Wiseman fitting is compared to the PCCA+ clustering from Chapter 4 and the model limitations are pointed out.

Chapter 6 is dedicated to the rebinding concept. Both, the biochemistry and the mathematical meaning of rebinding are discussed. Finally, the same experimental data from Chapter 3 are used to apply the concept of rebinding quantification.

---

The final chapter contains a summary of the thesis and a critical discussion of all the methods presented. Further, the outlook highlights some potential future research ideas.

## Aim of the Work

The following thesis highlights the stochastic implications of ligand receptor interactions. As mentioned above, it is ambiguous to tell whether a multivalent complex is bound or unbound. So far, classical ITC analysis assumed a constant binding rate over all titration steps. This thesis shows that in fact the binding rates are not constant, but follow a certain regime over time. This regime can be modeled by clustering. Therefore it is shown that applying a Galerkin discretization of the state space is a powerful tool. The key concept is a projection of the  $n$ -dimensional state space down to a two-dimensional subspace. This is done by using base functions having values between 0 and 1 referring to gradually bound states.

The present work serves as a contribution to the concentration dependent binding rates literature. It attempts to explain the relationship theoretically for all kinds of valencies from a mathematical perspective.

The insights of this thesis boil down to answer the following

**Research questions:**

1. *Is it possible to gain kinetic information from thermodynamic information?*
2. *How can we determine when a complex is bound or unbound?*
3. *Does rebinding quantification work for multivalent binding settings?*

## 1. INTRODUCTION

---



# Chapter 2

## Multivalency

### 2.1 Introduction

Multivalency is a natural phenomenon in biochemical systems and is characterized by the interaction of two molecules by more than one binding site [16; 17]. Through multiple functionalities on one entity and complementary functionalities on another one (such as intramolecular bonds). Even though they may be weak individually, the binding affinity as well as the complex stability are increased. The pioneers in multivalency research are among others Mammen, Choi and Whitesides [27], Gestwicki [28] and Cairo [15]. Due to its handy application in modern drug design the topic became popular within the last 20 years. Among today's researchers are Huskens, Haag, Prins and Javoo [17], as well as Fasting, Schalley, Weber et al. [16] to name just a few.

Studying multivalency is of practical use especially for molecular drug design. The receptor can be a surface protein receptor like for the Influenza A virus in case of an infection [29], or body cells of inflamed tissue. Inflamed tissue and cancer cells have a close relationship [30]. In both cases, influenza and cancer cells, as many receptor binding sites as possible are aimed to be occupied by the ligand, i.e. drug researchers strive to develop a multivalently bio-orthogonally designed drug, in order to enhance medication effects and minimize drug dosing. Multivalency can be exploited for targeting cancer immunotherapy in different ways, see [31; 32].

## 2. MULTIVALENCY

---

Another example from drug design is using the bivalency of the estrogen receptor for efficient contraceptives [33].

Multivalency is closely linked to cooperativity, also called allosterism. Curk et al. describe it as the change of the affinity of an interaction pair caused by the presence of a neighboring formed interaction pair [17]. Fasting et al. differentiate between non-cooperative (additive) and positively cooperative (synergistic) binding interactions. "Most multivalent systems are rather negatively cooperative (interfering), but still multivalent drugs can yield a bigger effect than analogue monovalent drug dosing" [16].

### 2.1.1 Ligand Structure

The *ligand moieties* are connected to a backbone via a molecular spacer or scaffold. *Spacers* can be rigid or flexible, rather long or short, limiting the number of reachable binding sites. In addition, the *scaffold* can determine the shape of the whole molecule by i.e. forming a spherical shape. The binding moiety is connected to the spacer or scaffold by a linker. A linker can also maintain a certain angle between spacer and moiety [34]. The shape of spacers also influence the conformation entropy of the entities. If the first bond is made, some entropy is lost because certain degrees of freedom are lost in terms of rotation and translation. If the spacers are too short, some binding sites cannot be reached after the first bond is made. If they are too long, the immediate effect is an increase in the number of unproductive degrees of freedom and therefore decrease the binding affinity [16]. An example of how spacers are manipulated with DNA such that more than one binding site can be accessed is described for carbohydrate-lectin interactions by Scheibe et al [35].

The number of moieties is mostly unequal on the ligand and receptor site. Technically, interactions between multivalent receptors and several monovalent ligands cannot be called multivalent. Further, the binding sites on the receptor are not necessarily the same. If they are, they are called *homo-multivalent*, if they differ, they are *hetero-multivalent* [16].

## 2.2 Examples of Multivalency

---

Another aspect in multivalent bindings is *sensitivity*. The binding strength is highly sensitive to external influences such as temperature, pH, and receptor concentration [17]. The latter determines the number of reachable binding sites and therefore influences the binding strength. The so-called *selectivity* denotes the ability of multivalent ligands and receptors to distinguish between substrates depending on the surface of binding sites. Sensitivity in turn influences the selectivity [17].

Two more concepts are affinity and avidity. *Affinity* denotes the interaction strength between two molecules [36]. In contrast, *avidity* is the combined strength of multiple affinities [17], and is therefore also called functional affinity. Avidity is more than the sum of the individual binding affinities [34]. However, a multivalent ligand can have two or more separate coordinate bonds to single atoms. In this case a so called chelate complex is formed [34].

In this work, we assume a one-to-one stoichiometry of ligand and receptor. Further, only homo-multivalent ligands of low valency numbers up to 10 are considered.

## 2.2 Examples of Multivalency

A prominent botanical example of multivalency is the burr, binding with many small hooks to hairy objects acting as loops [16; 34]. The more hooks are fastened, the more stable is the whole system. This principle was copied by humans as hook-and-loop-fastener, also known as Velcro. Further, one of the best known biological examples for multivalency is the binding of oxygen to haemoglobin. Haemoglobin has four binding sites acting cooperatively, i.e. once the first binding is made, all three subsequent binding interactions are facilitated. This phenomenon is called *allosteric cooperativity*. Thus, haemoglobin tends to bind either with all four binding sites or none at all [17].

The last example comes again from the medical field: the adhesion of a virus such as influenza to a mammalian cell [7]. As depicted in Figure 2.1, the influenza virus attaches itself to the bronchial epithel cell. Here, several trimers of hemagglutinin bind to moieties of sialic N-acetylneuraminic acid of the target cell. There are

## 2. MULTIVALENCY

600-1200 hemagglutinin binding sites on one virus particle and 50-200 binding sites per  $100 \text{ nm}^2$  of target cell.

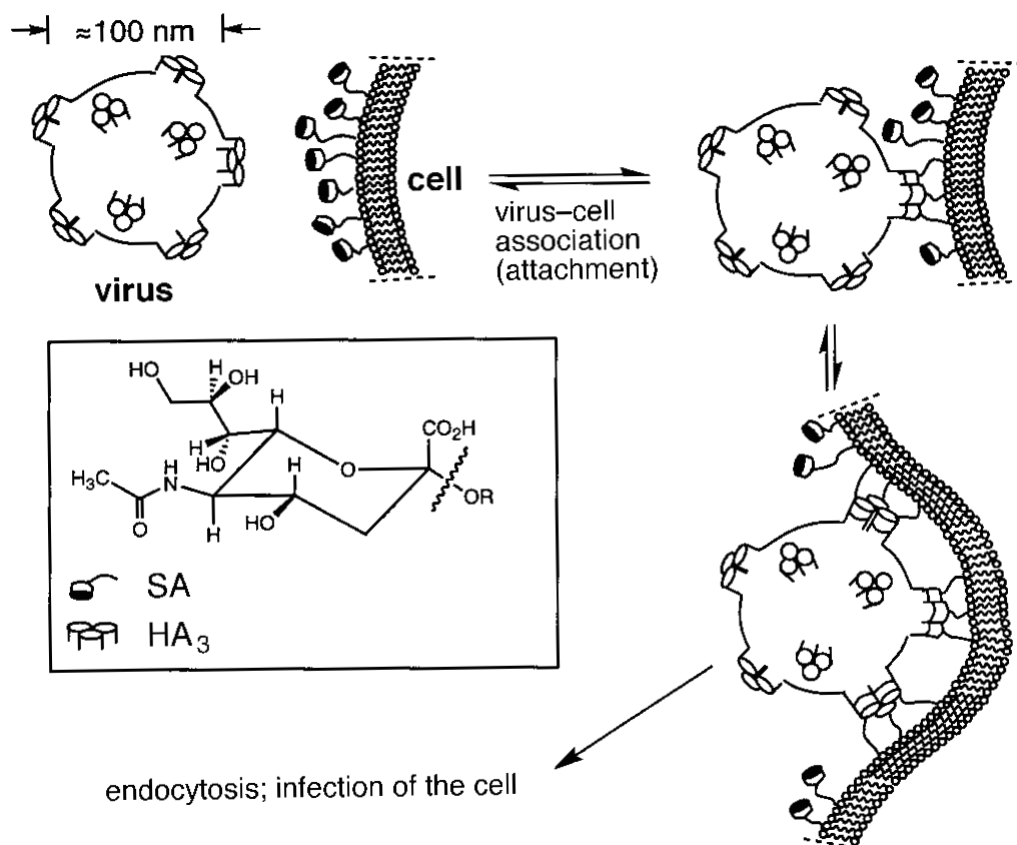


Figure 2.1: The influenza virus attaches to cells by interaction of trimeric hemagglutinin (HA<sub>3</sub>, shown as protruding cylinders on the virus) with sialic acid (SA, shown as caps on the cell). Only a few of the hemagglutinin trimers and SA groups are represented; neither it is to scale [7]. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

The multivalent nature of virus particles is exactly what is exploited to develop drugs. The more virus binding sites are blocked, the less it can attach itself to the host cells in the human body. Preferably, these virus binding sites are blocked by a scaffold hindering the virus to approach the cell surface as shown in Figure 2.2.

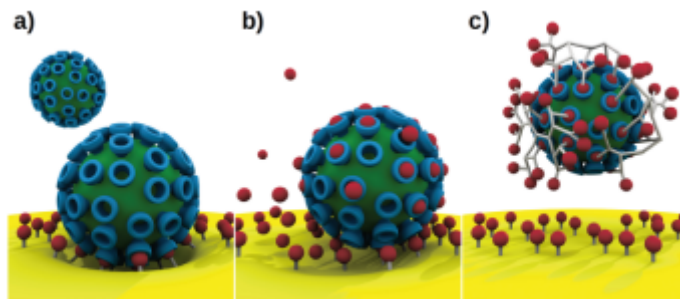


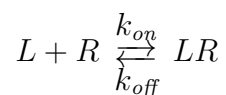
Figure 2.2: Multivalent interaction types of a virus particle to a cell surface. In a) the virus approaches the host. In b) a monovalent drug mimicking the host ligands is attempting to block as many virus receptors as possible. c) shows that a scaffold connecting several drug ligands is much more effective in hindering the virus from attaching itself to the cell surface. Figure taken from Huskens et al. [17]. Reprinted with kind permission from John Wiley and Sons, Inc.

## 2.3 Notation and Quantification

The following section first introduces the notation of multivalent binding steps and then highlights the most important quantities to estimate binding rates. Finally, one of the most common experiment techniques to determine binding rates is explained.

### 2.3.1 Notation of Monovalent and Multivalent Bindings

In general, ligand-receptor (L-R) interactions are described by



with a binding rate  $k_{on}$  in units of  $[M^{-1}s^{-1}]$  and unbind with the rate  $k_{off}$  in units  $[s^{-1}]$ . In equilibrium it holds

$$k_{on}[L][R] \leftrightarrow k_{off}[LR].$$

The squared brackets denote the molar concentrations, i.e. of the unbound ligands  $[L]$ , of the unbound receptors  $[R]$  and of the complex  $[LR]$ . The dissociation

## 2. MULTIVALENCY

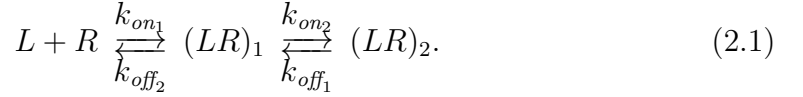
---

constant is the ratio of the unbinding rate to the binding rate, that means

$$K_d = \frac{k_{off}}{k_{on}} = \frac{[R][L]}{[LR]}$$

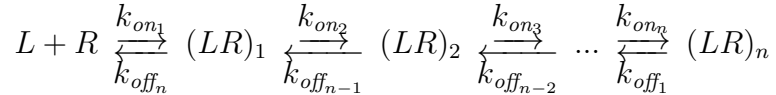
in units of molar [M]. The association rate is the inverse of the dissociation rate,  $K_a = 1/K_d = \frac{k_{on}}{k_{off}}$  in units of  $[M^{-1}]$ .

Adding one more intermediate step, the monovalent binding process can be extended into a bivalent one. That is



The subscript to (LR) denotes the number of bonds between the receptor and the ligand. In this case, there are two dissociation constants,  $K_{d_1}$  and  $K_{d_2}$ .  $K_{d_1}$  is similar to the monovalent binding rate:  $K_{d_1} = \frac{[R][L]}{[LR]}$ . The second dissociation constant is analogously the ratio of the concentrations of singly to doubly bound complexes according to  $K_{d_2} = \frac{[(LR)_1]}{[(LR)_2]}$ .

In a multivalent setting, given that  $n$  is the maximum number of possible bonds between two entities, one or more intermediate binding steps take place according to



with  $k_{on_i}$  the intermediate binding rates and  $k_{off_i}$  the intermediate unbinding rates. The subscript  $i$  in  $(LR)_i$  denotes how many bonds between ligand and receptor exist. Like in the monovalent case, the dissociation rate is the ratio of the unbinding to the binding rate in each binding step

$$K_{D_i} = \frac{k_{off_{n-i}}}{k_{on_i}}.$$

For the first binding step the dissociation constant is

$$K_{d_1} = \frac{[L][R]}{[(LR)_1]}.$$

The subsequent dissociation constants are

$$K_{D_i} = \frac{[(LR)_{i-1}]}{[(LR)_i]}.$$

Again, the association constant in each binding step  $K_{A_i}$  is the inverse, i.e.  $K_{A_i} = \frac{1}{K_{D_i}}$ .

### 2.3.2 Fundamental Equations and Quantification Formalisms

The following section deduces the equations used for ITC analyses as described in Wiseman et al. [37] and Goodrich and Kugel [36].

#### The $k_{on}$ rate

Of central importance is the observed  $k_{on}$  rate, also called  $k_{on-ITC}$ . For simplicity a 1 : 1 stoichiometry is assumed. The total ligand and receptor concentrations are the sum of the free species and the ones bound in the complex, respectively. The conservation equations state that  $[L]_{tot} = [L] + [LR]$  and  $[R]_{tot} = [R] + [LR]$ . The binding constant or association constant  $K_a$  is  $K_a = \frac{[LR]}{[L][R]}$ . With rearranging this equation for  $[LR]$  and differentiating with respect to  $[L]_{tot}$ , it can be related to the differential heat as

$$dQ = d([LR]\Delta H^0 V_0)$$

with  $\Delta H^0$  the molar enthalpy of binding and  $V_0$  the reaction cell volume.

#### The Wiseman parameter

The Wiseman parameter  $c$  is related to the number of binding sites  $r$ , but can also be computed with the total receptor concentration and the binding constant.

$$c = 1/r = [R]_{tot}K_a = [R]_{tot}/K_d.$$

## 2. MULTIVALENCY

---

For very high  $c$ -values, i.e.  $c = \infty$ , the binding is very tight and all added ligands are bound to saturation. Moderately tight bindings have  $c$ -values between 1 and 1000. Weak bindings have  $c$ -values close to 0 [37]. The  $c$ -value is connected to the slope of the isotherm. Like depicted in figure 2.3 it holds that the steeper the slope, the higher the  $c$ -value.

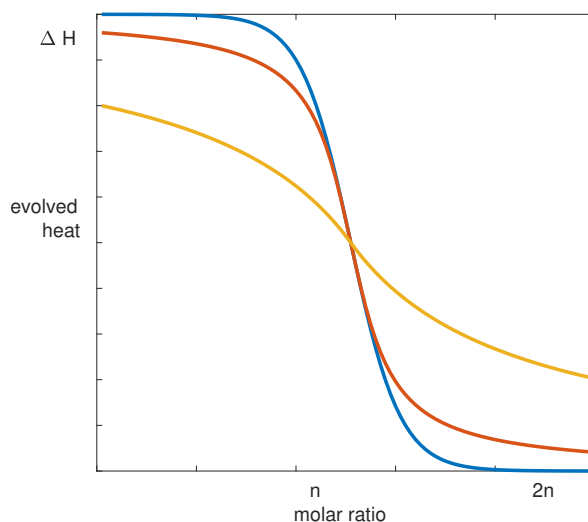


Figure 2.3: Isotherms for different  $c$ -values. The intercept is  $\Delta H$ . The higher the intercept on the y-axis, the higher the  $c$ -value. The blue curve corresponds to  $c = \infty$ . The flatter the slope of the curve and the lower the intercept, the smaller is the  $c$ -value. The molar ratio is approximately  $n$  at the inflection point of the curve.

Let us assume a binding process with two consecutive binding steps similar to the expression 2.1. The time derivative of the species concentration in the fully bound state is

$$\frac{\partial[(LR)_2]}{\partial t} = k_2[(LR)_1] - k_{-2}[(LR)_2].$$

The concentration change of the singly bound species is

$$\frac{\partial[(LR)_1]}{\partial t} = k_1[L][R] - k_{-1}[LR] - k_2[LR] + k_{-2}[(LR)_2].$$



## 2.3 Notation and Quantification

---

The observed  $k_{off}$  value from the ITC measurement is given by

$$\frac{[(LR)_2]_t}{[(LR)_2]_{max}} = 1 - e^{-k_{obs}t}.$$

$[LR]_t$  is the molar concentration of the complex at each time point and  $[LR]_{max}$  is the concentration of the complex at later time points where a plateau is reached. Assuming that the plateau is the equilibrium concentration is a proven natural scientific heuristic.  $k_{obs}$  is ligand concentration dependent, that is

$$\frac{1}{k_{obs}} = \frac{1}{k_2} + \frac{k_{-1}}{k_1 k_2 [L]_i} = \frac{K_{d1}}{k_2} \frac{1}{[L]_i} + \frac{1}{k_2}.$$

For increasing ligand concentration,  $k_{obs}$  depends only on the forward reaction rate,

$$\lim_{[L]_i \rightarrow \infty} \frac{1}{k_{obs}} = \frac{1}{k_2}.$$

The formation of the complex is given by

$$[(LR)_2]_t = [(LR)_2]_{max}(1 - e^{-k_{obs}t})$$

and the decay is  $[(LR)_2]_t = [(LR)_2]_{max} + [(LR)_2]_{min}(1 - e^{-k_{off}t})$ . This is due to the fact that  $\frac{[(LR)_2]_t}{[(LR)_2]_i} = e^{-k_{off}t}$  with the initial concentration  $[(LR)_2]_i$ .

### Effective Molarity

Effective molarity ( $EM$ ) is defined as the ratio between the rate constants of the intra- and intermolecular reactions, measured in  $[M]$ . The first bond between ligand and receptor entities is an intermolecular binding step. Any subsequent bond is intramolecular [17].  $EM$  is mainly a tool to compare different interaction settings.

The overall effective molarity is the product of the single association constants  $K_{ov} = K_{A_1} \cdot K_{A_2} \cdot \dots K_{A_n}$  in units  $[M^{-n}]$  [17]. It can be related to  $EM$  and the association rate in a monovalent setting by  $K_{ov} = K_{A_{mono}} \cdot EM$ . However, mostly these quantities are rather of theoretical nature. "For multivalent systems, experimental data usually provide access only to the final product, the concentrations

## 2. MULTIVALENCY

---

of the intermediates are too low to be observed” [17]. This is due to that fact, that the number of binding sites in biological systems is often unknown. On the other hand, for synthetic materials it is known [16].  $EM$  comes into play when comparing monovalent guests to a multivalent host with both multivalent guests and host such as in [16]. If two monovalent guests bind to a bivalent host, the binding coefficient only depends on the monovalent affinity constant. In the multivalent setting, it is a function of both the monovalent affinity constant and  $EM$ . Indeed, so far there is no theoretic relationship between the binding constant  $K$  and  $EM$ . In the papers by Kaufmann et al. [38] and Fasting et al. [16] a linear relationship was determined. However, depending on the setup of the experiment, there are different constant prefactors.

### The Enhancement Factor

The enhancement factor  $\beta$  examines the effect on the binding rate by using a multivalent ligand instead of a monovalent ligand with a multivalent receptor. It holds that  $\beta = \frac{K_{multi}}{K_{mono}}$  [16]. The concentration of free ligands and receptors influences the nature of binding. When the concentration of the free components approaches a value close to  $EM$ , the probability of intermolecular binding (polymerization) becomes equally likely as intramolecular binding (cyclization) [16]. Figure 2.4 depicts the relationship between dissociation constants and enhancement factor for monovalent and bivalent bindings.

### The Hill-Langmuir Curve

To measure the cooperativity described earlier in this chapter, the Hill-Langmuir curve  $\theta$  comes into play.  $\theta = \frac{[L]^n}{K_d + [L]^n}$  is a rectangular hyperbola denoting occupancy of receptors by ligands with  $\theta$  the ratio of the bound receptors to the total receptors in the experiment,  $[L]$  the unbound ligand concentration,  $K_d$  the dissociation rate and the Hill coefficient  $n$ . For  $n > 1$  there is positive cooperativity, i.e. if there already is a bond, the affinity for subsequent bonds is increased [39]. For  $n < 1$  there is negative cooperativity, i.e. the affinity decreases if there already is a bond. This may be due to unfavorable spacers, e.g. they are too short such that the next moiety cannot reach a binding site. For  $n = 1$  the single

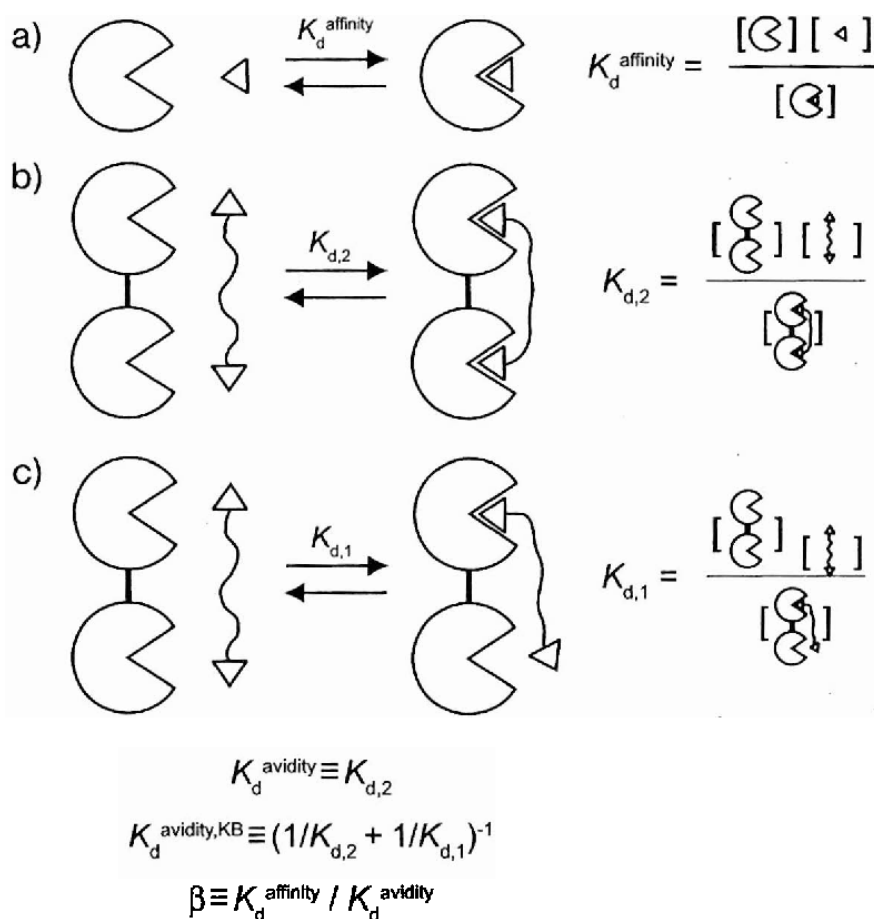


Figure 2.4: Thermodynamic equilibria used for the definitions of affinity, avidity, and enhancement.

a) A monovalent ligand binds a monovalent receptor with a dissociation constant of  $K_d^{\text{affinity}}$ .

b) The oligovalent (here bivalent) ligand binds a receptor of the same valency with a dissociation constant of  $K_{d,2}$  for the equilibrium between the fully complexed receptor and free receptor and ligand.

c) The bivalent receptor can also bind the bivalent ligand with only one receptor-ligand interaction; the complex has a dissociation constant of  $K_{d,1}$ . The enhancement  $\beta$  is the ratio of the affinity to the avidity. In this case, the enhancement contains a contribution from a statistical factor of 2.

Figure taken from Krishnamurty et al. [6]. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

## 2. MULTIVALENCY

---

binding steps take place independently.

### The Adair Equation

The cooperative binding is quantified with the Adair equation. Given a macroscopic association constant  $K_i$ , for any protein with  $n$  ligand binding sites, the fractional occupancy can be expressed as

$$\bar{Y} = \frac{1}{n} \frac{K_I[L] + 2K_{II}[L]^2 + \dots + nK_n[L]^n}{1 + K_I[L] + K_{II}[L]^2 + \dots + K_n[L]^n}$$

where  $[L]$  is the ligand concentration and each  $K_i$  is a combined association constant, describing the binding of  $i$  ligand molecules. The Adair equation and the Langmuir Hill plot have a connection, as Stefan and le Novère describe: "By combining the Adair treatment with the Hill plot, one arrives at the modern experimental definition of cooperativity. The resultant Hill coefficient, or more correctly the slope of the Hill plot as calculated from the Adair Equation, can be shown to be the ratio between the variance of the binding number to the ratio of the binding number in an equivalent system of non-interacting binding sites. Thus, the Hill coefficient defines cooperativity as a statistical dependence of one binding site on the state of other site(s)" [39].

### Degree of Binding

The degree of binding of a macromolecule  $M$  to a ligand  $L$  can be quantified by  $\nu = \frac{nK_a[L]}{1+K_a[L]} = \frac{[L]_{total} - [L]}{[M]_{total}}$ . Here, it is assumed that the macromolecule can host  $n$  ligands that have one binding site each. For the monovalent case, it holds that  $\nu_1 = \frac{[L]}{K_d + [L]}$ . For a detailed derivation and proof see Zumbansen [40]. This concept is interesting for bindings of one macromolecule to several ligands such as the haemoglobin-oxygen example. Because of this limitation, this concept is not useful for one-to-one stoichiometries and multivalent ligands.

## 2.4 Thermodynamic and Kinetic Studies of Multivalency

Both enthalpic and entropic behavior influence the multivalent binding mechanisms. The degree of influence is still a piece of research. As Goodrich and Kugel describe, "kinetic studies measure changes in the concentrations of reactants and/or products that occur in reactions over time. To study the kinetics of a binding reaction, the equilibrium between free and bound species is perturbed and changes in their concentrations are measured over time as the reaction approaches a new state of equilibrium" [36].

In order to determine general association and dissociation constants there are several techniques such as fluorescence spectroscopy, total internal reflection fluorescence spectroscopy, microscopy, and quartz crystal microbalance measurements. Atomic force microscopy (ATF) is a method to measure binding strength. For quantifying thermodynamic and kinetic parameters, surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) are the methods of choice, see [16] and references therein. However, SPR is better suited to determine binding constants for monovalent ligands due to the often heterogeneous binding sites of multivalent ligands [41]. ITC is better suited for complex formation studies than SPR and fluorescence spectroscopy in terms of cost and ease of use [42], therefore the thesis is focused on ITC. The most important thermodynamic quantities are entropy and enthalpy as discussed in the next two sections.

### Entropy

Entropy  $S$  of a macro state depends on the number of possible micro states  $i$  and their respective probabilities  $p_i$ . With the Boltzmann constant  $k_B = 1.38065 \cdot 10^{-23} J/K$ ,  $S$  is defined as

$$S = -k_B \sum p_i \ln p_i.$$

Generally, with every (additional) binding, the system loses degrees of freedom and therefore entropy. Often it is referred to as disorder or chaos. A monovalent binding eliminates both translational and rotational entropy. "The translational entropy of a molecule arises from its freedom to translate independently through

## 2. MULTIVALENCY

---

space; the value of  $\Delta S_{trans}$  is related to the logarithm of its mass  $M$  ( $\Delta S_{trans} \propto \ln M$ ), and inversely to the logarithm of its concentration  $\Delta S_{trans} \propto \frac{1}{\ln[L]}$ . The rotational entropy,  $\Delta S_{rot}$ , arises from the freedom of the particle to rotate around all three of its principle axes, and is related logarithmically to the product of its three principle moments of inertia  $I_x, I_y$  and  $I_z$ ,  $\Delta S_{rot} \propto (I_x I_y I_z)$ . The values of  $\Delta S_{trans}$  and  $\Delta S_{rot}$  for a particle are, therefore, only weakly (logarithmically) dependent on its mass and dimensions” [7]. In the bivalent setting, the first binding eliminates translational entropy and the second binding the rotational entropy. Further, in multivalent cases it must be differentiated between one or more entitites of ligand binding to one multivalent receptor. Figure 2.5 illustrates these differences. In theory, one must distinguish between rigid (conformational entropy  $\Delta S = 0$ ) and flexible linking groups ( $\Delta S \neq 0$ ). However, in reality most linking groups are somewhat flexible.

### Enthalpy

Enthalpy  $H$  quantifies the heat content of a chemical system. Thus,  $\Delta H$  is the total energy evolved or absorbed by a chemical reaction at constant pressure and temperature. It only depends on the unbound and bound state of the system [34]. If the system absorbs heat energy, enthalpy increases. Inversely, it decreases if the system emits heat. Ligand-receptor bindings produce heat, whereas unbinding absorbs heat.

Gibb’s free energy difference  $\Delta G$  depends on the enthalpic and entropic components, as well as temperature  $T$  [7].

$$\Delta G = \Delta H - T\Delta S.$$

In the multivalent setting there are several sources of enthalpic penalties  $\Delta H^{strain}$ . Firstly, if the spacer is rigid and not all the receptor-ligand pairs can reach each other, some enthalpic potential is forgone. This setting is rather unrealistic. Secondly, there may be unfavourable angles between the bonds that may hinder adjacent ligand-receptor pairs to reach each other. Thirdly, the receptors’ selectivity play a great role - adding enthalpic penalty by not accepting any random ligand. These penalties can partly be overcome by spacer, scaffold and linker

## 2.4 Thermodynamic and Kinetic Studies of Multivalency

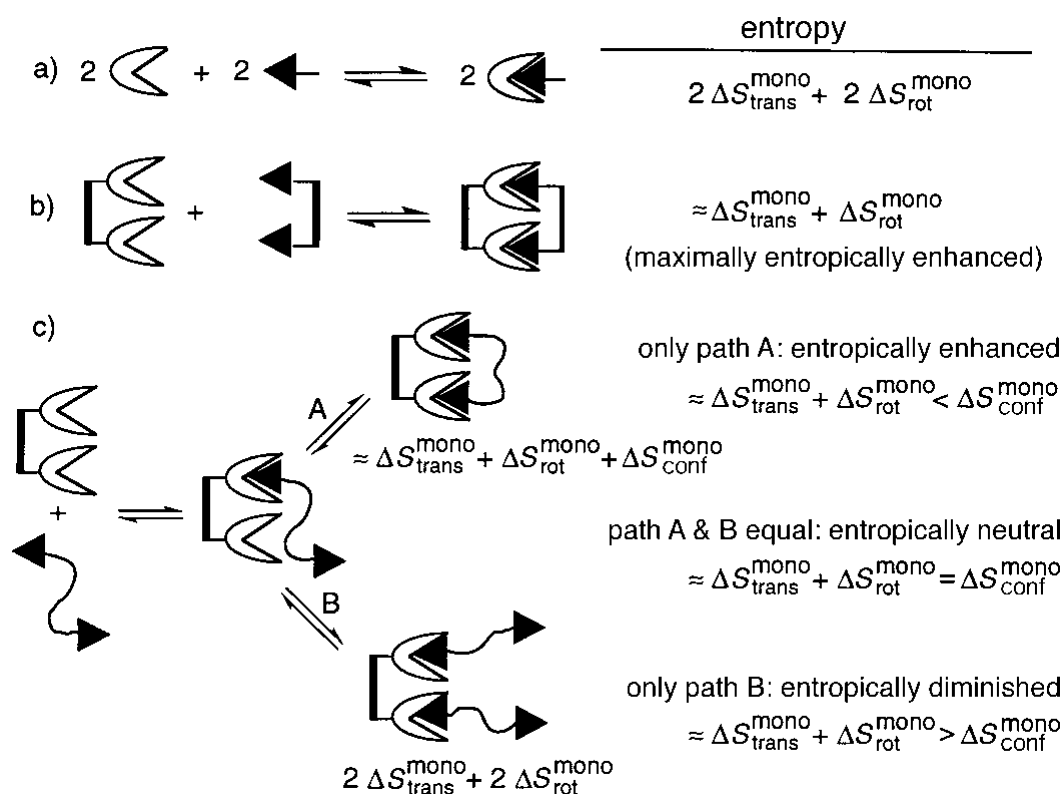


Figure 2.5: Relationships among translational, rotational, and conformational entropies for a bivalent system with a rigid and flexible linking group.

a) The total entropic cost of association of two monovalent receptors with two monovalent ligands is  $2\Delta S_{trans} + 2\Delta S_{rot}$ .

b) With a rigid linking group and the correct spacing to match the two receptor and ligand sites, the entropic cost of assembling both bivalent species is  $\Delta S_{trans} + \Delta S_{rot}$ .

c) Here it is distinguished between the conformational cost being less (only path A), equal (path A and B) or higher (path C) than the total translational and rotational cost.

Figure taken from Mammen et al. [7]. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

## 2. MULTIVALENCY

---

design. Theoretically, it holds that  $\Delta H^{multi} = n\Delta H^{mono}$  with  $n$  the number of binding sites. However, due to the strains mentioned, it is hardly ever reached. With favorable molecular design, the bivalent enthalpic balance may become  $\Delta H^{bi} = 2\Delta H^{mono} + \Delta H^{strain} + \Delta H^{spacer}$  with  $\Delta H^{spacer} \leq -\Delta H^{strain}$  [34]. The penalties spacer rigidity and distortion  $\theta$  from the system's most stable conformation and their effects on enthalpy are depicted in Figure 2.6.

Entropy and enthalpy can partly compensate each other in terms of affinity of multivalent bindings. A reaction is favored if entropy increases. It is also favored if enthalpy decreases [43].

The enthalpic considerations so far were of rather theoretic nature. In the following, enthalpy quantification is discussed. As titrations are one of the most common tools to measure the enthalpy of a complex, the next subsection is focusing on the ITC method.



## 2.4 Thermodynamic and Kinetic Studies of Multivalency

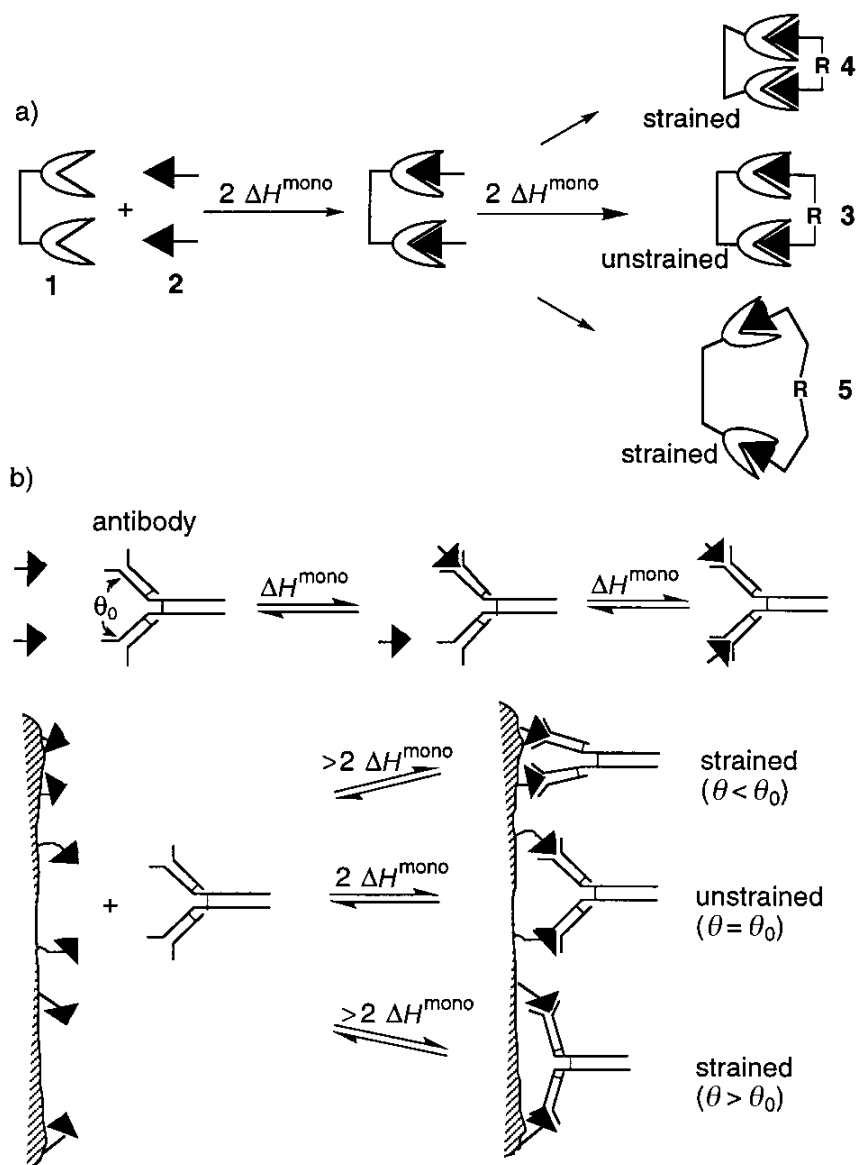


Figure 2.6: Enthalpy of binding for multivalent interactions.

a) Possible binding modes of monovalent and divalent ligands to a divalent receptor. If the two receptor sites are independent and non interfering, the binding of two monovalent ligands occurs with twice the enthalpy of binding one ligand. 1: bivalent receptor, 2: two monovalent ligands, 3: rigid group spacer R fits the spacing of 1, 4: R is too short for the spacing of 1, 5: R is too long to fit the spacing of 1. 4 and 5 do not hinder binding per se, but lead to a strained conformation.

b) The binding of a divalent antibody to ligand present at variable densities on the surface may be enthalpically diminished due to strain induced by distortion of the antibody from its most stable conformation ( $\theta_0$ ).

Figure taken from Mammen et al. [7]. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

## 2. MULTIVALENCY

---

### 2.4.1 Isothermal Titration Calorimetry

In general, titrations are designed such that the ratio of the two components are gradually changed and their behavior is observed until the equilibrium is reached. The concentration of one component is fixed, while the concentration of the counterpart is increased and their equilibrium is observed.

Specifically, the Isothermal Titration Calorimetry (ITC) setup consists of two adiabatic cells: a sample or titration cell and a reference cell with a buffer solution such as water, as depicted on the left hand side of Figure 2.7. In the sample cell a certain receptor concentration is fixed on a plate and remains unchanged. At regular time intervals, a solution with a known ligand concentration is injected from a syringe. For theoretical aspects and specific applications of ITC see [44] and [45] respectively. "In the absence of a reaction, the feedback power will be constant at the resting baseline value. Exothermic reactions will temporarily decrease and endothermic reactions temporarily increase feedback power" [37]. Complex formation produces heat in the sample cell. A thermostat regulates the temperature in the titration cell in order to keep it equal to the reference cell temperature. The data output of an ITC experiment is a thermogram, which records the time-course of this compensatory power  $Q$  required to maintain a constant temperature differential between the titration cell and the reference cell. As shown in the upper right diagram of Figure 2.7, the heat peaks show a sigmoidal shape with a decreasing tail as biomolecules reach their saturation. The enthalpy change is the total heat effect:  $\Delta H = H_2 - H_1 = Q_{tot}$ . If  $\Delta H > 0$ , the reaction is exothermic, otherwise endothermic. The entropy  $S$  describes the degrees of conformational freedom of the system. If  $\Delta S > 0$ , the reaction is irreversible, if it is zero, it is reversible. The sign of the Gibb's free energy change  $\Delta G$  indicates whether the reaction is spontaneous ( $\Delta G < 0$ ) or not ( $\Delta G > 0$ ). For  $\Delta G = 0$ , the system reached thermodynamic equilibrium [46].

The heat curve is described by  $Q = \sum_{k=1}^n \Delta H_k \Delta c_k V$  with  $Q$  being the heat evolved after titrating ligands to the reaction cell,  $\Delta H_k$  the complexation enthalpy of complex  $LR$ ,  $\Delta c_k$  the concentration change of species  $k$ ,  $V$  is the cell volume of the titration equipment, see Figure 2.7 upper right.

As described above, the shape of the binding isotherm is directly linked to the

## 2.4 Thermodynamic and Kinetic Studies of Multivalency

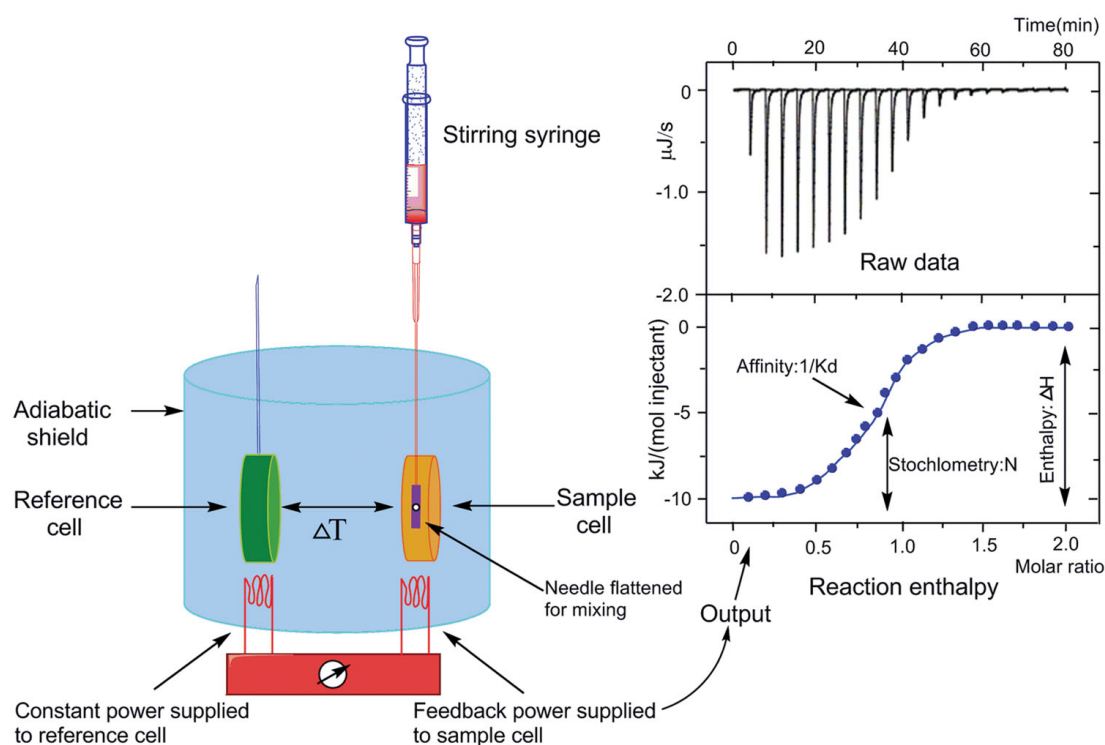


Figure 2.7: Basic principle of isothermal titration calorimetry. Schematic representation of the isothermal titration calorimeter (left) and a characteristic titration experiment (upper right) with its evaluation (lower right). In the upper right picture, the titration thermogram is represented as heat per unit of time released after each injection of the ligand into the protein (black), as well as the dilution of ligand into buffer (blue). In the lower right picture, the dependence of released heat in each injection versus the ratio between total ligand concentration and total protein concentration is represented. Circles represent experimental data and the line corresponds to the best fitting to a model considering  $n$  identical and independent sites. The syringe is inserted in the sample cell and a series of injections are made. Figure taken from Song et al. [47]. Licensed by Creative Commons Attribution 4.0 International Public License.

## 2. MULTIVALENCY

---

Wiseman parameter  $c = \frac{R_{tot}}{K_d}$ . "For very tight binding ( $c = \infty$ ), all added ligand is bound until saturation occurs so that a rectangular curve of height  $\Delta H$  is seen. For moderately tight binding with  $c$  values between 1 and 1000 the shape of the binding isotherms are very sensitive to small changes in  $c$  values. The intercept of these curves on the ordinate is no longer exactly equal to  $\Delta H$  but this parameter is still easily obtained by deconvolution from the total area under the curve and its shape. Very weak binding (cf.  $c = 0.1$ ) yields a nearly horizontal trace which again, like very tight binding, yields little information of the precise value of  $K$ " [37]. Further, the Wiseman parameter is linked to the equilibration time  $\theta$  [48]. The first injection leads to a minimum equilibration time  $\theta_1$  and reaches its maximum at mid-titration time  $\theta_{max}$ . Their ratio is approximated by the Wiseman parameter by the relationship  $\theta_{max}/\theta_1 \approx \sqrt{c}/2$ . The following relationship helps approximating  $k_{off}$ :  $\theta_{max} \approx 2k_{off}^{-1}/\sqrt{c}$ .

ITC simultaneously determines the thermodynamic parameters: binding affinity  $K_a$ , enthalpy changes  $\Delta H$ , and the binding stoichiometry  $n$ . From the isotherm, see Figure 2.7 lower right, the binding constant  $K_a$  can be deduced. The binding constant  $K_a$  is the curve slope at the inflection point. The binding mechanism denotes  $\Delta H$ , i.e. the maximum enthalpy change. The stoichiometry  $n$  is the value on the x-axis where the curve has  $\Delta H/2$  on the y-axis.<sup>1</sup> From the relationship  $\Delta G = -RT \ln K_a = RT \ln K_d = \Delta H - T\Delta S$  we can calculate Gibb's free energy  $\Delta G$  and the entropy change  $\Delta S$  of the equilibria with the absolute temperature  $T$  and the gas constant  $R = 8.3145 J/molK$  [37].

Another method to gain the complexation enthalpies is the Van't Hoff method, where multiple titrations are performed at varying temperatures [16], but the equilibrium constants are not only temperature-dependent [17]. That is why in the subsequent work only the common ITC experiment is considered.

---

<sup>1</sup>For sigmoidal curves of the shape  $f(x) = \frac{e^{ct+d}}{1+e^{ct+d}}$  with constant  $c$  and  $d$ , the inflection point  $x_i$  is equal to  $x$  such that  $f(x) = (f_{max} - f_{min})/2$ . However, depending on the binding affinity, the heat curves do not necessarily have a sigmoidal shape and an inflection point.

### 2.4.2 Kinetic ITC

Kinetic ITC (kinITC) is a relatively new method on how to obtain kinetics in addition to the thermodynamic information from ITC experiments. The biggest advantage is that kinetic binding and unbinding rates  $k_{on}$  and  $k_{off}$  can be obtained. Its pioneers are Burnouf, Dumas and co-workers, and Vander Meulen and Butcher [18; 48; 49]. In the classical ITC measurements, only the peaks of the heat curve were considered. Apart from that, it is important how fast the curve decreases from the peak back to the baseline. Therefor the area under each peak is integrated until the baseline is reached again. In fact, the apparatus measures the heat power  $P$  in [ $\mu cal/s$ ] and not directly the heat  $Q$ . They are related by the time derivative  $P(t) = \frac{\partial Q}{\partial t}$ . An equilibration time curve (ETC) has to be determined, i.e. where each injection starts and ends effectively. For each titration experiment a weighted average  $k_{on}$  value of the single titrations is determined. The change of the complex concentration over time is

$$\frac{\partial[LR]}{\partial t} = -\frac{\partial\lambda}{\partial t} = k_{on}([L]_{eq} + \lambda)([R]_{eq} + \lambda) - k_{off}([LR]_{eq} - \lambda)$$

with  $\lambda$  the difference between the equilibrium concentration of the complex and its current concentration  $\lambda = [LR]_{eq} - [LR]$ .  $\lambda$  cannot be determined analytically, but together with  $k_{on}$  in a least squares approximation. The heat evolution function is [49]

$$Q_{ev} = -\Delta H V_0 (\lambda_{t=0} - \lambda).$$

The equilibrium constants are determined by the Van't Hoff equation:

$$\frac{\partial \ln K_a}{\partial T} = \frac{\Delta H}{RT^2}$$

with  $R$  the gas constant and  $T$  the temperature.  $\Delta H$  is known from the ITC experiment. The binding constants  $k$  are temperature dependent and are determined with the Arrhenius equation:

$$k_{on} = A e^{\frac{-E_a}{k_B T}}$$

## 2. MULTIVALENCY

---

with  $A$  some pre-exponential factor depending on the type of chemical reaction,  $E_a$  the activation energy,  $k_B$  the Boltzmann constant and  $T$  the temperature.  $k_{off}$  is obtained by the relationship

$$k_{off} = \frac{k_{on}}{K_a} \quad (2.2)$$

Since these binding rates depend on the temperature, usually several ITC experiments are run at different temperatures and an average is used.

# Chapter 3

## Clustering Multivalent Binding Data

The third chapter is dedicated to the mathematical fundamentals of Markov processes and clustering. These concepts will be applied to experimental data of Isothermal Titration Calorimetry (ITC), the only measurement technique providing the complexation enthalpies [17] in the next chapter. Therein, it is explained how thermodynamic information is gathered from the output. Then, the assessment is enhanced in order to gain kinetic information. From a mathematical perspective, there is even more one can read out of the output. Setting up a theoretical rate matrix, discretizing it and clustering it, gives rise to the individual binding and unbinding coefficients.

### 3.1 Mathematical Preliminaries

Chemical reactions such as binding and unbinding of receptors and ligands can be modeled as rare stochastic events. In the following sections, we define the necessary operators and functions. The ligand-receptor binding interactions are Markov processes. Let  $(X_t)_{t \in T}$  be a stochastic process,  $(E, \Sigma)$  a measurable space for a given set  $E$  and  $\Sigma$  is a  $\sigma$ -algebra on  $E$ .  $(\Omega, \mathcal{A}, \mathbb{P})$  is the probability space.  $\Omega$  is the sample space,  $\mathcal{A}$  the event space and  $\mathbb{P}$  is the probability function.

**Definition 1** (Stochastic Process). [50]

### 3. CLUSTERING MULTIVALENT BINDING DATA

---

A stochastic (or random) process  $(X_t)_{t \in \mathbb{T}}$  is a collection of random variables  $X_t : \Omega \rightarrow E$  defined on a common probability space  $(\Omega, \mathbf{A}, \mathbb{P})$  and indexed by the elements of a parameter set  $T$ . Element  $t \in T$  is usually thought of as time.

**Definition 2** (Stochastic Transition Function). [10]

A stochastic transition function  $p : \mathbb{T} \times E \times \Sigma \rightarrow [0, 1]$  gives the probability for a transition to a state  $A$

$$p(t, x, A) = \mathbb{P}[X_{t+s} \in A | X_s = x] \quad (3.1)$$

and fulfills the following properties for all  $s, t \in \mathbb{T}, x \in E$  and  $A \subset E$ :

- i)  $x \mapsto p(t, x, A)$  is measurable  $\forall t \in \mathbb{T}$  and  $A \in \Sigma$
- ii)  $A \mapsto p(t, x, A)$  is a probability measure  $\forall t \in \mathbb{T}$  and  $x \in E$
- iii)  $p(0, x, E \setminus x) = 0 \quad \forall x \in E$
- iv) the Chapman-Kolmogorov equation

$$p(t + s, x, A) = \int_E p(t, x, dz) p(s, z, A) \quad (3.2)$$

holds  $\forall t, s \in \mathbb{T}, x \in E, A \subset \Sigma$ .

**Definition 3** (Markov Process). A stochastic process  $(X_t)_{t \in \mathbb{T}}$  on a state space  $E$  is a Markov process if equation 3.1 is fulfilled for all  $s, t \in \mathbb{T}$  and  $A \in \Sigma$ .

An invariant measure, also called stationary measure, with respect to a Markov process describes the situation that the probability to be in a certain subset is equal to the probability to get into this very subset for any fixed transition time.

**Definition 4** (Invariant Probability Measure). [51]

A Markov process  $X_t$  admits an invariant probability measure  $\mu$ , or  $\mu$  is invariant with respect to the Markov process, if

$$\int_E p(t, x, A) \mu(dx) = \mu(A) \quad \forall t \in \mathbb{T}, A \in \Sigma.$$



**Definition 5** (Transient and Recurrent State). [52; 53]

Consider a Markov chain  $X_n : n \in N_0$  on state space  $E$  with transition matrix  $P$ .

A state  $i \in E$  is called recurrent if  $P_i[X_n = i \text{ for infinitely many } n] = 1$ .

A state  $i \in E$  is called transient if  $P_i[X_n = i \text{ for infinitely many } n] = 0$ .

In the discrete setup, a recurrent state is sure to be returned to. Irreducible chains on finite space always satisfy this notion.

A Markov process is memoryless, i.e. every state depends only on the current state. If  $E$  consists of a finite number  $s$  of states, we can construct a transition probability matrix  $P \in \mathbb{R}^{s \times s}$ . A stochastic process is in equilibrium if the detailed balance condition is fulfilled.

**Definition 6** (Detailed Balance Condition). [51]

A stochastic transition matrix  $Q$  fulfills the detailed balance condition if

$$\pi_k Q_{kl} = \pi_l Q_{lk}.$$

$\pi \in \mathbb{R}^n$  denotes the stationary distribution and satisfies  $\pi^T = (\pi P)^t$ .

In other words, in equilibrium one cannot tell if a process is going forward or backward in time. Thus, the detailed balance condition is directly linked to the reversibility property of a Markov process. A Markov process  $(X_t)$  is reversible with respect to an invariant probability measure  $\mu$  if the detailed balance condition holds. If it holds and the state space is finite, then all the operator's eigenvalues are real [54]. Considering continuous space and time, the reversibility definition can be written as an integral.

**Definition 7** (Reversibility). [10]

Let  $(X_t)_{t \in \mathbb{T}}$  be a Markov process with invariant probability measure  $\mu$ . Then  $X_t$  is reversible with respect to  $\mu$  if

$$\int_A p(t, x, B) \mu(dx) = \int_B p(t, x, A) \mu(dx)$$

$\forall t \text{ in } \mathbb{T} \text{ and } A, B \in \Sigma$ .

## 3.2 The Transfer Operator

A transfer operator propagates probability densities across time. The time can be considered forward or backward. Detailed overviews of transfer operators and their properties can be found in Schütte and Sarich [10]. Nielsen showed that Markov operators and transfer operators are identical [20]. Further, the class of adjoint transfer operators can be characterized by generalized Koopman operators.

**Definition 8** (Markov Operator). [20]

A linear operator  $P : L^1(\mu) \rightarrow L^1(\mu)$  satisfying

- $Pf \geq 0 \quad \forall f \geq 0, f \in L^1(\mu)$
- $\|Pf\|_1 = \|f\|_1 \quad \forall f \geq 0, f \in L^1(\mu)$

is called a Markov operator.

We consider Banach spaces of equivalence classes of measurable functions weighted by the invariant measure  $\mu$ :

$$L^r(\mu) = \{f : E \rightarrow \mathbb{C} : \int_E |f(x)|^r \mu(dx) < \infty\}$$

with  $1 \leq r < \infty$  and

$$L^\infty(\mu) = \{f : E \rightarrow \mathbb{C} : \mu\text{-ess sup}_{x \in E} |f(x)| < \infty\}$$

with the corresponding norms  $\|\cdot\|_r$  and  $\|\cdot\|_\infty$ , respectively [51]. Due to Hölder's inequality it holds that  $L^r(\mu) \subset L^s(\mu) \quad \forall \quad 1 \leq s \leq r < \infty$  [51]. We are interested in transfer operators  $P^t$  propagating sub-ensembles relative to the weighting with  $\mu$  in time according to

$$\nu_o \mapsto \nu_t = P^t \nu_o$$

The following transfer operators are well-defined on Banach spaces  $L^r(\mu)$ ,  $1 \leq r \leq \infty$ . Specifically,  $L^2(\mu)$  is a Hilbert space with scalar product

$$\langle f, g \rangle_\mu = \int_E f(x)g(x)\mu(dx).$$

## 3.2 The Transfer Operator

---

**Definition 9** (Forward Transfer Operator). [10; 51]

The forward transfer operator (or propagator)  $P^t : L^r(\mu) \rightarrow L^r(\mu)$  with  $t \in \mathbb{T}$  and  $1 \leq r \leq \infty$  is defined by

$$\int_A P^t v(y) \mu(dy) := \int_E v(x) p(t, x, A) \mu(dx) \quad (3.3)$$

for measurable  $A \in \Sigma$  and  $v \in L^r(\mu)$ .

**Remark.** Because  $\mu$  is invariant, the characteristic function  $\mathbb{1}$  of the entire space is invariant under  $P^t$ , that is  $P^t \mathbb{1}_E = \mathbb{1}_E$ .  $P^t$  is a Markov operator and thus norm-conserving, that is  $\|P^t v\|_1 = \|v\|_1$ . Further, the operator conserves positivity, i.e.  $P^t v \geq 0$  if  $v \geq 0$ . The forward transfer operator propagates probability densities according to a given Markov process. Propagators of reversible Markov processes are well-defined for the Hilbert space  $L^2(\mu)$ , too.

**Definition 10** (Backward Transfer Operator). [10]

The backward transfer operator  $T^t : L^r(\mu) \rightarrow L^r(\mu)$  with  $t \in \mathbb{T}$  and  $1 \leq r \leq \infty$  is defined by

$$T^t u(x) = \int_E u(y) p(t, x, dy) \quad (3.4)$$

for every  $t \in \mathbb{T}$ .

**Remark.** Like for the propagator, for the backward transfer operator it holds  $T^1 \mathbb{1}_E = \mathbb{1}_E$ .

**Definition 11** (Self-adjointness). [24]

An operator  $O$  on  $L^2(\mu)$  is self-adjoint if for all  $f, g \in L^2(\mu)$  it holds

$$\langle f, Og \rangle_\mu = \langle Of, g \rangle_\mu.$$

**Theorem 1.** [51]

Let  $T : L^2(\mu) \subset L^1(\mu) \rightarrow L^2(\mu)$  be the transfer operator corresponding to the Markov process  $(X_t)_{t \in \mathbb{T}}$ . Then  $T$  is self-adjoint with respect to the scalar product  $\langle \cdot, \cdot \rangle_\mu$  in  $L^2(\mu)$  if and only if  $(X_t)_{t \in \mathbb{T}}$  is reversible.

### 3. CLUSTERING MULTIVALENT BINDING DATA

---

Both operators are linked by the  $\mu$ -weighted scalar product:

$$\langle v, u \rangle_\mu := \int_E v(x)u(x)\mu(dx)$$

The backward transfer operator is adjoint to the forward transfer operator,  $(P^t)^* = T^t$ . Therefore, it holds

$$\langle P^t v, u \rangle_\mu = \langle v, T^t u \rangle_\mu. \quad (3.5)$$

The backward transfer operator is also referred to as Koopman operator in the literature. It is the solution operator of the backward Kolmogorov equation:

$$K_t : L^\infty(\mathbf{X}) \rightarrow L^\infty(\mathbf{X}) : \quad K_t f(x) = \int_{\mathbf{X}} p_\tau(x, y) f_t(y) dy = \mathbf{E}[f_t(\mathbf{X}_{t+\tau}) | \mathbf{X}_t = x]$$

and describes the evolution of observables [55].

#### 3.2.1 Discrete Transfer Operators

In order to apply the concepts of transfer operators and generators which act on continuous time and space, we need to discretize both. One possible discretization is the Galerkin projection.

By discretizing the transfer operator we get a transition probability matrix.

**Definition 12** (Transition Probability Matrix). [54]

*The matrix  $P = P(x, y)$ ,  $x, y \in X$  is called a Markov transition matrix if*

$$P(x, y) \geq 0, \quad \sum_{z \in X} P(x, z) = 1 \quad x, y \in X.$$

The entries  $P_{ij}$  denote the probability to jump from state  $i$  to state  $j$  in within one time step  $t$ .

### Example

The transition probability matrix

$$P = \begin{pmatrix} 0.8 & 0.2 & 0 \\ 0.1 & 0.7 & 0.2 \\ 0 & 0.5 & 0.5 \end{pmatrix}$$

describes the probabilities of a molecule that can be in three different states: unbound, partly bound and fully bound to another protein. Suppose that in the beginning, the system is unbound and thus the state vector  $x$  is  $x_0 = (1, 0, 0)^T$ . How does the equilibrium distribution look like? Since  $x_t = x^T P$  and as  $t \rightarrow \infty$  the molecule is most likely partly bound with a little higher propensity to dissociate than associate further, i.e.  $x_{1000}^T = (0.26, 0.53, 0.21)$ .

In the example above, the molecule had a certain probability to be in all of the three states. The question, if all states of the state space can be visited when starting at a specific position at a specific time, is captured by ergodicity, a concept coined by Boltzmann.

### Definition 13 (Ergodicity). [10]

Let  $(X_t)_{t \in \mathbb{T}}$  be a Markov process with invariant measure  $\mu$ . Then  $X_t$  is ergodic with respect to  $\mu$  if for all functions  $v : E \rightarrow \mathbb{R}$  with  $\mathbb{E}_\mu(|v|) = \int_E |v| \mu(dx) < \infty$  it holds that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T v(X_t) dt = \int_E v(x) \mu(dx).$$

In other words, ergodicity means that the space average equals the time average if time goes to infinity. All subspaces can be visited by the stochastic process. Huisinga connects ergodicity for Markov processes to irreducibility. In his thesis [51] he puts "that it is possible to move from (almost) every state to every 'relevant' subset within a finite time". This leads us directly to communication and irreducibility. Communicating states mean that state  $x$  leads to state  $y$ . That means,  $\sum_{n=0}^{\infty} P^n(x, y) > 0$  and  $\sum_{n=0}^{\infty} P^n(y, x) > 0$  [54].

### Definition 14 (Irreducible Spaces and Absorbing Sets). [54]

If for a set  $C$  and a space  $X$  it holds that  $C(x) = X$  for some  $x$ , then we say that

### 3. CLUSTERING MULTIVALENT BINDING DATA

---

$X$  is irreducible.

We say  $C(x)$  is absorbing if  $P(y, C(x)) = 1 \quad \forall y \in C(x)$ .

**Semigroup Property:** It holds  $P_{\tau+\sigma} = P_\tau P_\sigma$  and  $K_{\tau+\sigma} = K_\tau K_\sigma$  for  $\tau, \sigma \geq 0$ , i.e. these operators describe time-stationary Markovian dynamics [55].

The transition matrix  $P(k)$  of the  $k^{\text{th}}$  time step meets the semi group property given by [56]:

$$P(k) = (P(1))^k = P^k.$$

**Definition 15** (Infinitesimal Generator). [24]

For the semigroup of transfer operators  $\mathcal{P} : L^r(\mu) \rightarrow L^r(\mu)$  with  $t \in \mathbb{T}$  and  $1 \leq r \leq \infty$ ,  $\mathcal{D}(\mathcal{Q})$  is defined as the set of all  $f \in L^r(\mu)$  such that the limit

$$\mathcal{Q}f := \lim_{t \rightarrow \infty} \frac{\mathcal{P}^t f - f}{t}$$

exists. Then the operator  $\mathcal{Q} : \mathcal{D}(\mathcal{Q}) \rightarrow L^r(\mu)$  is called the infinitesimal generator of  $\mathcal{P}^t$ .

The propagator  $P$  propagates probability densities over time, while the infinitesimal generator  $\mathcal{Q}$  describes the Markov process in infinitesimal time. By discretizing  $\mathcal{Q}$ , we get the transition rate matrix  $Q$ .

**Definition 16** (Rate Matrix). [57]

The rate matrix  $Q$  with entries  $Q = q_{ij}$  is also called intensity matrix or infinitesimal generator and has the following properties:

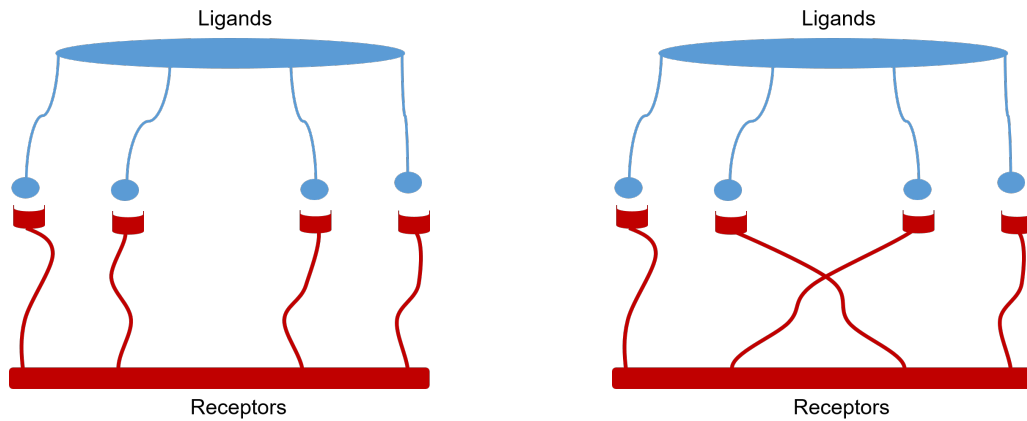
- the off-diagonal entries are positive:  $q_{ij} \geq 0$  for  $i \neq j$
- each column sum is 0 :  $\sum_i q_{ij} = 0$  for all  $j$
- each diagonal element is the negative sum of the column entries:  $q_{ii} = -\sum_{i \neq j} q_{ij}$ .

**Remark.** The rate matrix  $Q$  and the probability matrix  $P$  are connected by the matrix exponential

$$P_t = \exp(tQ) \tag{3.6}$$

### 3.2.2 Defining the Rate Matrix

In the following subsection, the multivalent binding setting is going to be translated into a rate matrix. This procedure is described in Åberg et al. [58]. First, some necessary assumptions are made to model the binding process as simplified as possible, yet realistically. In the model we assume that the molecule backbone and spacers are somewhat rigid, i.e. the radius of spacer movement is bounded and not every spacer arm can bind with any binding pocket of the counterpart, as illustrated by Figure 3.1.



Ligand-receptor binding with rigid spacers

Ligand-receptor binding with flexible spacers

Figure 3.1: In this thesis it is assumed that the spacers of ligands and receptors are rigid, respectively, as depicted in the left image. No cross-linking like in the right image can occur.

This fact limits the possible combinations in that way that a single binding site can either bind with a counterpart or with its left or right neighbor (if it is unbound of course). The (transition) rate matrix, is a square matrix with as many rows and columns as there are different states in the state space. It is closely linked to the probability matrix which is defined first.

If the number of binding sites on the receptor and ligand respectively differ, we will use the minimum  $n = \min(n_{ligand}, n_{receptor})$  like in [16]. As an example, let us assume a bivalent receptor with binding sites  $A$  and  $B$  interacting with a bivalent ligand with binding sites 1 and 2. There are four different possibilities to have

### 3. CLUSTERING MULTIVALENT BINDING DATA

---

one binding and two possibilities for two bindings. Further, for each receptor and ligand pair, there is only one situation without any bindings. Altogether, there are seven different states in the bivalent case, as depicted in Figure 3.2.

The order of the states is arbitrary. The order used below is listed in Table 3.1.  $[LR^1]$  denotes the total concentration of singly bound ligand-receptor complex and comprises the sum of the conformational states  $II, \dots, V$ .  $[LR^2]$  denotes the total concentration of the doubly bound ligand-receptor complex and comprises the sum of states  $VI$  and  $VII$ .

State Number	Combination	Number of Bindings
I	1- 2-	0
II	1 - A 2-	1
III	1 - B 2-	1
IV	1- 2 - B	1
V	1- 2 - A	1
VI	1 - A 2 - B	2
VII	1 - B 2 - A	2

Table 3.1: Order of the seven possible states for bivalent bindings.

How many possible states  $Z(n)$  do exist in the  $n$ -valent case? There is always one state for the unbound case and  $n$  states for the fully bound case. For the  $n - 2$  states in between, i.e. singly bound, doubly bound, ...,  $n - 1$  bound, it is a Bernoulli trial. For each of these 'partly bound states' there are  $n$  possibilities, how to allocate these Bernoulli trials. Per partly bound states there exist  $\sum_{i=1}^{n-1} \binom{n}{i}$  possibilities to bind. If the valency increases by 1, the state space size increases by  $1 + (n + 1) \sum_{i=1}^n \binom{n+1}{i} - n \sum_{i=1}^{n-1} \binom{n}{i}$  states.



### 3.2 The Transfer Operator

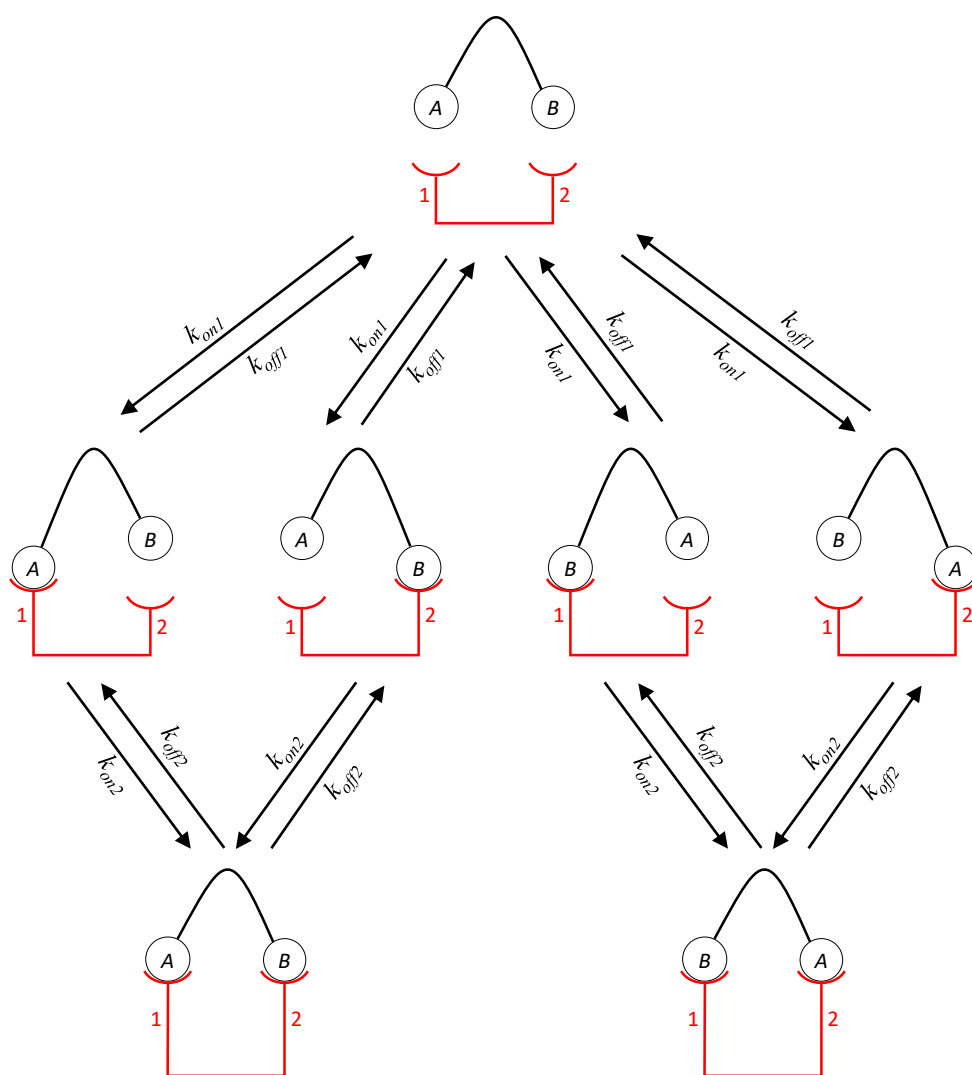


Figure 3.2: This scheme shows how all the seven possible states in a bivalent binding setting are connected by  $k_{on}$  and  $k_{off}$  rates.

### 3. CLUSTERING MULTIVALENT BINDING DATA

---

**Proposition:** the number of possible states for an  $n$ -valent ( $n \in \mathbb{N}_+$ ) system is:

$$Z(n) = 1 + n + n \sum_{i=1}^{n-1} \binom{n}{i}.$$

*Proof.* Let  $Z(n)$  be the statement  $1 + n + n \sum_{i=1}^{n-1} \binom{n}{i}$ .

If the valency  $n$  is increased by 1, there are altogether  $Z(n+1)$  different states with  $Z(n+1) = 2 + n + (n+1) \sum_{i=1}^n \binom{n+1}{i}$ .

This can be written as  $Z(n+1) = Z(n) + 1 + (n+1) \sum_{i=1}^n \binom{n+1}{i} - n \sum_{i=1}^{n-1} \binom{n}{i}$

$$\begin{aligned} Z(n+1) &= 2 + n + (n+1) \sum_{i=1}^n \binom{n+1}{i} \\ &\stackrel{!}{=} Z(n) + 1 + (n+1) \sum_{i=1}^n \binom{n+1}{i} - n \sum_{i=1}^{n-1} \binom{n}{i} \\ &= 1 + n + n \sum_{i=1}^{n-1} \binom{n}{i} + 1 + (n+1) \sum_{i=1}^n \binom{n+1}{i} - n \sum_{i=1}^{n-1} \binom{n}{i} \\ &= 2 + n + (n+1) \sum_{i=1}^n \binom{n+1}{i} \\ &= Z(n+1) \end{aligned}$$

□

Thus, in a bivalent case, we have  $Z(2) = 7$  different states, in the trivalent case  $Z(3) = 22$ ,  $Z(4) = 61$  etc. The number of states increases exponentially with the valency number as depicted in Figure 3.3.

In the following, the rate matrix is filled in the bivalent case and concluding it is shown how to fill the  $n$ -valent rate matrix. The left column represents the first binding: from unbound to singly bound. These rates depend on the first binding constant  $k_{on_1}$  and the ligand concentration that changes with time. In the case of ITC, the time is linked to the titration steps at certain time intervals. There is no direct way to get from unbound to doubly bound, therefore the last two entries in the first column are zero. The first row is filled in an analogous way. Going

## 3.2 The Transfer Operator

---

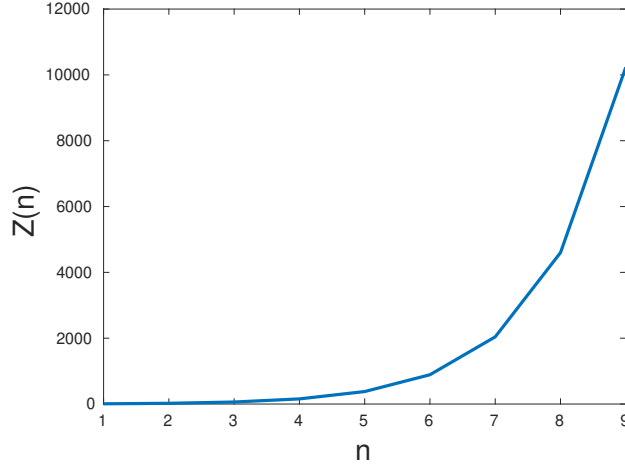


Figure 3.3: Number of micro states  $Z$  depending on the valency number  $n$ .

from singly bound to unbound, only the unbinding rate  $k_{off_1}$  matters. Again, there is no way to move from the fully bound state to unbound in one step, that is why the last two entries in the first row are zero. The submatrix representing the transition from singly bound to fully bound is of size  $2 \times 4$  and depends only on the  $k_{on_2}$  value. In accordance to the first unbinding event, that means moving from double bound to the singly bound state, the submatrix is of size  $4 \times 2$  and depends on the  $k_{off_2}$  value.

$$Q = \begin{pmatrix} d_1 & k_{off_1} & k_{off_1} & k_{off_1} & k_{off_1} & 0 & 0 \\ k_{on_1} \cdot c_L & d_2 & 0 & 0 & 0 & k_{off_2} & 0 \\ k_{on_1} \cdot c_L & 0 & d_2 & 0 & 0 & 0 & k_{off_2} \\ k_{on_1} \cdot c_L & 0 & 0 & d_2 & 0 & k_{off_2} & 0 \\ k_{on_1} \cdot c_L & 0 & 0 & 0 & d_2 & 0 & k_{off_2} \\ 0 & k_{on_2} & 0 & k_{on_2} & 0 & d_3 & 0 \\ 0 & 0 & k_{on_2} & 0 & k_{on_2} & 0 & d_3 \end{pmatrix} \quad (3.7)$$

with  $d_1 = -4k_{on_1} \cdot c_L$ ,  $d_2 = -k_{off_1} - k_{on_2}$  and  $d_3 = -2k_{off_2}$ .

The positions of the entries in these two submatrices depend on the order of the states as in Table 3.1. The order of the states could be different, but it has no effect on the final result. However, the entry positions have to be the transpose

### 3. CLUSTERING MULTIVALENT BINDING DATA

---

of the corresponding submatrix.

The rate matrix for the  $n$ -valent binding is only sketched schematically by the order of its submatrices.

$$Q = \begin{pmatrix} * & k_{off} & 0 & 0 & \dots & 0 \\ k_{on} \cdot c_L & * & X_1 & 0 & \dots & 0 \\ 0 & Y_1 & * & X_2 & \dots & 0 \\ 0 & \dots & \ddots & \vdots & & 0 \\ 0 & 0 & \dots & Y_2 & * & X_{n-1} \\ 0 & 0 & 0 & \dots & Y_{n-1} & * \end{pmatrix} \quad (3.8)$$

**Legend:**

$*$	diagonal matrix with the negative column sum such that the column sum is 0
$k_{off}$	row vector of length $n^2$ with entry $k_{off_1}$
$k_{on} \cdot c_L$	column vector of length $n^2$ with entry $c_L \cdot k_{on_1}$
$X_i \quad i = 1, \dots, n - 1$	sparse matrix of size $\binom{n}{i} \cdot n \times \binom{n}{i+1} \cdot n$ with entries $k_{off_2}$ to $k_{off_n}$ respectively. The position of these entries depend on the theoretic order of conformational states.
$Y_i \quad i = 1, \dots, n - 1$	sparse matrix of size $\binom{n}{i+1} \cdot n \times \binom{n}{i} \cdot n$ with entries $k_{on_2}$ to $k_{on_n}$ , respectively. The positions of the entries are the transpose of the respective $X_i$ submatrices.

The following subsection deals with the clustering of the rate matrix defined above using PCCA+.

### 3.3 Clustering with PCCA+

In the previous subsection it is shown how to discretize the operator practically by determining the number of states of a ligand-receptor binding system. The next step is a projection onto an invariant subspace. The ultimate goal is to filter the essence of this Markov process by clustering. One way is the PCCA+ clustering algorithm. PCCA+ stands for 'Robust Perron Cluster (Cluster) Analysis' and was developed by Weber et al. It is the more robust successor of the PCCA (Perron Cluster Cluster Analysis) algorithm, developed by Deuffhard et al. [59]. PCCA results in a crisp clustering which is applicable for decomposable Markov chains. For some applications such as molecular dynamics, the transition matrix cannot be decomposed completely because there exist transition states due to entropic and/or enthalpic energy barriers [60]. PCCA+ uses the simplex structure of the eigenvectors of the transition matrix. The core part is a Galerkin projection of the transfer operator on the invariant subspace.

**Definition 17** (Galerkin Projection). [10]

Let  $\chi = \{\chi_1, \dots, \chi_n\} \subset L^2(\mu)$  be a finite family of measurable functions. Let  $\chi_k$  be non-negative, linearly functions, summing up to unity, i.e.

$\sum_{k=1}^n \chi_k(x) = 1 \quad \forall x \in E$ . Then the Galerkin projection onto the associated finite-dimensional ansatz space  $D = \text{span}\{\chi_1, \dots, \chi_n\}$  by  $\Pi : L^2(\mu) \rightarrow D$  has the form

$$\Pi v = \sum_{k,j=1}^n (S^{-1})_{kj} \langle \chi_k, v \rangle_{\mu} \chi_j, \quad (3.9)$$

where  $S$  is a non-negative, symmetric square matrix with entries  $S_{kj} = \langle \chi_k, \chi_j \rangle_{\mu}$ .

$S$  is a Gramian matrix of a set of linearly independent functions and is thus invertible.

If  $\{\chi_1, \dots, \chi_n\}$  are the characteristic functions  $\{\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}\}$  belonging to a full partition of the state space, equation 3.9 becomes

$$\Pi v = \sum_{\mu(A_k)}^n \langle \chi_k, v \rangle_{\mu} \chi_k.$$

$\chi$  can be thought of as membership functions of the discrete spaces and  $S$  then

### 3. CLUSTERING MULTIVALENT BINDING DATA

---

refers to the respective overlap matrix of these membership functions. The discretized propagator  $\Pi P \Pi$  induces an approximate eigenvalue problem:

$$\Pi P \Pi = \lambda v.$$

Next we compute the eigenvectors  $X$  of  $P$  corresponding to eigenvalues close to one,  $PX = X\Lambda$ ,  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{n_C})$ ,  $1 \geq \lambda_i \geq 1 - \epsilon$  and the eigenvectors are normalized such that  $X^T D^2 X = I$  with  $I$  the identity matrix and  $D^2$  the diagonal matrix with the stationary distribution  $\pi$  on its diagonal  $D^2 = \text{diag}(\pi)$ .

The aim of the PCCA+ algorithm is to decompose the state space  $\Omega$  into almost invariant membership functions  $\chi_1, \dots, \chi_{n_C}$  with  $n_C$  the number of clusters

$$\chi_i : \Omega \rightarrow [0, 1].$$

The membership function form a partition of unity

$$\sum_{i=1}^{n_C} \chi_i = 1.$$

The result is a fuzzy clustering, i.e. each object is assigned to all clusters with certain probabilities [60]. To find the number of clusters there are three common approaches [21]:

1. lower boundary for the Perron cluster
2. spectral gap approach
3. minChi Indicator for Simplex Structure

However, for the present use case we can assume that  $n_C = 2$  because we are only interested in the two clustered states 'bound' and 'unbound'. Experimental results from the laboratory suggest that in fact there exist at least two clusters and hence, it is safe to assume that there are indeed two clusters.

PCCA+ is defined for any row-stochastic matrix as input such as a transition probability matrix of a discrete-time Markov chain [60]. In the following, PCCA+ is applied with the rate matrix instead of the transition probability matrix. The

rate matrix and the transition matrix are connected by the matrix exponential  $P = \exp(tQ)$ . This procedure is also valid as shown in the next proof.

**Corollary 1.1.** *The PCCA+ algorithm can be run with the rate matrix as well as with the transition matrix.*

*Proof.* It has to be shown that  $PX = X\Lambda = X \exp(\tau\Lambda)$  for time  $\tau$ .

With the matrix exponential  $\exp(X) = \sum_{k=0}^{\infty} \frac{X^k}{k!}$ , the left hand side of the equation can be expanded to

$$\begin{aligned}
 PX &= \underset{P=\exp(tQ)}{=} \exp(\tau Q)X \\
 &= (I + \tau Q + \frac{1}{2}\tau^2 Q^2 + \dots)X \\
 &= \tau(I + Q + \frac{1}{2}Q^2 + \dots)X \\
 &= \tau(X + QX + \frac{1}{2}Q^2X + \dots) \\
 &= \tau(X + X\Xi + \frac{1}{2}X\Xi^2 + \dots) \\
 &\underset{QX=X\Xi}{=} X \exp(\tau\Xi) \\
 &= X\Lambda
 \end{aligned}$$

$$\begin{aligned}
 PX &\underset{P=\exp(tQ)}{=} (I + Q\tau + \frac{(Q\tau)^2}{2} + \frac{(Q\tau)^3}{6} + \dots)X \\
 &\underset{Q=X\Lambda X^{-1}}{=} X + X\Lambda X^{-1}\tau + (X\Lambda X^{-1})^2 X \frac{\tau^2}{2} + (X\Lambda X^{-1})^3 X \frac{\tau^3}{6} + \dots \\
 &= X \left( I + \Lambda\tau + \Lambda^2 \frac{\tau^2}{2} + \Lambda^3 \frac{\tau^3}{6} + \dots \right) \\
 &= X \exp(\Lambda\tau)
 \end{aligned}$$

□

Thus, the matrix of eigenvalues of the transition and rate matrices respectively, are connected by the matrix exponential, too. It holds that  $\Lambda = \exp(\tau\Xi)$ . The rate matrix can therefore be used for a similar eigenvalue problem. Practically,

### 3. CLUSTERING MULTIVALENT BINDING DATA

---

the first step of PCCA+ is the computation of the eigenvalues of  $Q$  according to

$$QX = X\Xi$$

with  $X$  the matrix of all the eigenvectors and a diagonal matrix of eigenvalues  $\Xi = \text{diag}(\xi_1, \dots, \xi_{n+1})$ . The unique biggest real eigenvalue is called the Perron eigenvalue [60]. For transition probability matrices, the Perron eigenvalue is always 1, for rate matrices it is 0 due to the relationship (3.6). The biggest eigenvalues refer to the slow states and describe the metastabilities. Usually, there are several eigenvalues only slightly smaller than the maximum eigenvalue and then a spectral gap occurs. For the purpose of this work, only the two biggest eigenvalues are of interest because all the states will be clustered into two states only. The leading eigenvalue and its corresponding leading eigenvector are selected. The second eigenvector has to satisfy the criterion that the first entry and the respective last one have the maximum distance, to make sure that the two states *unbound* and *bound* are as distinct as possible. For regular matrices  $Q$  the clustering algorithm gives a membership matrix  $\chi$  by

$$\chi = XA$$

with  $A$  being a non-singular matrix of linear factors computed by the optimization process.  $A$  has to fulfill the following two conditions:

$$A(1, j) \geq - \sum_{i=2}^{n_C} x_i(l) A(i, j), \quad \forall j \in 1, \dots, n_C. \quad l \in 1, \dots, N \quad \text{positivity} \quad (3.10)$$

$$A(i, 1) = \delta_{i,1} - \sum_{j=2}^{n_C} A(i, j), \quad \forall i \in 1, \dots, n_C \quad \text{partition of unity} \quad (3.11)$$

with  $\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$  the Kronecker delta.

In theory there is an uncountable number of transformation matrices  $A$  [21]. To maximize metastability, the following objective function has been used to find  $A$



such that

$$[\text{trace} D_C^{-2} \chi^T D^2 P \chi] \rightarrow \max$$

with  $D_C^2 = \text{diag}(\chi^T \omega)$  with  $\omega$  the stationary distribution.

**Definition 18** (Stationary Distribution). *The stationary (or invariant) distribution of a Markov Chain with transition matrix  $P$  is some vector  $\pi$ , such that*

$$\pi P = \pi, \quad \sum_{i=1}^n \pi_i = 1$$

That means that the stationary distribution is a starting distribution that does not change if the Markov process continues. "The existence of an invariant measure prevents the probability mass from 'escaping to infinity'" [53].

In the bivalent application presented in the next section, the matrix  $\chi$  of membership functions is a  $7 \times 2$  matrix stating the degree of membership of each of the seven states to the unbound and bound states, respectively. Its row sum is always one. In [21] it has been proven that for the case  $n_c = 2$ , the algorithm presented therein provides a feasible unique solution. Finally,  $Q$  is Galerkin discretized weighted by the stationary distribution  $\pi$  into  $Q_c$  (clustered  $Q$ ) using  $\chi$ .

The PCCA+ algorithm only works for reversible Markov processes, showing real eigenvalues. Complex eigenvalues are a guarantee for non-reversibility. However, the opposite conclusion is not true. Non-reversible processes can also have real eigenvalues only. Non-reversible processes can be tackled with a generalization of PCCA+ (GenPCCA): using the Schur decomposition instead of the eigenvalue decomposition. For more details on GenPCCA, refer to [61]. No matter which decomposition of the rate matrix is used, the next step is to project this state space down on an invariant subspace. The result is the clustered rate matrix, also called stochastic coupling matrix.

**Definition 19** (Stochastic Coupling Matrix). [21]

*The stochastic coupling matrix  $Q \in \mathbb{R}^{n_c \times n_c}$ , which provides the transition probabilities between the transition rates between the conformations  $\chi_1, \dots, \chi_{n_c} : \Omega \rightarrow$*

### 3. CLUSTERING MULTIVALENT BINDING DATA

---

$[0, 1]$ , is defined as

$$Q_c = \left( \frac{\langle \chi_i Q^T \chi_j \rangle_\pi}{\langle \chi_i \rangle_\pi} \right)_{i,j=1,\dots,n_C}$$

where  $Q^T$  is the matrix exponential of the transfer operator. The trace of  $Q$  are the metastabilities of the conformations  $\chi_1, \dots, \chi_{n_C}$ .

In matrix notation this definition is

$$Q_c = (\chi^T \Pi \chi)^{-1} \chi^T \Pi Q \chi$$

How does this coupling matrix look like in practice? In [62] Berberan-Santos and Martinho describe a matrix formulation of kinetic rate constants. Recall how equation 2.3.2 showed the time derivative of the one bond complex in a micro perspective. Now we are interested in the time derivative of the complex concentration and the ligand concentration in the macro perspective.

$$\frac{\partial [R]}{\partial t} = -k_{on}[L][R] + k_{off}[LR]$$

and

$$\frac{\partial [LR]}{\partial t} = -k_{off}[LR] + k_{on}[L][R].$$

Thus, we have an ODE of the form

$$\begin{pmatrix} \frac{\partial [R]}{\partial t} \\ \frac{\partial [LR]}{\partial t} \end{pmatrix} = Q_c \begin{pmatrix} [L] \\ [LR] \end{pmatrix}.$$

Note that this setting is observed in a macro perspective.  $k_{on}$  is the overall binding rate and  $k_{off}$  is the overall unbinding rate.  $[LR]$  is the bound complex. Here, it does not play a role how many individual bonds there are exactly. Since we project on a two-dimensional subspace, our  $Q_c$  matrix is of size  $2 \times 2$  and

### 3.3 Clustering with PCCA+

---

according to the two ODE's stated above<sup>1</sup>, it must have the following entries:

$$Q_c = \begin{pmatrix} -k_{on} \cdot c_L & k_{off} \\ k_{on} \cdot c_L & -k_{off} \end{pmatrix}$$

By dividing the lower left matrix entry by the respective ligand concentration  $c_L$  of the titration step, we determine the overall binding rate  $k_{on}$ . It is an interesting finding that the evolution of the overall binding rate is in fact concentration dependent. So far this concentration dependence has not been acknowledged by the ITC community.

**This new method of incorporating the concentration dependence to the kinITC method will be called kinITC+.**

---

<sup>1</sup>Note, that one could also use the ODE  $\frac{\partial[L]}{\partial t} = -k_{on}[L][R] + k_{off}[LR]$  in lieu of  $\frac{\partial[R]}{\partial t}$ . Then  $Q_c$  changes to  $\begin{pmatrix} -k_{on} \cdot c_R & k_{off} \\ k_{on} \cdot c_R & -k_{off} \end{pmatrix}$  with  $c_R$  the receptor concentration per titration step. Technically, the result is the same.

### **3. CLUSTERING MULTIVALENT BINDING DATA**

---

# Chapter 4

## KinITC+

The theoretical framework for describing a multivalent binding process and the logic behind the clustering method PCCA+ have been demonstrated in Chapter 3. This chapter shows their application with ITC experimental data. The applications in Section 3.3 are based on Erlekam et al. [63] if not indicated differently.

### 4.1 Application of PCCA+ to Kinetic Binding Data

Our ligand-receptor interaction process is space and time continuous and needs to be discretized for numerical results. Therefore, all the possible states were determined and the resulting transition rate matrix was set up like described in subsection 3.2.2. The Markov process shall be projected on an invariant subspace. Since we are only interested if the complex is unbound or bound, this subspace is two-dimensional. For example in the bivalent case we project the  $7 \times 7$  rate matrix  $Q$  on a  $2 \times 2$  clustered matrix  $Q_c$ .

As a proof of concept, overall  $k_{on}$  rates from experimental ITC data sets were compared to the  $k_{on}$  rates obtained from the clustered rate matrices. All ITC measurements and subsequent kinITC data presented in the following were taken from Igde et al. [64]. In their paper Igde et al. synthesized a library of mannose-functionalized oligo(amidoamines) varying the valency of the ligands from mono- to decavalent, introducing different linkers between the mannose and the oligomer

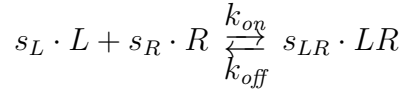
#### 4. KINITC+

---

backbone and varying the position of mannose ligands along the backbone. Binding of the glycooligomers to model lectin Concanavalin A (Con A) was studied by ITC performing normal titration where the glycooligomers were titrated into the sample cell containing the protein [64]. As described in the Supporting Information of [63], ITC data were then evaluated for thermodynamic information on the ligand-receptor complex formation and kinetic rate constants were extracted from the heat flow signals of the ITC isotherms following the work by Dumas et al. [18], Vander Meulen and Butcher [49], and Egawa et al. [19]. For each heat profile of the isotherm, the change in complex enthalpy  $\Delta H$  is determined with

$$P(t) = \frac{dQ}{dt} = -\Delta H V_0 \frac{d[LR](t)}{dt}$$

with  $V_0$  the starting volume of the solution in the sample cell and  $[LR](t)$  the ligand-receptor complex at time  $t$ .  $[LR]$  is the molar complex concentration and depends on the volume and concentration of the titrated ligands. From the equilibrium constant  $K_d$  which is the slope of the thermogram, the concentrations of the free ligands and free receptors as well as the complex concentration can be deduced for every injection. Based on these known concentrations, the reaction scheme



and the reaction rate equations can be set up.  $s_L, s_R$  and  $s_{LR}$  are the stoichiometric prefactors. For one-to-one bindings, they can be assumed to be 1, but are included in the following equations for the sake of generality. The rate equations with their respective boundary conditions are:

$$\begin{aligned} \frac{d[L](t)}{dt} &= -k_{on} \cdot s_L [L][R] + k_{off} \cdot s_L \cdot [LR] & [L](0) &= [L]_0 \\ \frac{d[R](t)}{dt} &= -k_{on} \cdot s_R [L][R] + k_{off} \cdot s_R \cdot [LR] & [R](0) &= [R]_0 \\ \frac{d[LR](t)}{dt} &= k_{on} \cdot s_{LR} [L][R] - k_{off} \cdot s_{LR} \cdot [LR] & [LR](0) &= [LR]_0 \end{aligned}$$

The starting concentrations of  $[L]$ ,  $[R]$  and  $[LR]$  were known a priori. From  $P(t)$  one binding isotherm and one  $k_{on}$  value for every heat flow were determined

## 4.1 Application of PCCA+ to Kinetic Binding Data

using a covariant Gauss Newton method, specifically the NLSCON algorithm. NLSCON stands for Numerical solution of nonlinear (NL) least squares (S) problems with nonlinear constraints (CON), which was especially designed for numerically sensitive problems that work with an error oriented convergence criterion [65]. The titration time period was excluded, only the relaxation time period of each heat signal was considered. After having accounted the uncertainty quantification like shown by Vander Meulen et al. [49], the dilution in the sample cell and the instrument response time, one gets one  $k_{on}$  rate per titration step. According to equation 2.2  $k_{off}$  can be deduced for each titration.

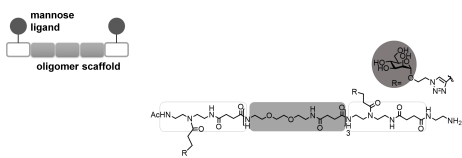
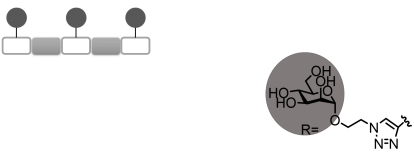
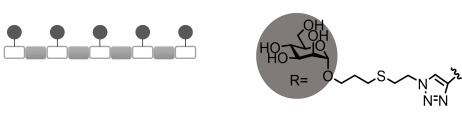
Variance (n)	Structure	Compound name
bivalent (2)		Man(1,5)-5
trivalent (3)		Man(1,3,5)-5
pentavalent (5)		Man(1,3,5,7,9)S-9

Table 4.1: Ligands used in the study [64; 66].

Figures taken from Erlekam et al. [63]. Copyright MDPI (Molecular Diversity Preservation International and Multidisciplinary Digital Publishing Institute). Reproduced with permission.

In the following, a PCCA+ clustering has been performed for the rate matrices, respectively. Because these rate matrices depend on the single binding and un-

## 4. KINITC+

---

binding rates, an inverse problem had to be solved. In order to fit the overall binding rate from the ITC experiments, we need to choose the unknown binding and unbinding rates,  $k_{on_i}$  and  $k_{off_i}$ , respectively. There are infinitely many combinations and hence as a goodness of fit measurement, the correlation coefficient has been used.

For the following three subsections, data sets of bi-, tri- and pentavalent ligands binding to tetrameric Con A (four binding sites) were selected from the series of measurements [64] to be compared with the model derived from the PCCA+ clustering. Table 4.1 shows the ligands that were used for comparing the experimental data to the mathematical model.

### 4.1.1 Bivalent Ligand Example

Even though the receptor is tetrameric Con A, this interaction is bivalent, because the ligand has only two binding sites. Assuming a one-to-one stoichiometry, only two bindings can be made maximum. The graph in Figure 4.1 shows a very

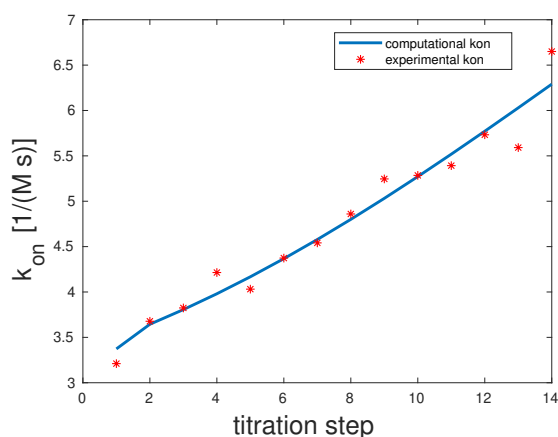


Figure 4.1: Comparison of the computed and experimental [64]  $k_{on}$  rates for bivalent Man(1,5)-5, see SI for additional information on experimental data. Figure taken from Erlekam et al. [63]. Copyright MDPI (Molecular Diversity Preservation International and Multidisciplinary Digital Publishing Institute). Reproduced with permission.

good fit with a correlation coefficient of 0.98. The theoretical  $k_{on}$  rates are 100 and 1,  $k_{off}$  are 1 and 1 in this particular setting. However, this solution is by



## 4.1 Application of PCCA+ to Kinetic Binding Data

far not unique. Considering the data, it is obvious that any upward sloping straight line can fit. Hence, there are many combinations of single  $k_{on}$  and  $k_{off}$  rates possible to achieve such a slope. Further, the sensitivity to find these unknown rates were only in logarithmic terms, i.e. they could have the values  $10^0, 10^1, 10^2, 10^3, 10^4$ . The absolute value of the overall  $k_{on}$  rates did not match the experimental value, only the behavior of the slope was considered. It was shifted for comparison purposes in the figure. Obviously, a stricter goodness of fit criterion is necessary. A second assessment of the absolute deviation of the clustered  $k_{on}$  rate and the ITC  $k_{on}$  showed that the model shown above is also the best fitting model. However, if the least squares deviation is considered, a different model showed the minimum deviation, such as shown in Figure 4.2. In

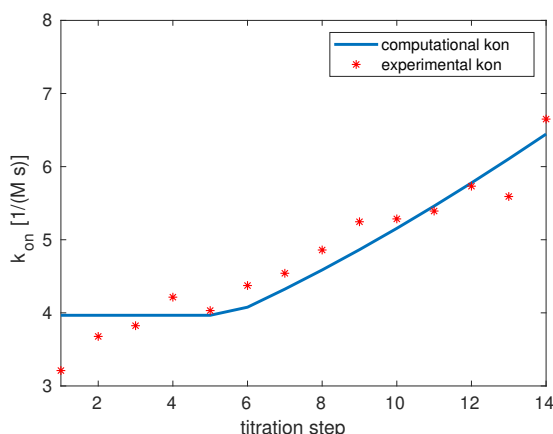


Figure 4.2: Comparison of the computed and experimental [64]  $k_{on}$  rates for bivalent Man(1,5)-5 with a least squares deviation measure for goodness of fit.

this particular scenario,  $k_{on_1} = 1$ ,  $k_{on_2} = 100$  and  $k_{off_2} = 10$ ,  $k_{off_1} = 1$ .

The correlation coefficient is still very high with 0.94. The interpretation of these values tell a slightly different story than the model above. Here, the first binding is very reluctant and the second one is more likely. Both unbinding processes are very unlikely.

The MSE of the bivalent ligand was used as an example to compare the result to the correlation coefficient fitting. We focus on the shape of the overall  $k_{on}$  curve rather on absolute or squared differences of the  $k_{on}$  values and therefore the following examples will be fitted by correlation coefficient only.

## 4. KINITC+

---

### 4.1.2 Trivalent Ligand Example

The valency in the next example is limited again by the ligand's number of binding sites. Even though the receptor Con A has four binding sites, there cannot be more than three bindings assuming a one-to-one stoichiometry. This particular

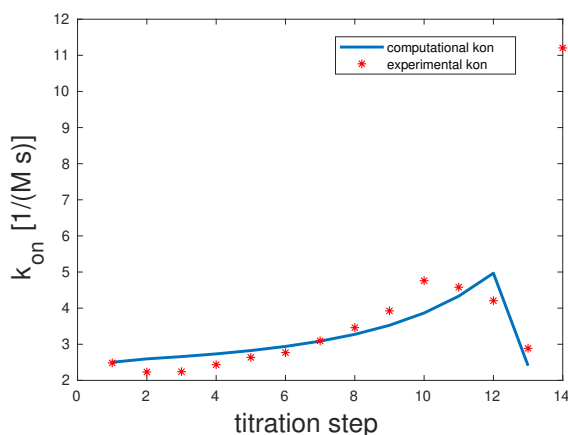


Figure 4.3: Comparison of the computed and experimental [64]  $k_{on}$  rates for bivalent Man(1,3,5)-5. Figure taken from Erlekan et al. [63]. Copyright MDPI (Molecular Diversity Preservation International and Multidisciplinary Digital Publishing Institute). Reproduced with permission.

model for the trivalent example reached a correlation coefficient of 0.87. The last data point is clearly an outlier as it does not seem to fit to the downward slope of the bell shaped curve. Therefore, the last titration has been omitted for fitting the overall  $k_{on}$  rate. In Figure 4.3 it has been plotted for the sake of completeness, though. The fitted  $k_{on}$  rates are 1000, 1000 and 1000,  $k_{off}$  are 1, 1 and 1.

Again, the proposed solution is not unique. For the inverse problem,  $6^5 = 15625$  different combinations of  $k_{on}$  and  $k_{off}$  rates were tested. Six because there were three different  $k_{on}$  rates and three different  $k_{off}$  rates. Each rate could take five different values:  $10^0, 10^1, 10^2, 10^3, 10^4$ . Out of these 15625, 25 combinations showed the maximum correlation coefficient of 0.87. An interesting finding is the fact that of these 25 optima is that all the three  $k_{on}$  rates are 1000 and  $k_{off_1} = 1$ . Both  $k_{off_2}$  and  $k_{off_3}$  range between 1 and 10000, thus there are 25 combinations and their influence seems irrelevant.

### 4.1.3 Pentavalent Ligand Example

The third example shows binding of a pentavalent mannose ligand to tetrameric Con A. Since  $n_{valency} = \min(n_{ligands}, n_{receptors}) = 4$ , there can be at most four bindings at a time. Even though the ligand is pentavalent, the binding is at most tetravalent. The  $k_{on}$  rates are 1, 1, 10000, and 10 and the  $k_{off}$  rates are 10000,

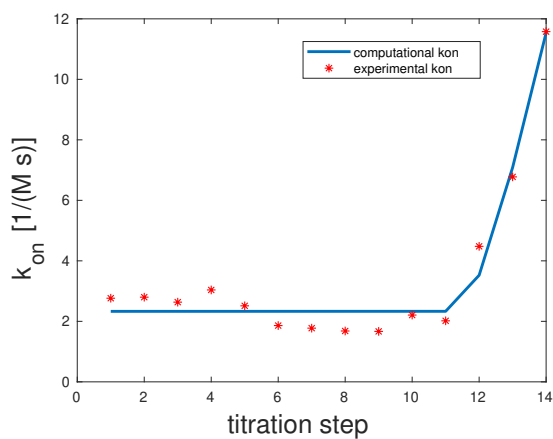


Figure 4.4: Comparison of the computed and experimental [64]  $k_{on}$  rates for pentavalent Man(1,3,5,7,9)S-9, see SI for additional information on experimental data. Figure taken from Erlekam et al. [63]. Copyright MDPI (Molecular Diversity Preservation International and Multidisciplinary Digital Publishing Institute). Reproduced with permission.

1, 1 and 1. The correlation coefficient for this example is 0.98. The scenario representing the best fit for this model shows that the first and second binding seem to be highly unlikely with  $k_{on_1} = k_{on_2} = 1$ , but once they took place, the third binding happens almost surely with  $k_{on_3} = 10000$ . The last binding is again unlikely with  $k_{on_4} = 10$ .  $k_{off_1} = 10000$  and  $k_{off_2} = k_{off_3} = k_{off_4} = 1$ . That means that overall unbinding is very low if there is more than one binding. If there is only one bond between the mannose and Con A, dissociation is very likely.

## 4.2 Parameter Interdependency

In this section it will be assessed which input values determine the clustered rate matrix most and how this affects the overall binding rate. First, it will

## 4. KINITC+

---

be analysed which shapes for the overall  $k_{on}$  curve are possible at all. Then an alternative enhancement factor is introduced to gain insights of the valency and the enhancement relationship. Finally, the general relationships between input and output rates are studied by a sensitivity analysis.

### 4.2.1 Shapes of $k_{on}$ Curves

In Section 4.1 the focus was on finding the optimal input parameters for the  $k_{on}$  and  $k_{off}$  rates to achieve the best fitting overall  $k_{on}$  curve. In the following, the same mannose-Con A binding examples are considered. In contrast to the previous subsections, in this subsection not only the best but all the possible curve shapes are assessed. For a better visibility, their slopes are only sketched schematically. For comparison reasons, all curves are scaled such that overall  $k_{on} \in [0, 1]$ .

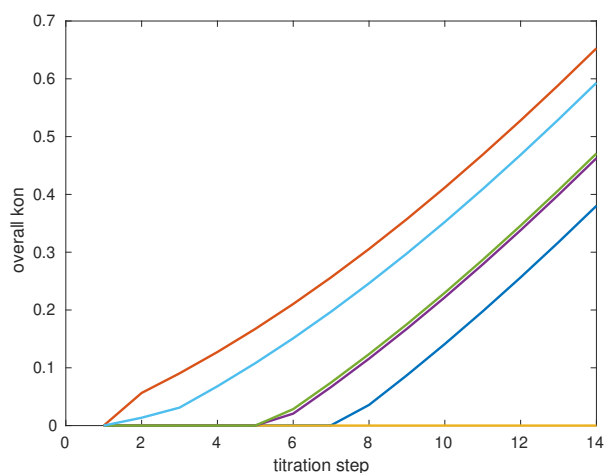


Figure 4.5: Possible overall  $k_{on}$  curves shapes for Man15-5 for  $k_{on}$  and  $k_{off}$  ranging between 1,10,100,1000 and 10000. The colors of the curves are only a visual distinction and have no mathematical meaning.

In the bivalent case, there is basically only an upward trend possible as depicted in Figure 4.5. The main difference is at which titration step the kink in the curve comes up. It is also possible that the upward slope is so small that the curve appears to be a constant. Thus, if the experimental  $k_{on}$  curve is upward sloping, it is very easy to fit. Possibly, the higher the valency, the more diverse the shapes

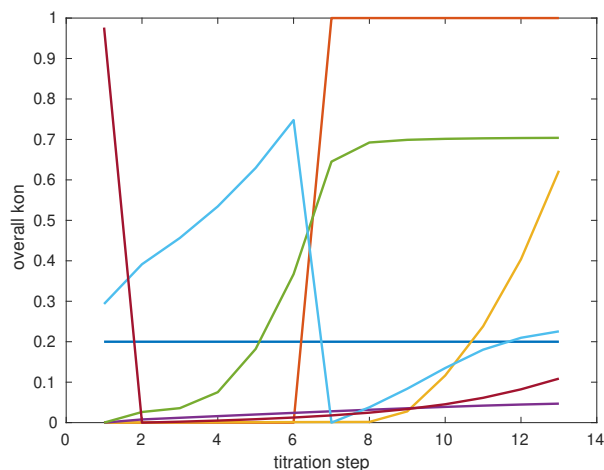


Figure 4.6: Possible overall  $k_{on}$  curves shapes for Man135-5 for  $k_{on}$  and  $k_{off}$  ranging between 1,10,100,1000 and 10000. The colors of the curves are only a visual distinction and have no mathematical meaning.

can be. This assumption is true at least for the trivalent ligand example in Figure 4.6.

As expected, the variety of possible  $k_{on}$  curve shapes is much wider in the trivalent example than in the bivalent one. Most curves show an upward trend, too. Within the upward sloping curves, there is also a variety. The curves can be convex, concave or have a sigmoidal shape. However, a constant is also possible as well as abruptly changing zigzag curves. The latter may indicate a sudden change in the binding process, i.e. that one process gets significantly faster than the other one.

As depicted in Figure 4.7, in the pentavalent example the range of possible shapes is even more diverse than in the previous two examples. It can be constant, a smooth upward slope or bumpy up and down curves. A sigmoidal shape is also achievable. This supports the expectation that the more input variables are involved, the more complex the process and the more diverse the results can be. Again, abrupt changes such as inflection points indicate that one micro process takes control over the previous processes. Since these binding and unbinding events happen on a micro scale, they may have an effect on the macro scale.

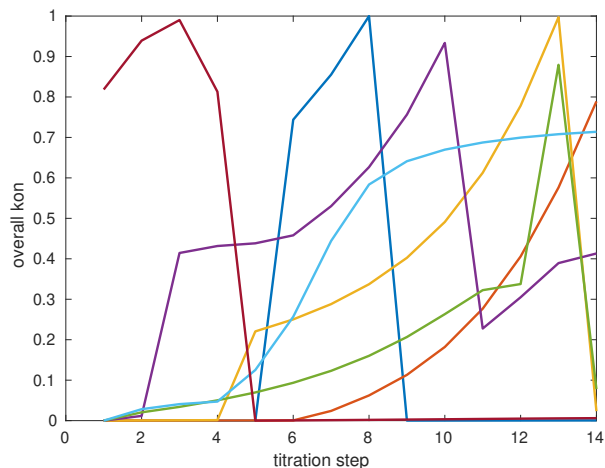


Figure 4.7: Possible overall  $k_{on}$  curves shapes for Man13579-9S for  $k_{on}$  and  $k_{off}$  ranging between 1,10,100,1000 and 10000. The colors of the curves are only a visual distinction and have no mathematical meaning.

### 4.2.2 General Relationship of the Single and Overall Binding Rates

This section analyses how changing the microscopic  $k_{on}$  and  $k_{off}$  rates influence the macroscopic  $k_{on}$  rate. We start the assessment with the bivalent rate matrix. Recall the shape of the bivalent rate matrix (3.7) from section 3.2.2. It is a sparse matrix with non-zero sub-matrices across the diagonal. The diagonal is the negative column sum. The following subsections discuss the impact of altering the input parameters of  $Q$  and how the macroscopic binding and unbinding rates of the clustered rate matrix evolve. For simplicity, the ligand concentration is assumed to be constant and equal to 1. Thus,  $Q$  depends only on the microscopic  $k_{on}$  and  $k_{off}$  rates as input parameters.

## 4.2 Parameter Interdependency

$$Q = \begin{pmatrix} d_1 & k_{off_1} & k_{off_1} & k_{off_1} & k_{off_1} & 0 & 0 \\ k_{on_1} & d_2 & 0 & 0 & 0 & k_{off_2} & 0 \\ k_{on_1} & 0 & d_2 & 0 & 0 & 0 & k_{off_2} \\ k_{on_1} & 0 & 0 & d_2 & 0 & k_{off_2} & 0 \\ k_{on_1} & 0 & 0 & 0 & d_2 & 0 & k_{off_2} \\ 0 & k_{on_2} & 0 & k_{on_2} & 0 & d_3 & 0 \\ 0 & 0 & k_{on_2} & 0 & k_{on_2} & 0 & d_3 \end{pmatrix} \quad (4.1)$$

with  $d_1 = -4k_{on_1}$ ,  $d_2 = -k_{off_1} - k_{on_2}$  and  $d_3 = -2k_{off_2}$ .

In this case study,  $k_{on}$  and  $k_{off}$  could take values of  $10^i$ ,  $i \in \{0, 1, 2, 3, 4\}$ . Thus, there are  $5^4 = 625$  different combinations. The overall  $k_{on}$  rate ranges from 1.41 to  $1.9999e + 04$  as shown in Figure 4.8.

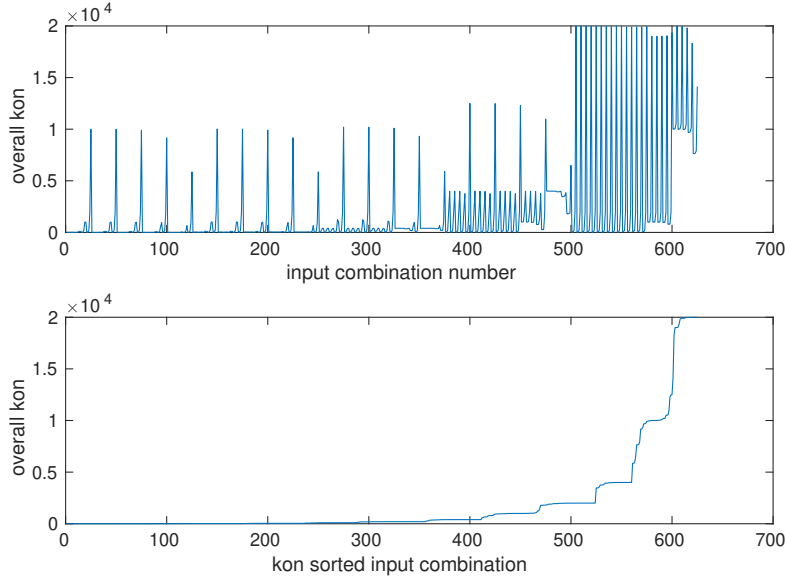


Figure 4.8: Evolution of the overall  $k_{on}$  rate depending on the input combination of  $k_{on_1}, k_{on_2}, k_{off_1}, k_{off_2}$ . The upper figure shows the overall  $k_{on}$  rate in the order of the input combination that was tested. In the lower figure the overall  $k_{on}$  rates are sorted from min to max in order to get insights into which input combinations cause it to decrease or increase, respectively.

We are interested in the evolution of the overall binding rates, therefore they are

#### 4. KINITC+

---

also depicted sorted. The interesting parts are the extremes:  $\arg \min k_{on}$  and  $\arg \max k_{on}$ . What input values minimize and maximize  $k_{on}$ ? The 35 smallest  $k_{on}$  values are smaller than 2. Their input variables are listed in Table 4.2.

Table 4.2: Bivalent parameter study. These 35 different input combinations cause the overall binding rate to be minimal.

$k_{on_1}$	$k_{on_2}$	$k_{off_1}$	$k_{off_2}$	overall $k_{on}$
1	1	1	1	1.4142
1	1000	10000	1	1.8182
1	100	1000	1	1.8185
1	10	100	1	1.8211
1	1	10	1	1.8443
1	10	10	1	1.9002
10	1	1	1	1.9488
10	1	10	1	1.9585
1	100	10000	1	1.9802
1	10	1000	1	1.9802
10	100	10000	1	1.9803
1	1	100	1	1.9806
10	10	1000	1	1.9809
10	1	100	1	1.9856
1	100	100	1	1.9900
100	1	1	1	1.9950
100	1	10	1	1.9951
100	1	100	1	1.9960
1	10	10000	1	1.9980
1	1	1000	1	1.9980
10	10	10000	1	1.9980
10	1	1000	1	1.9981
100	10	10000	1	1.9981
100	1	1000	1	1.9986

Continued on next page



## 4.2 Parameter Interdependency

**Table 4.2 – continued from previous page**

$k_{on_1}$	$k_{on_2}$	$k_{off_1}$	$k_{off_2}$	overall $k_{on}$
1	1000	1000	1	1.9990
1000	1	1	1	1.9995
1000	1	10	1	1.9995
1000	1	100	1	1.9995
1000	1	1000	1	1.9996
1	1	10000	1	1.9998
10	1	10000	1	1.9998
100	1	10000	1	1.9998
1000	1	10000	1	1.9999
1	10000	10000	1	1.9999
10000	1	1	1	1.9999

What is striking is that for all these values,  $k_{off_2} = 1$ . It seems like  $k_{off_2}$  is a dominant input variable. The maximum output values for the overall binding rate is  $1.99e+4$ , which occurs 17 times. Their input factors are shown in Table 4.3 .

Table 4.3: Bivalent parameter study. These 17 different input combinations cause the overall binding rate to be maximal.

$k_{on_1}$	$k_{on_2}$	$k_{off_1}$	$k_{off_2}$	overall $k_{on}$
10000	100	1	10000	1.9900e10+4
10000	100	10	10000	1.9900e10+4
10000	100	100	10000	1.9900e10+4
10000	100	1000	10000	1.9905e10+4
10000	100	10000	10000	1.9933e10+4
10000	10000	10	10000	1.9980e10+4
10000	10	1	10000	1.9990e10+4
10000	10	10	10000	1.9990e10+4
Continued on next page				

#### 4. KINITC+

---

**Table 4.3 – continued from previous page**

$k_{on_1}$	$k_{on_2}$	$k_{off_1}$	$k_{off_2}$	overall $k_{on}$
10000	10	100	10000	1.9990e10+4
10000	10	1000	10000	1.9990e10+4
10000	10	10000	10000	1.9993e10+4
10000	10000	1	10000	1.9998e10+4
10000	1	1	10000	1.9999e10+4
10000	1	10	10000	1.9999e10+4
10000	1	100	10000	1.9999e10+4
10000	1	1000	10000	1.9999e10+4
10000	1	10000	10000	1.9999e10+4

In the table above it is remarkable, that in all cases where the overall  $k_{on}$  rate takes the maximum, both  $k_{on_1}$  and  $k_{off_2}$  are maximal, i.e. 10000. Both  $k_{off_1}$  and  $k_{on_2}$  vary,  $k_{off_1}$  even more than  $k_{on_2}$ .

In order to assess the individual contribution of each input factor, a sensitivity study is presented. Again, the ligand concentration remains fixed at 1 for all time steps. All input variables are also held constant while only one is being changed. First, the influence of  $k_{on_1}$  is depicted in Figure 4.9. When varying only  $k_{on_1}$ , the overall  $k_{on}$  curve has a sigmoidal shape with its inflection point around  $\log(k_{on_1}) = 0$ . The overall  $k_{off}$  curve first rises linearly with a peak around  $\log(k_{on_1}) = 0$  and decreasing somewhat afterwards before reaching an equilibrium.

Secondly, we assess the influence of  $k_{on_2}$  in Figure 4.10. Varying  $k_{on_2}$  distorts the sigmoidal shape of the overall  $k_{on}$  rate with a downward kink around  $\log(k_{on_2}) = 0$ . The overall  $k_{off}$  rate peaks exactly there. Both curves reach their equilibria when  $\log(k_{on_2}) > 5$ .

Now the unbinding coefficients are focused on. Thirdly, the influence of  $k_{off_1}$  is shown in Figure 4.11. The variation of  $k_{off_1}$  leads to an overall  $k_{on}$  rate with a minimum at  $\log(k_{off}) = 0$ . The overall  $k_{off}$  rate is maximal at this point and plummets afterwards. Like in the sensitivity study of  $k_{on_2}$ , both overall  $k_{off}$  curve and the  $K_d$  curve reach their equilibria when  $\log(k_{off_1}) > 5$ .

## 4.2 Parameter Interdependency

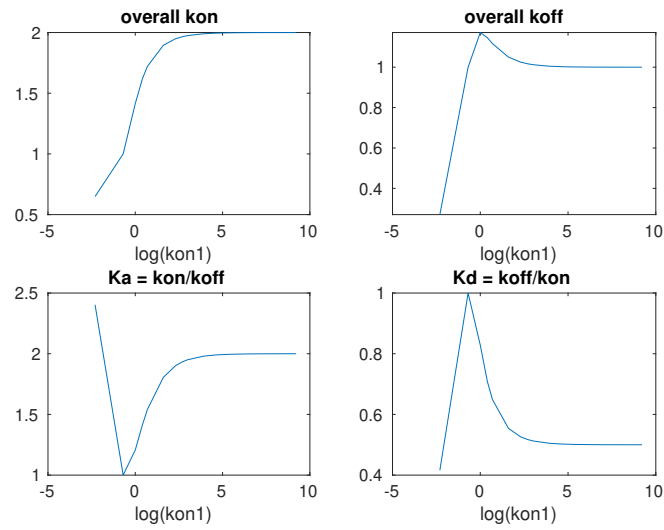


Figure 4.9: Possible output when all input parameters are held constant except  $k_{on1}$ .

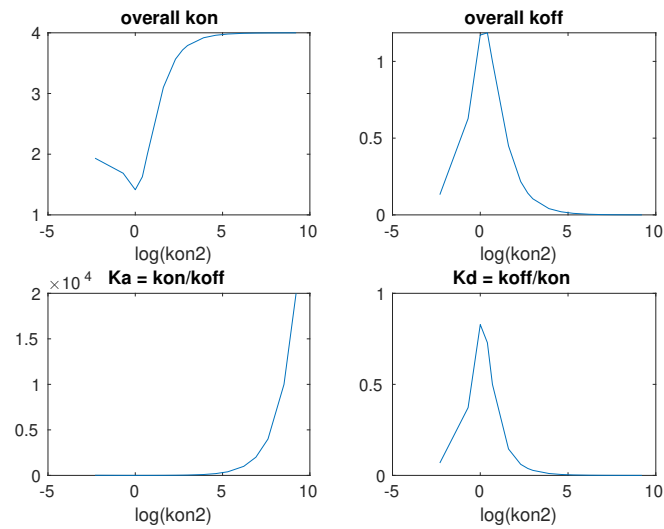


Figure 4.10: Possible output when all input parameters are held constant except  $k_{on2}$ .

#### 4. KINITC+

---

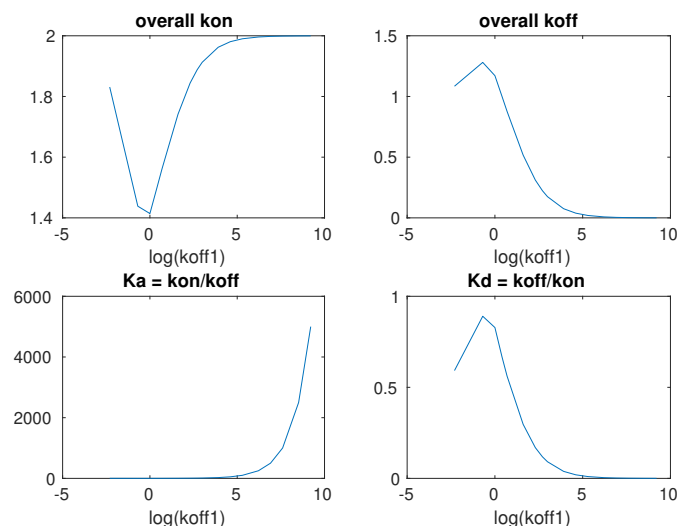


Figure 4.11: Possible output when all input parameters are held constant except  $k_{off_1}$ .

Lastly, Figure 4.12 shows the impact of varying  $k_{off_2}$ . The overall  $k_{on}$  curves of Figures 4.9 and 4.12 have their sigmoidal shape in common. Both Figures 4.10 and 4.11 show a global minimum at the input value  $\log(1) = 0$  and both curves converge to their maxima at  $k_{on} = 4$  or  $2$ , respectively. Concerning the overall  $k_{off}$  rates, in all four cases the curves have a peak around the input value of  $\log(1) = 0$ . After the peak the overall  $k_{on}$  rate converges to 1 for varying  $k_{on_1}$  and  $k_{off_2}$ , and to 0 for varying  $k_{on_2}$  and  $k_{off_1}$ , respectively.

Examining Tables 4.2 and 4.3 one would expect that varying  $k_{on_1}$  and  $k_{off_2}$  would yield the most impact. That means that the overall  $k_{on}$  and  $k_{off}$  rates would show the biggest spectra. By varying  $k_{on_2}$  and  $k_{off_1}$  one might expect almost stable overall binding and unbinding rates with a limited spectrum respectively. However, this is obviously not the case. Determining the reason for these cause and effect relationships is out of the scope of this thesis, but it has been shown that the mechanisms of the moments when a certain micro process takes control over the macro process is much more complex than previously anticipated.

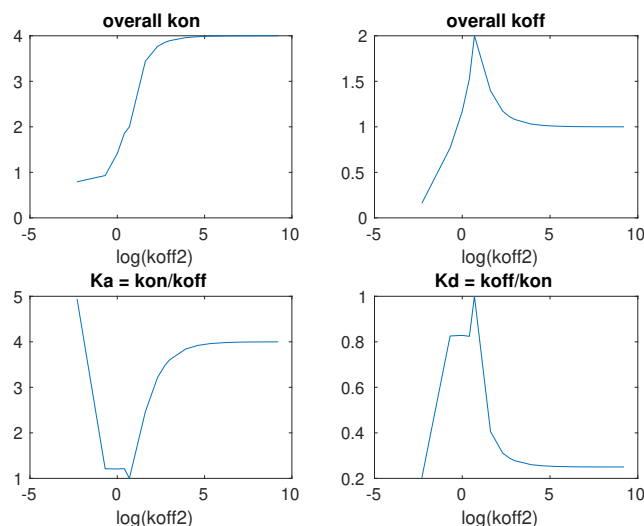


Figure 4.12: Possible output when all input parameters are held constant except  $k_{\text{off}2}$ .

### 4.2.3 Alternative Enhancement Factor

Recall the enhancement factor  $\beta$  describing the binding enhancement comparing monovalent bindings to multivalent ones from Section 2.3. In this subsection we will discuss an alternative metric for that enhancement factor. Therefore we will compare the binding coefficients of the first bond to the overall binding coefficients. A possible alternative enhancement factor could be  $\beta = \frac{k_{\text{on}1} k_{\text{off}}}{k_{\text{off}1} k_{\text{on}}}$ . The second fraction uses the overall binding and unbinding rates determined from the PCCA+ clustering. The alternative  $\beta$  was chosen this way such that the units will cancel out. This would not be the case if for example  $\frac{k_{\text{on}2}}{k_{\text{off}2}}$  was used instead of  $\frac{k_{\text{on}1}}{k_{\text{off}1}}$ . In the following, the alternative enhancement factor will be calculated and plotted over time for each of the three titration experiments discussed above.

Table 4.4 assesses the alternative enhancement factor from the bivalent mannose-Con A fitting from Section 4.1.1. The macroscopic fraction of  $k_{\text{on}}/k_{\text{off}}$  is 100 and the microscopic fraction is calculated per titration step.

#### 4. KINITC+

---

ligand valency	$\frac{k_{on1}}{k_{off1}}$	$\frac{k_{off}}{k_{on}}$	$\beta$
bivalent	$\frac{100}{1}$	0.0407	4.0663
		0.0354	3.5364
		0.0328	3.2841
		0.0305	3.0505
		0.0284	2.8355
		0.0264	2.6388
		0.0246	2.4596
		0.0230	2.2972
		0.0215	2.1503
		0.0202	2.0178
		0.0190	1.98986
		0.0179	1.7912
		0.0169	1.6946
0.0161	1.6076		

Table 4.4: Alternative enhancement factor  $\beta$  for bivalent mannose Man(1,5)-5

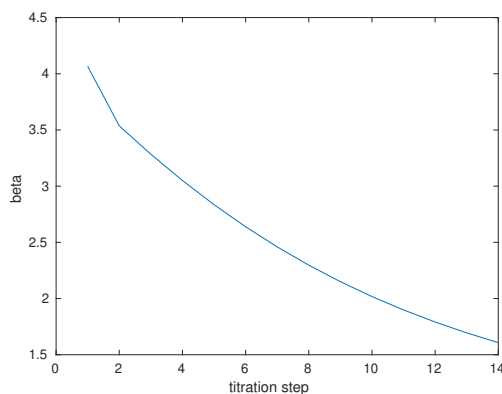


Figure 4.13: Alternative enhancement factor  $\beta$  for bivalent mannose Man(1,5)-5

The  $\beta$  -column from Table 4.4 is plotted in Figure 4.13.  $\beta$  has an overall downward trend with a kink at the second titration step. A possible explanation for the kink is that first binding in the first two titrations happen very fast and then some other molecular micro-process such as second arm binding or unbinding

## 4.2 Parameter Interdependency

---

takes over control and happens faster.

The alternative enhancement factor for the trivalent example is also downward sloping like depicted in Figure 4.14 and Table 4.5. Numerically, it is higher because the first fraction of  $\beta$  is ten times higher, i.e.  $k_{on_1}/k_{off_1} = 1000$ . Interestingly, there is also a kink at the second injection, but the slope is much steeper than in the bivalent example.

ligand valency	$\frac{k_{on_1}}{k_{off_1}}$	$\frac{k_{off}}{k_{on}}$	$\beta$
trivalent	$\frac{1000}{1}$	0.0629	62.9083
		0.0413	41.2577
		0.0325	32.5307
		0.0253	25.2900
		0.0193	19.3237
		0.0145	14.4566
		0.0105	10.5413
		0.0075	7.4523
		0.0051	5.0797
		0.0033	3.3235
		0.0021	2.0865
		0.0013	1.2676
0.0000	0.000		

Table 4.5: Alternative enhancement factor  $\beta$  for trivalent mannose Man(1,3,5)-5

Similar to Figure 4.4, the graph in Figure 4.15 is constant until including the 11<sup>th</sup> titration step. Afterwards the graph has a downward slope like the other enhancement factors in Figures 4.13 and 4.14. Due to the big difference between  $k_{on_1}$  and  $k_{off_1}$ ,  $K_d$  is multiplied by almost 0 and therefore  $\beta$  is almost 0 too. The enhancement effect is negligible in this example.

#### 4. KINITC+

---

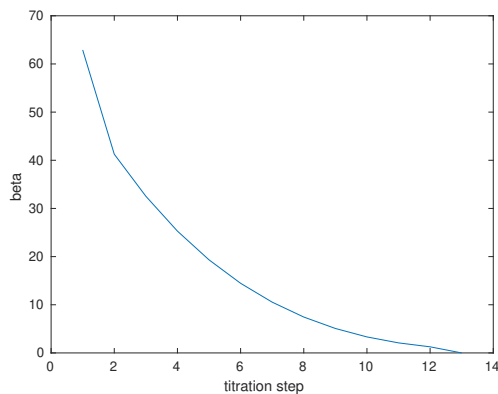


Figure 4.14: Alternative enhancement factor  $\beta$  for trivalent mannose Man(1,3,5)-5

ligand valency	$\frac{k_{on1}}{k_{off1}}$	$\frac{k_{off}}{k_{on}}$	$\beta$
pentavalent	$\frac{1}{10000}$	0.0179	0.0179
		0.0179	0.0179
		0.0179	0.0179
		0.0179	0.0179
		0.0179	0.0179
		0.0179	0.0179
		0.0179	0.0179
		0.0179	0.0179
		0.0179	0.0179
		0.0179	0.0179
		0.0179	0.0179
		0.0147	0.0147
		0.0090	0.0090
0.0055	0.0055		

Table 4.6: Alternative enhancement factor  $\beta$  for pentavalent mannose Man(1,3,5,7,9)-9

In Figure 4.4, the overall  $k_{on}$  curves show also an upward trend on the first glance. However, because there are more possible input parameter permutations, there are also more curves clouding that not all curves are strictly upward sloping.



### 4.3 Limitations of the Model

---

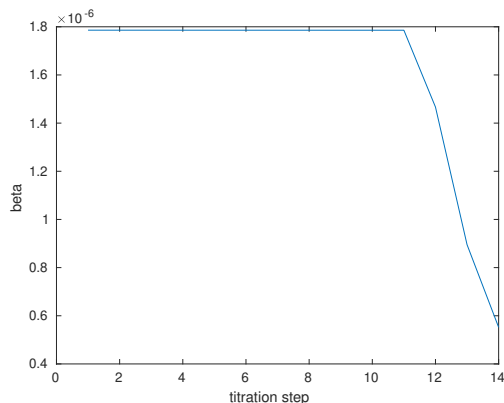


Figure 4.15: Alternative enhancement factor  $\beta$  for pentavalent mannose Man(1,3,5,7,9)-9

Many of them show kinks, or only local maxima. Again, the most likely reason is that some microscopic process such as a single binding step is much faster than another one, such as a sudden unbinding step. The first binding step seems to be slow and the complex' stability stems from the subsequent bindings.

### 4.3 Limitations of the Model

As already lined out in the numerical bivalent example, there are many ways to find suitable input rates to gain the same slope behavior of the  $k_{on}$  curve. As the the curve is a simple upward slope without any kinks or inflections, it is easy to find some combination using four input variables, namely  $k_{on1}$ ,  $k_{on2}$ ,  $k_{off1}$  and  $k_{off2}$ . Thus, there exists a solution, but it is not unique. The number of macroscopic binding and unbinding rates is linearly dependent on the valency. Thus, the higher valency, the more complicated the shape of the overall binding curve and the more justifiable our solution.

In terms of validation, the choice of measuring the goodness of fit is ambiguous. It determines which models of the assessed sample mimic the overall  $k_{on}$  rate best. However, different methods to measure the fit do not necessarily come up with the same output. Compare for instance the correlation coefficient used above with the mean squared error. A model with exactly the same slope but with

#### 4. KINITC+

---

shifted values in y-direction would yield a much lower fit than a graph that is numerically close to the original curve, but not having the same slope. One way of tackling this problem is to apply more than one goodness of fit measure, but then the question of the order of measure applications remains.

# Chapter 5

## Wiseman Fitting

Chapters 2 and 3 explained how to obtain thermodynamic and kinetic information from binding processes using ITC and kinITC. The following chapter is investigating the derivation of kinetic parameters from a different angle: the Wiseman fitting. Like ITC, the Wiseman fitting extracts binding information such as stoichiometry  $n$ , the heat released upon binding, as well as the overall  $k_{off}$  rate. Also similar to ITC, the classical Wiseman fitting does not take intermediate binding steps into account, but regards even a multivalent binding process as one direct binding. This chapter will apply the PCCA+ algorithm to the classical Wiseman fitting to account for intermediate binding steps and kinetic parameters like obtained from kinITC+. The Wiseman fitting will be applied to two multivalent binding experiments taken from [64] and compared to an amended version of kinITC+ presented in Chapter 3.

The theory, notation and experimental results presented in this chapter are adapted from the paper of Erlekam et al. [67] if not specified differently. The code used for the numerical examples belongs to the same paper and is referenced therein.

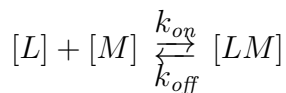
### 5.1 The classical Wiseman Fitting

In 1989 Wiseman et al. were one of the pioneers to describe the method of how to extract binding parameters of bivalent and multivalent bindings with titration

## 5. WISEMAN FITTING

---

calorimeters [68]. Like in the previous chapter's notation, we again model a ligand - macromolecule binding by



with  $[\cdot]$  denoting the respective concentrations.

Given an experimental ITC output data set with  $T \in \mathbb{N}$  injections, the total concentrations of the ligand  $L$  and the macromolecule  $M$  are:  $L_t = \{l_1, \dots, l_T\}$  and  $M_t = \{m_1, \dots, m_T\}$ . Then the *Wiseman function* is defined as

$$W(K_a, n; l, m) := \frac{1}{2} \left( 1 + \frac{n - \frac{l}{m} - \frac{1}{mK_a}}{\sqrt{\left(n + \frac{l}{m} + \frac{1}{mK_a}\right)^2 - 4n\frac{l}{m}}} \right),$$

for  $l \in L_t$  and  $m \in M_t$ .

The peaks of the thermogram are integrated to determine the transition heat  $q_{trans}$

$$q_{trans} := \{(q_{trans})_1, \dots, (q_{trans})_T\}.$$

Practically, the integration can be achieved by approximating the upward slope as a spline curve of a higher degree, i.e. 3 to 5. Finally, the Wiseman fitting is performed, which is a search algorithm that chooses the numerical values for the stoichiometry  $n$ , the enthalpy  $\Delta H$  and the association constant  $K_a$  that minimizes the difference between  $q_{trans}$  and the Wiseman function with a Frobenius norm:

$$\min_{K_a, n, \Delta H} \|q_{trans} - W(K_a, n; L_t, M_t)\Delta H\| \quad (5.1)$$

In this formula the term  $W$  has no physical unit. As a consequence,  $q_{trans}$  has the same unit as  $\Delta H$ . Scaling  $q_{trans}$  just scales  $\Delta H$ . To make the units fit,  $K_a$  needs to be without a physical unit too, thus it has to be the inverse of the physical unit of the macromolecule concentration  $m$ . Rescaling both  $m$  and  $l$  would only change the physical unit of  $K_a$ .

## 5.2 The $Q_c$ Fitting

The  $Q_c$  fitting is done the same way as described in Section 3.3. First, a transition rate matrix is set up as in 3.8. Then the clustering is denoted as

$$Q_c^T = (\chi^T \Pi \chi)^{-1} \chi^T \Pi Q^T \chi,$$

Again in the end the ODE

$$Q_c = \begin{pmatrix} -k_{on} \cdot c_L & k_{off} \\ k_{on} \cdot c_L & -k_{off} \end{pmatrix}$$

must be solved. For this clustered rate matrix the ligand concentration  $[L]$  is needed per titration step. It is computed iteratively as

$$[L]_i = \begin{cases} l_1, & \text{if } i=1, \\ [L]_{i-1} + l_i - (L_b)_{i-1}, & \text{otherwise,} \end{cases}$$

where

$$(L_b)_i := \frac{nm_i + l_i + \frac{1}{K_a} - \sqrt{\left(nm_i + l_i + \frac{1}{K_a}\right)^2 - 4nm_i l_i}}{2}$$

is the *bound ligand concentration for injection  $i$* . A proof for the formula of ligand concentration can be found in Zumbansen [40]. For an  $s$ -valent binding, the binding parameter for each titration step is

$$Q(k_{on_1}, k_{off_1}, \dots, k_{on_s}, k_{off_s}; [L]) = K_1, \dots, K_T,$$

with the set of free ligand concentrations  $[L]$  and, on the right side, the resulting association constants  $K_i$  for each injection  $i \in \{1, \dots, T\}$ . Then the  $Q_c$  fitting is defined with the Frobenius norm as

$$\min_{\substack{k_{on_1}, \dots, k_{on_s} \\ k_{off_1}, \dots, k_{off_s} \\ n, \Delta H}} \left\| q_{trans} - W(Q(\dots; [L]), n; L_t, M_t) \Delta H \right\|. \quad (5.2)$$

## 5. WISEMAN FITTING

---

We are thus looking for binding parameters such that the Euclidian distance between the  $q_{trans}$  vector and the Wiseman function is minimal.

The main difference between the kinITC+ method and this  $Q_c$  fitting is that kinITC+ only searches for optimal binding and unbinding parameters on a logarithmic scale of base 10<sup>0</sup> up to 10<sup>4</sup> and finds the best fitting overall  $k_{on}$  curve, whereas the  $Q_c$  fitting in this chapter starts the search randomly and the  $k_{on}$  and  $k_{off}$  rates can take up any float number. The kinITC+ code was written in Matlab, the  $Q_c$  code in Python 3.7.

### 5.3 Extracting experimental ITC Data

For the Wiseman fitting the total concentrations of ligands  $L_t$  and macromolecule  $M_t$  have to be extracted from the ITC data set. Egawa et al.[19] describe in their Appendix I how to obtain both:

$$l_i = c_L \left( \sum_{k=1}^i V_k \right) \frac{2V_0 - \sum_{k=1}^i V_k}{2(V_0)^2}$$
$$m_i = c_M \frac{2V_0 - \sum_{k=1}^i V_k}{2V_0 + \sum_{k=1}^i V_k},$$

with  $V_0$  the main cell volume containing the macromolecules,  $V_i$ ,  $i \in \{1, \dots, T\}$  the titrated volumes containing the ligands,  $c_L$  the ligand concentration and  $c_M$  the macromolecule concentration. The total volume is  $V_{Tot} = V_0 + \sum_{i=1}^T V_i$ .

### 5.4 Application of $Q_c$ and extended Wiseman Fitting

In the following, a bivalent binding and a trivalent binding experiment are fit. The data is again taken from Igde et al. [64]. The bivalent example is the binding between bivalent Concanavalin A (Con A) with decavalent oligo(amidoamines) with pendant mannose side chains. Assuming a one to one stoichiometry and since  $\min(10, 2) = 2$ , the binding can be maximum bivalent. The trivalent example is

## 5.4 Application of $Q_c$ and extended Wiseman Fitting

tetravalent Con A binding with trivalent mannose macromolecules. Here, the same stoichiometry is assumed and due to  $\min(4, 3) = 3$  there can be a trivalent binding at most. Again, only the first 14 titration steps are taken into account. For any computation, we need to extract the concentration and volume data as described in the previous section. For the bivalent case, the values are:

$$\begin{aligned}c_M &= 0.071 \text{ mmol/l} & V_0 &= 1.442 \text{ ml} \\c_L &= 0.7 \text{ mmol/l} & V_1 &= 0.001 \text{ ml}, V_j = 0.01 \text{ ml}\end{aligned}$$

and for the trivalent case:

$$\begin{aligned}c_M &= 0.0918 \text{ mmol/l} & V_0 &= 1.442 \text{ ml} \\c_L &= 0.978 \text{ mmol/l} & V_1 &= 0.001 \text{ ml}, V_j = 0.01 \text{ ml}\end{aligned}$$

with  $j \in \{2, \dots, 14\}$ .

### 5.4.1 Bivalent Binding Example

Figure 5.1 shows the first 14 heat curve peaks of the bivalent ITC experiment.

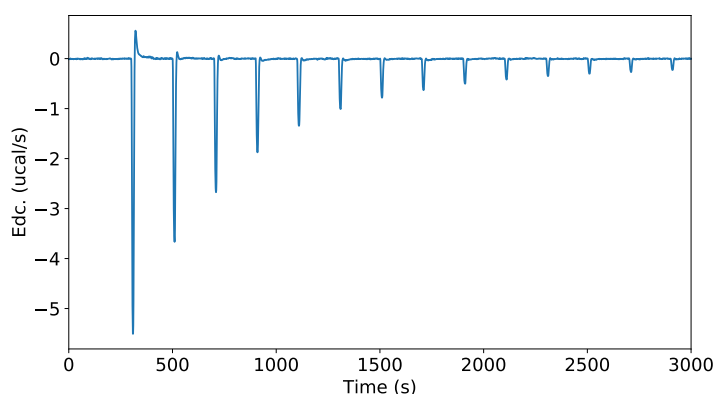


Figure 5.1: Deconvolved power trace of an ITC experiment with bivalent Con A and decavalent mannose

The results of the *Wiseman fitting* for the bivalent example are:

## 5. WISEMAN FITTING

$K_a = 93.3295$	$n = 0.1998$
$\Delta H = -132.3069$ kcal/mol	<b>error = 5.4838</b>

The  $Q_c$  **fitting** not only delivers  $n$ ,  $\Delta H$  and  $K_a$ , but also the intermediate  $k_{on}$  and  $k_{off}$  rates:

$k_{on1} = 136.4185$	$k_{off1} = 2.3716$
$k_{on2} = 33.4088$	$k_{off2} = 18.5925$
$n = 0.3208$	$\Delta H = -73.4629$ kcal/mol
	<b>error = 3.435</b>

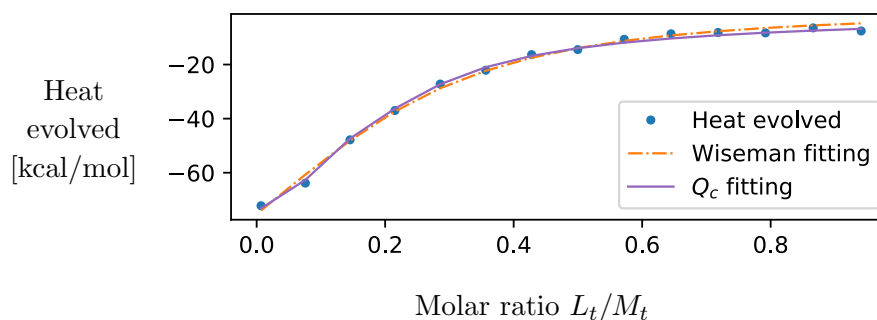


Figure 5.2: Wiseman plot of the bivalent ITC experiment with Wiseman fitting and  $Q_c$  fitting

Both fittings give different results and have their strengths and weaknesses. The heat of binding from the Wiseman fitting is 80% higher than  $\Delta H$  from the  $Q_c$  fitting. The number of binding sites from the Wiseman fitting is only 62% of the  $n$  from the  $Q_c$  fitting. From the sigmoidal curve of the Wiseman graph one can see that the inflection point is around 0.2. Therefore, the Wiseman fitting shows an  $n$  value closer to the true number of binding sites.

The Wiseman fitting gives only an overall  $K_a$  rate, thus we cannot directly compare the microscopic association constants the  $Q_c$  fitting delivers. These are:

$$K_1 = k_{on1}/k_{off1} = 57.5217$$

$$K_2 = k_{on2}/k_{off2} = 1.7969.$$

These values suggest a stronger first binding and a much weaker second binding. Note that due to the lack of physical units, these results are of qualitative nature.



## 5.4 Application of $Q_c$ and extended Wiseman Fitting

We can only learn about the binding affinity of the first binding relative to the subsequent ones.

The  $Q_c$  fitting error is smaller than the Wiseman error. Thus, in terms of absolute error, the  $Q_c$  produces a better fit to the actual ITC data than the Wiseman fitting does.

### 5.4.2 Trivalent Binding Example

Figure 5.3 shows the first 14 heat curve peaks of the trivalent ITC experiment. For the fitting procedures, the first titration has been omitted, because it is an outlier.

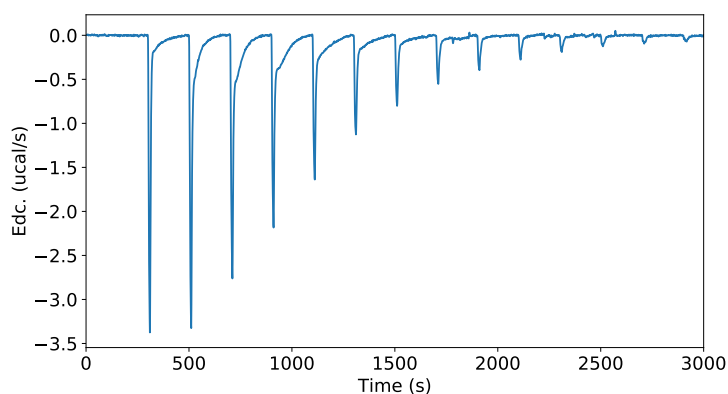


Figure 5.3: Deconvolved power trace of an ITC experiment with trivalent Con A and trivalent mannose

The results of the Wiseman fitting for the trivalent example are:

$K_a = 410.2917$	$n = 0.3586$
$\Delta H = -127.7664$ kcal/mol	<b>error = 13.1455</b>

The  $Q_c$  fitting for the trivalent example produced the following results:

## 5. WISEMAN FITTING

$k_{on1} = 559.1484$	$k_{off1} = 1844.5886$
$k_{on2} = 489.7665$	$k_{off2} = 81.966$
$k_{on3} = 1056.7834$	$k_{off3} = 1338.633$
$n = 0.3624$	$\Delta H = -115.1639$ kcal/mol
<b>error = 11.1979</b>	

Figure 5.4 shows both the Wiseman and the  $Q_c$  fitting.

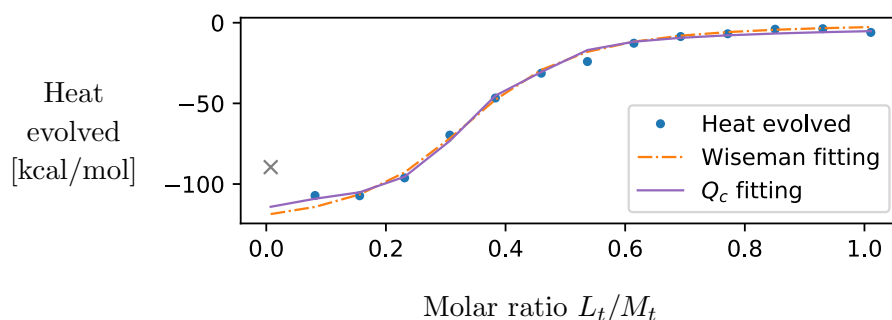


Figure 5.4: Wiseman plot of the trivalent ITC experiment with Wiseman fitting and  $Q_c$  fitting

Again, the  $Q_c$  has a lower error in absolute terms than the Wiseman fitting. Both findings determine a number of binding sites of 3.6. The microscopic binding rates as a result of the  $Q_c$  fitting are:

$$K_1 = k_{on1}/k_{off1} = 0.3031$$

$$K_2 = k_{on2}/k_{off2} = 5.9752$$

$$K_3 = k_{on3}/k_{off3} = 0.7894.$$

Once again, we can only discuss the qualitative aspect of these figures. The first binding is a very hesitant one. Once the first bond is made, the second one is much more likely. A third one is again rather unlikely.

### 5.5 Limitations of the Model

There is only a limited choice for setting the parameters of both models. The Wiseman fitting has become an established tool within the community, so the ob-

## 5.5 Limitations of the Model

---

tained association and dissociation constants can be trusted. The  $Q_c$  fitting depends on the chemical substances: their stoichiometry and the therefore possible binding events and their concentrations. The parameter that could be discussed critically is the dimensionality of the subspace that the rate matrix is projected onto. So far it was assumed, that there is simply bound and unbound and therefore the dimension of the rate matrix was reduced to  $2 \times 2$  by clustering. One may argue if this is the right method for all kinds of valencies.

There are two timescales at play: the macro perspective, i.e. the overall binding event, runs on a slow timescale. The micro perspective, all the single binding and unbinding processes of all binding sites happen on the fast timescale. How can these timescales be separated? Swope et al. show in their article how the fastest and the second fastest process can be separated using eigenvalues [69]. The smallest eigenvalues are associated with the fastest processes. The clustered rate matrix  $Q_c$  has the smallest eigenvalue  $\lambda_1 = 0$ . The timescale separation can be measured by the ratio of the first non null eigenvalues, thus  $\frac{\lambda_3}{\lambda_2}$ . This ratio is observed over all titration steps.

In the bivalent ITC experiment, the eigenvalue ratio is 1.1 in the first titration step and grows to 2 in the last one. This can be interpreted that the second fastest process is almost twice as fast as the slowest process in the end of the experiment. These ratios are plausible and show a good timescale separation.

In contrast to that, for the trivalent ITC experiment the ratio of the first non null eigenvalues is 1.4 for all titration steps. There is obviously no separation in timescales. Possibly the projection of a trivalent model onto a two-dimensional subspace is not suitable? Maybe, a two-step macroprocess models this trivalent example better than the previously used one-stop process? To test this we can observe the ratio of the third and fourth eigenvalues, i.e.  $\frac{\lambda_4}{\lambda_3}$  instead of  $\frac{\lambda_3}{\lambda_2}$ .  $\frac{\lambda_4}{\lambda_3} = 1$  for the first titration step and  $\frac{\lambda_4}{\lambda_3} = 4.7$  for the last one. This growth of eigenvalue ratio fits better the expected behavior. However, one drawback is that clustering on a three-dimensional  $Q_c$  matrix would be impossible to compare to the results obtained from the Wiseman fitting.

Another point to critically discuss is the robustness of the fittings. This can be tested by disturbing the input parameters before clustering and compare the disturbed to the undisturbed output. Every value of the heat curve (input) was

## 5. WISEMAN FITTING

---

disturbed by up to 1%. Disturbing the bivalent ITC example yields changes of the output below 1%. Compared to the undisturbed model  $n$  is 0.33% higher,  $\Delta H$  decreased by 0.31% and the best norm decreased by 0.18%. The binding and unbinding parameters change even less:  $k_{on1}$ ,  $k_{on2}$ ,  $k_{off1}$  and  $k_{off2}$  differ by a factor of  $10^{-6}$  compared to the undisturbed model. Therefore,  $K_1$  and  $K_2$  do not differ either. This method is very robust for a bivalent setting.

Disturbing the input of the trivalent experimental data yields a difference of the binding and unbinding parameters of up to 743%. Their ratios  $K_1$ ,  $K_2$  and  $K_3$  thus differ substantially, too: up to 653%.

However, the stoichiometry and the heat of binding are similar. In the disturbed model, the stoichiometry  $n$  is only 0.04% higher, and  $\Delta H$  is 1.5% smaller. The best norm decreased by 1.36% compared to the undisturbed model. The spectral gap analysis already indicated the fact that a  $Q_c$  fitting distinguishing only between bound and unbound does not deliver satisfying results. The error estimates are another confirmation. It needs further study, but likely the Wiseman fitting is not the suitable fitting for trivalent bindings.

This chapter has presented a different way of fitting data from ITC experiments in order to gain kinetic information of intermediate binding steps using the Wiseman fitting. Both ways of fitting may help to explain more complex heat curves where a Wiseman curve would be too simplistic. This method has been applied to a bivalent and a trivalent binding process. Both ITC data sets have been fit with the Wiseman fitting method and compared to the  $Q_c$  fitting method, which resembles the previously shown kinITC+ method. It turns out, that the goodness of fit is better for the  $Q_c$  fitting method, at least at first sight for these two examples. One must be careful to generalize these results, as it was only applied for two examples with low valencies. Firstly, the evaluation of the so called better fitting method was in terms of minimal absolute errors. One could argue, if this is the best decision criteria. Secondly, the spectral gap analysis and the error estimates have shown that the  $Q_c$  fitting is suitable for the bivalent case but not necessarily for the trivalent case.

Lastly, the algorithm for the  $Q_c$  fitting in this chapter is using a random search to find the optimal binding parameters. Therefore, the algorithm does not always produce the exact same results. The algorithm could be enhanced further such

## 5.5 Limitations of the Model

---

that the best starting point is used. Further, again a one to one stoichiometry is assumed, which does not always model true binding mechanisms happening naturally.

## 5. WISEMAN FITTING

---

# Chapter 6

## Rebinding

In the previous two chapters, the overall binding rates were quantified. This chapter is dedicated to the so-called rebinding effect where the probability of a monovalent binding event of multivalent ligands is increased by the high local concentration of ligands with no regard to potential multivalent receptor binding. For advanced drug therapy the rebinding effect is particularly interesting. What if a high level of target occupancy can be achieved with only short and thus negligible interruptions? If a drug molecule stays around the pain receptor in the human tissue after it unbound, there is a certain likelihood that molecule and ligand will reattach. In this case, the drug will work again. Exploiting this kind of rebinding may be useful in drug design. If it was known if and to what extent rebinding does occur, the drug doses might be lowered, which in turn also reduced undesirable side effects.

So far there was no distinction made between rebinding and new binding in this work. There already exist a quantification of the lower bound of rebinding in the system [11; 70]. In this chapter, such quantification will be applied numerically for the ITC data presented in Chapter 4.

### 6.1 Biochemical and Mathematical Perspectives on Rebinding

In the literature, there exist different definitions of rebinding. From the pharmacologists' point of view, rebinding takes place when the amount of free ligands is not reduced by e.g. wash-out experiments. Another point of view, which the author also shares, is that rebinding means a consecutive binding of freshly dissociated ligands to the same receptors or those in the immediate vicinity [22]. Vauquelin states in [71] that "drug-target interactions are represented as a reversible single-step bimolecular process and pharmacokinetic elimination rates rely on drug concentrations in the plasma, not in the vicinity of the target itself." In the last years, the role of binding kinetics has been evolved as a relevant research topic. So far it has been focused on the  $k_{\text{off}}$  rate as the key property of many marketed drugs because it is assumed that slow drug dissociation is equivalent to low  $k_{\text{off}}$  rates [71]. According to Zhang, there is a missing link between the measured unbinding rates in vitro and the duration of the drug's therapeutic effect in vivo [71; 72].

The rebinding effect is commonly known as a 'short-time memory' effect [11; 22]. Mathematically, it occurs when projecting a system onto a smaller state-space. By clustering of a molecular process onto the macro states 'bound' and 'unbound', it loses information about the exact spatial arrangement of the receptors and ligands in consideration as illustrated in Figure 6.1. In the macroscopic view, the process is Markovian. That means that the system's state does not depend on its historical states. In the microscopic view, Markovianity is lost. That is due to the fact that shortly after the dissociation of a receptor-ligand-complex, both molecules are still spatially close and therefore much more likely to bind again, i.e. their rebinding is more probable than binding of two molecules far apart. Thus, the spatial situation of ligand and receptor is no longer negligible on the molecular scale and the current state does depend on the past [11].

This effect is discussed especially in the context of multivalent molecules, where the spatial connection of molecules can lead to increased rebinding events. This can be imagined such that one dissociated ligand is 'held at its place' by its joint



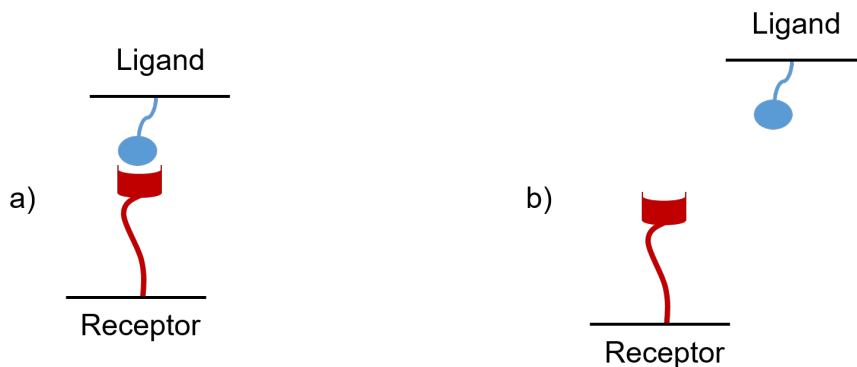


Figure 6.1: Rebinding effect: a) Spatial constellation after dissociation. b) Spatial arrangement at arbitrary time. These two configurations of a receptor and a ligand represent the same macro state ('unbound') and are not distinguishable in the clustered model, even though different binding probabilities are expected by the receptor-ligand-distance on the microscopic scale.

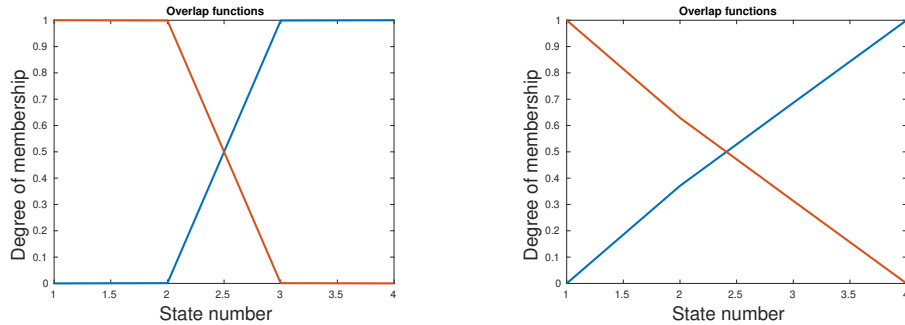
molecules. However, the strength of this effect depends on different parameters such as the rigidity or flexibility of the spacer [23]. The quantity of such rebinding events is investigated in [11] for reversible and in [70] for general processes, i.e. including non-reversible processes. This quantity depends on the choice of the projection, more clearly on the fuzziness of the membership functions that are chosen.

## 6.2 Quantification of Rebinding

As described in Section 3.3 by applying the PCCA+ clustering algorithm, we project a high dimensional rate matrix  $Q$  onto a two-dimensional space: 'bound' and 'unbound'. As part of PCCA+, the two membership functions  $\chi_1, \chi_2 : \Omega \rightarrow [0, 1]$  for the unbound and bound states are calculated. The columns of  $\chi$  give good insight into the rebinding behavior of the binding process. If there is a relatively big overlap between two consecutive states, i.e. from unbound to singly bound, then rebinding is very likely. As the process belongs to two states, it can easily jump back and forth between both of them. If the clustering is crisp, i.e. there is no or very little overlap, rebinding is unlikely. For example, there are four different macro states in a trivalent binding process: unbound, one bond, two

## 6. REBINDING

bonds and three bonds (fully bound). In this example we focus on the transition from state 2 to state 3, i.e. singly bound to doubly bound. In Figure 6.2 the left hand graph shows a crisp clustering, where the membership is either 0 or 1 for the unbound and fully bound state (state number 1 and 4, respectively). The transition is very clear cut with little overlap between state 2 and 3 and thus there is no rebinding. The right hand graph of Figure 6.2 shows that the membership functions have flatter slopes, they meander from one state to the other. Therefore rebinding is very likely.



(a) Crisp clustering.

(b) High overlap of the clusters.

Figure 6.2: Rebinding effect from the membership function perspective: a high overlap of membership functions yield much rebinding, while a crisp clustering makes rebinding unlikely.

Where does this overlap come from? When clustering reversible Markov processes, the system needs to commute between time propagation and projection onto the invariant subspace.

$$\begin{array}{ccc}
 Q(\tau) & \xrightarrow{\tau \rightarrow \tau^k} & Q(\tau)^k \\
 \downarrow & & \downarrow \\
 Q_C(\tau) & \xrightarrow{\tau \rightarrow \tau^k} & (Q_C(\tau))^k
 \end{array}$$

If the diagram was commutative, the order of time propagation and Galerkin projection would not play a role. Otherwise, there is the inequality

$$\Pi(Q^k) \neq (Q_c)^k$$

leading to two kinds of possible errors:

## 6.2 Quantification of Rebinding

---

- Iteration error:  $e_i = \Pi(Q^k) - (\Pi(Q))^k$
- Rebinding error: There is a memory effect between two consecutive states, thus the process is not Markovian.

The overlap can be determined by the scalar product  $\langle \chi, \chi \rangle$ . To quantify the rebinding we need a weighted mass matrix  $S$ .

$$S = D^{-1} \langle \chi, \chi \rangle_\pi \quad (6.1)$$

with  $D = \text{diag}(\pi_c)$ . Note that the scalar product  $\langle \chi, \chi \rangle$  is weighted by the normalized distribution  $\pi$  *before* the clustering and the two diagonal elements of  $D$  are the normalized distribution *after* the clustering.

If  $\chi$  was unknown because neither PCCA+ nor GenPCCA was performed,  $S$  can be obtained by the optimization process stated below [11]:

We know that  $A^{-1}A = I$  and that the first column of  $A^{-1}$  is the first eigenvector (constantly one) of our discretized rate matrix  $Q$ , whereas the other columns of  $A^{-1}$  are arbitrary multiples of the other eigenvectors.

$$A^{-1} = \begin{pmatrix} 1 \\ \vdots \\ \alpha_2 X_2 & \cdots & \alpha_n X_n \\ 1 \end{pmatrix}, \alpha_2, \dots, \alpha_n \in \mathbb{R}.$$

By the following optimization,  $S$  is determined:

$$\begin{aligned} & \min_{\alpha_1, \dots, \alpha_n \in \mathbb{R}} |\det(S) - 1| \\ & \text{such that } \alpha_1 = 1, \\ & A_{ij}^{-1} = \alpha_i X_{ij} \quad \forall i, j, \\ & S = D^{-1} A^T A, \\ & S_{ij} \geq 0 \quad \forall i, j \end{aligned} \quad (6.2)$$

$S$  can be interpreted as a weighted row-stochastic overlap matrix. The characteristics of  $S$  give insights into the degree of rebinding. Using  $S$  we can calculate

## 6. REBINDING

---

the transition matrix  $T$  by

$$T = D^{-1} \langle \chi, Q\chi \rangle_\mu = D^{-1} A^T \Lambda A$$

with  $A$  the matrix solving the optimization problem from equations 3.10 and 3.11. Both  $S$  and  $T$  give us information about the rebinding. After clustering, one can obtain  $T$  by  $T = SQ_c$ . The determinants of both  $S$  and  $T$  are linked by

$$\begin{aligned} \det(T) &= \det(S) \det(A^{-1} \Lambda A) \\ &= \det(S) \det(\Lambda) \\ &= \det(S) \prod_{i=1}^n \lambda_i \end{aligned} \tag{6.3}$$

$T$  is the matrix denoting the movement from the micro state to the macro state.  $T_{ij}$  is the portion of micro states starting in the macro state  $\chi_i$  that ends in the macro state  $\chi_j$  within time  $\tau$  [11]. The state transition matrix  $T$  and the overlap matrix  $S$  are responsible for the transition probability matrix  $P$  by

$$P = S^{-1}T$$

which in turn is linked to the transition rate matrix by

$$P(\tau) = \mathbf{exp}(\tau Q).$$

The eigenfunctions of  $P$  and  $Q$  are equivalent. The eigenvalues  $\psi$  of  $Q$  and  $\lambda$  of  $P$  differ according to the relationship  $\exp(\psi_i) = \lambda_i$ . Thus the stable macro states show the maximum eigenvalues of 1 and 0, respectively:  $\max \lambda_i = 1$  and  $\max \psi_i = 0$ .

For actually quantifying the rebinding effect, we need to assess the magnitude of overlap between the conformations.

## 6.2 Quantification of Rebinding

---

Let  $F := -\text{trace}(Q_c)$  be the stability indicator of the molecular system

$$\begin{aligned}
 F &= -\text{trace}(Q_C) \\
 &= -\tau^{-1} \log(\exp(\text{trace}(\tau Q_C))) \\
 &= -\tau^{-1} \log(\det(\exp(\tau Q_C))) \\
 &= -\tau^{-1} \log(\det(P_C \tau)) \\
 &= \tau^{-1} (\log(\det S) - \log(\det T))
 \end{aligned}$$

For a proof, see Weber and Fackeldey in [11].  $F$  was chosen this way because it equals the negative sum of the leading eigenvalues of  $Q$ . In their paper [11] the authors show how to optimize  $S$  because the transformation matrix  $A$  is not known.  $A$  determines the clustered rate matrix by

$$Q_c = A^{-1} \Sigma A = (\langle \chi, \chi \rangle_\mu)^{-1} \langle \chi, Q \chi \rangle_\mu.$$

In our practical example, we already applied PCCA+ and hence we know  $A$ . Therefore we do not need to optimize  $S$ , instead we can compute it directly by

$$S = \chi^T D_\mu \chi.$$

Since  $S$  is a stochastic matrix, the determinant lies between 0 and 1.  $F$  is low, if the determinant of  $S$  is low. The lower  $F$ , the more stable is the system. Stability means in this context that the probability of a ligand-receptor system to stay in a certain state is close to 1. If  $S$  is diagonal dominant, there is little rebinding taking place. If  $\text{trace}(S) \approx 0$ , there is much rebinding. If  $\text{trace}(S) = \text{rank}(S)$  there is no rebinding at all. Another hint is given by the determinant. If  $\det(S)$  is close to 1, the system is not stable, if  $\det(S)$  is close to 0, it is stable [70]. Since we projected onto a two-dimensional subspace, the overlap matrix  $S$  is a stochastic  $2 \times 2$  matrix. In this case it is equivalent to assess the determinant or the trace because  $\text{trace}(S) = 1 + \det(S)$ .

*Proof.* Assume a two-dimensional stochastic matrix  $S = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}$  with

## 6. REBINDING

---

$s_{11} + s_{12} = 1$  and  $s_{21} + s_{22} = 1$ .

$$\begin{aligned} \text{trace}S - \det S &= 1 \\ s_{11} + s_{22} - (s_{11}s_{22} - s_{12}s_{21}) &= 1 \\ s_{11} + s_{22} - s_{11}s_{22} + (1 - s_{11})(1 - s_{22}) &= 1 \\ s_{11} + s_{22} - s_{11}s_{22} + 1 - s_{22} + s_{11}s_{22} - s_{11} &= 1 \\ 1 &= 1 \end{aligned}$$

□

In [11; 70], the authors state that a two-dimensional clustered rate matrix is per se reversible and that therefore it is always possible to find a transformation matrix  $A$  yielding  $\det S = 1$ , and consequently yielding no rebinding. However, this does not apply if we use the transformation matrix obtained from the previous clustering.

### 6.3 Numerical Examples

The examples presented in the following section are the same as previously presented in Chapter 3. Originally the numerical data stem from Igde et al. [64]. After the clustering process, we are interested in the share of rebinding taking place over time, i.e. in our case in each titration step.

#### 6.3.1 Bivalent Ligand Example

The following example shows the minimum rebinding effect of interactions of bivalent mannose to tetrameric Con A. Within the clustering algorithm, the overlap matrix  $S$  and its trace have been computed. Figure 6.3 shows that there is almost no rebinding because  $\text{trace}(S) = \text{rank}(S) \approx 2$ . Only at later titration steps, the trace of  $S$  decreases a little, thus there is some rebinding taking place. As discussed in Subsection 4.1.1, the solution presented is not a unique solution of the optimization problem. However, all the combinations of possible  $k_{on}$  and  $k_{off}$  rates showed the same low degree of rebinding effect. The evolution of the trace of  $S$  similar for every combination yielding the maximum correlation coefficient.

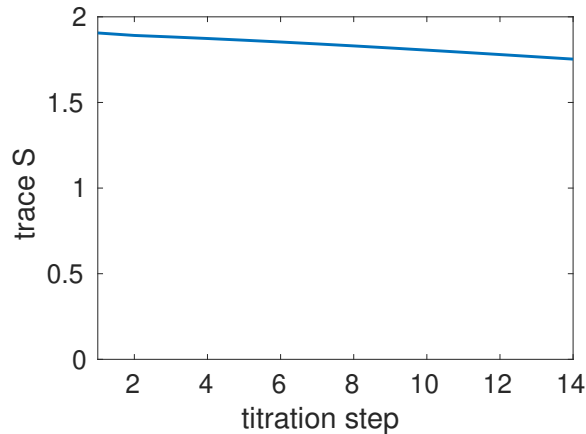


Figure 6.3: Rebinding effect for bivalent Man(1,5)-5 and tetrameric Con A, experimental data taken from [64] .

### 6.3.2 Trivalent Ligand Example

The next example illustrates the rebinding for the trivalent mannose Man(1,3,5)-5 binding with tetrameric Con A from Subsection 4.1.2. Even though, there were

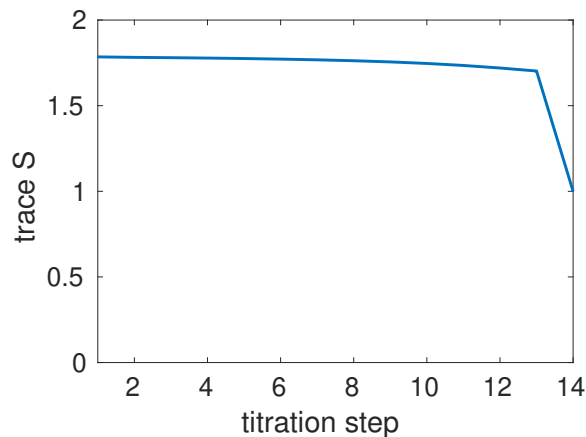


Figure 6.4: Rebinding effect for trivalent Man(1,3,5)-5 and tetrameric Con A, experimental data taken from [64] .

25 out of 15625 input combinations yielding the same correlation coefficient in the optimization process, all of these 25 possibilities yielded the same rebinding information. The trace of  $S$  is almost constant until the 13th titration step and then drops dramatically. That means, that the rebinding is very low for almost

## 6. REBINDING

---

the whole process and increases in the end after the receptor concentration passes a certain threshold.

Another reason for the kink in the rebinding curve might be the assumption that the last titration step was an outlier. Therefore, the trace of  $S$  was calculated again without it. The following figure shows the rebinding for only 12 titration steps. This rebinding behavior is interesting because it first goes up as

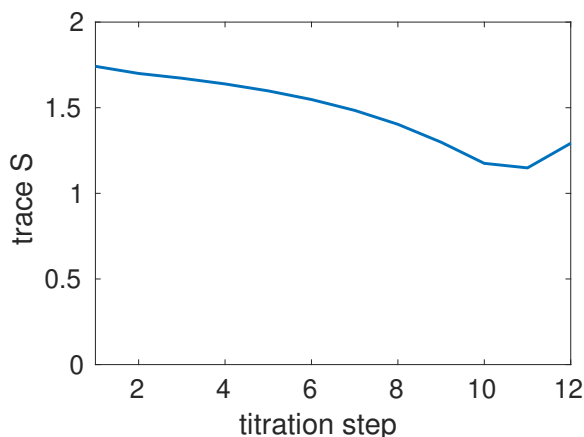


Figure 6.5: Rebinding effect of the first 12 titrations steps of trivalent Man(1,3,5)-5 and tetrameric Con A, experimental data taken from [64] .

$\text{trace}(S)$  goes down, and at a certain threshold it goes down again. The turning point is exactly where the experimental  $k_{on}$  rate curve has its inflection point. Of course, caution must be taken with the input parameters. The optimization algorithm process with only 12 titration steps shows slightly different values than for 14 steps, namely  $k_{on_i} = [1000, 1000, 10000]$  and  $k_{off_i} = [1, 1, 1]$  instead of  $k_{on_i} = [1000, 10000, 1000]$  and  $k_{off_i} = [1, 1, 1]$  as before. Still the interesting finding is that an initially low rebinding turns out to be high toward the end of the experiment.

### 6.3.3 Pentavalent Ligand Example

Comparing the bivalent and the trivalent examples with each other, the trace of the overlap matrix decreased throughout all titration steps, thus in the trivalent



case, there is more rebinding taking place. These two examples might lead to the proposition that rebinding activities increase the higher the valency number is. Due to the fact that there are more ligand-receptor couples that can quickly bind and unbind, contributing to a higher stability of the whole complex, one may suppose. However, the trace of the overlap matrix of the binding of pentavalent mannose to tetrameric Con A tells a different story as depicted in Figure 6.6.

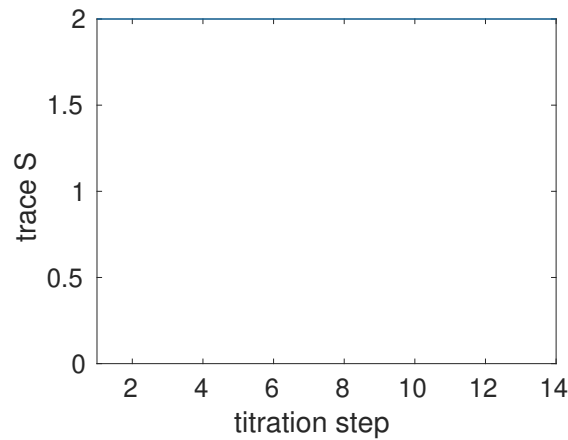


Figure 6.6: Rebinding effect for pentavalent Man(1,3,5,7,9)-9 and tetrameric Con A, experimental data taken from [64] .

Throughout all the 14 titration steps,  $\text{trace } S = \text{rank } S = 2$ . Thus, it appears that there is no rebinding at all. How is this possible? In their paper Fackeldey and Weber show a limit of the rebinding quantification theory with the following theorem:

**Theorem 2.** *Let the matrix  $Q_C \in \mathbb{R}^{n \times n}$  be reversible. Furthermore, let  $Q_C$  stem from a clustering with positive definite overlap matrix  $S$ . Then there exists a feasible matrix  $A \in \mathbb{R}^{n \times n}$  in the optimization problem 3.1 with  $\det(S_{opt}) = \det(D^{-1}A^T A) = 1$ .*

For a proof, cf. the rebinding paper [11]. In this case, the result would mean that we cannot tell if there is rebinding or not. The clustered rate matrix  $Q_C$  is indeed reversible. The big difference is that in their paper, the matrix  $Q$  is unknown. From the known matrix  $Q_C$  all possible  $\chi$ 's were calculated and their resulting determinant of the overlap matrix to find an upper boundary. In the

## 6. REBINDING

---

present case,  $\chi$  and  $S$  do not have to be optimized, but are obtained from the actual  $Q$  matrix. Concluding, the example above does not show the limitation of the rebinding quantification, but there really is *no rebinding*.

How can that happen if there are so many binding sites that may quickly associate and dissociate with one another? Rebinding entails that the molecule is jumping quickly many times between two processes before staying in one particular state. Eventually, the molecule may move to yet another state. If there was no rebinding, that would imply that there is no or hardly no overlap between two states. Instead of a step function,  $\chi$  shows a crisp clustering, as depicted in Figure 6.7. An

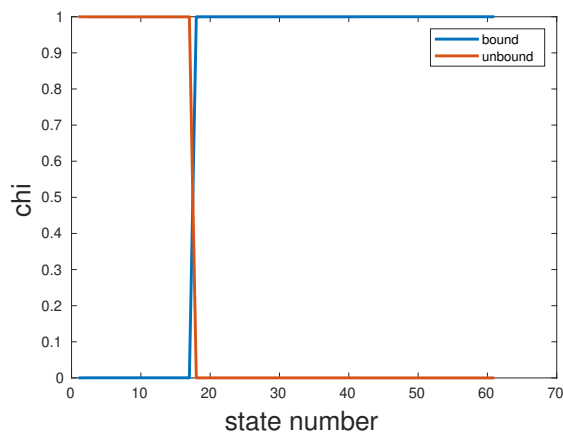


Figure 6.7: Membership functions of the binding of pentavalent Man(1,3,5,7,9)-9 to tetrameric Con A.

interesting finding is the moment when the bound state becomes predominant over the unbound state: it is exactly at the transition from state 17 to state 18. From the definition of the rate matrix in Subsection 3.2.2 and the number and order of states, we know that there is one unbound state (state number one) and sixteen different singly bound states (states two to 17). At state 18, the bivalently bound states start. When the system is unbound or has only one bond, the macro state is 'unbound'. If there are two or more bonds, the macro state instantly jumps to 'bound'. Thus, once the first bond is made, there is no jumping back and forth between macro-states, but the complex is bound.

## 6.4 Conclusion

The simple assumption, that the rebinding is higher, the higher the valency, is not true. There may be cases with a positive correlation of valency number and rebinding, but it is not a direct relationship. Other factors such as ligand type, assay, length and rigidity of the spacers etc., influence binding, and thus the rebinding.

There are two different time scales at work. High rebinding stems from fast micro processes, i.e. fast binding and unbinding of ligand-receptor couples. The binding and unbinding of the whole complex run on a slow timescale. In fact, the complex stays a complex very long even though the single micro processes change very fast. Thus, rebinding stabilizes the whole complex.

## 6. REBINDING

---

# Chapter 7

## Conclusion

This thesis aimed to discuss a way to model receptor-ligand binding processes from a stochastic perspective in contrast to the existing experimental studies. It has been shown that it is possible to cluster a high-dimensional state space into two states only: 'bound' and 'unbound'. This is achieved by projecting the  $n$ -dimensional state space of multivalent ligands and receptors on an invariant subspace without losing any crucial information. The subspace is 2-dimensional and represents the macroscopic binding states 'bound' and 'unbound'. For the clustering procedure, the following simplifying assumptions have been made to the model:

- The scaffold and its spacers are somewhat rigid. That means that not every ligand arm could bind to any arbitrary receptor binding site. No physical crossing of binding arms was allowed.
- The stoichiometry was assumed one to one. The complex can be formed with only one ligand entity and one receptor entity, respectively.
- The microscopic binding and unbinding rates have assumed to range on a logarithmic scale of base  $10^0$  up to  $10^4$ .

Respecting these assumptions, the rate matrix  $Q$  was set up. Importantly, this matrix is depending on the ligand concentration. Then  $Q$  was used to perform the clustering algorithm PCCA+ to gain the clustered rate matrix  $Q_C$ . In the course of the clustering, the membership vectors  $\chi$  have been computed. These can be used to calculate the overlap matrix  $S$ . The trace of  $S$  gives insight into

## 7. CONCLUSION

---

the thermodynamic contribution of the so-called rebinding. This relationship is inverse, that means the higher the trace of  $S$ , which has a minimum of 0 and a maximum of the matrix dimension, in this case 2, the lower the rebinding. As a proof of concept, this method has been applied to one set of experimental ITC data with bivalent, trivalent and pentavalent mannose ligands and dimeric and tetrameric Con A. This application has brought the following key findings:

1. The kinITC algorithm only averages binding rates. kinITC+ shows the change of the species in each time step.
2. The macroscopic binding rate is concentration dependent. kinITC+ confirmed, that overall  $k_{on}$  varies over time with each titration step.
3. Bindings of ligands and receptors are macroscopically slow processes. kinITC+ can zoom in and show processes of faster timescales.
4. The common assumption that rebinding increases with valency might be wrong. At least this relationship is not always verifiable as the experimental data presented in this thesis is one counterexample.

Especially the third finding may have a practical application in molecular dynamics research. So far it was difficult to determine at least a magnitude of the distinct binding coefficient per time step. The conformational changes happen too fast to be captured. By using the PCCA+ fitting we have a differential equation for the concentration dependent binding coefficients for each titration step. With the pseudo inverse of the clustered rate matrix we can compute the binding coefficients per time step. This gives us the concentration of the unbound and bound species that are in reality too transient to be measurable by an apparatus such as a spectrometer.

## Discussion

Simplifying the model came with some cost. The described modeling and clustering has some weaknesses that are discussed below.

---

One of the biggest restrictions of the stochastic model in this thesis is the assumption of a one to one stoichiometry. Experimental settings such as titrations however have shown that this is not always the case. Especially for large valencies such as decavalent ligands, aggregation does occur. That means that not a one to one complex is formed but rather a chain of several ligand and receptor entities. Another critical point is the generality of the method. This thesis is a proof of concept for the idea of invariant subspace projection, but the method has only been tested for one experimental study. It was difficult to gain ITC data with kinetic information. In most studies, the classical ITC method is being used without the information of free and bound ligands, receptors and complex per titration step. This information is crucial for applying kinITC and to obtain the macroscopic binding rate which is being compared to the clustered  $k_{on}$  rates.

Within the fitting method of clustered  $k_{on}$  there is the weakness of ambiguity of goodness of fit measures. Depending on the method chosen, such as mean squared differences or correlation coefficient, some other slightly different input combination is favorable. As the bivariate parameter study showed, one cannot tell that a specific microscopic binding or unbinding rate influences the macroscopic rates the most.

In comparison to the classical Wiseman fitting, the kinITC+ method delivers better results in terms of absolute errors. However, it has to be emphasized that the resulting binding and unbinding parameters are rather of qualitative than of quantitative nature. This method is not suitable to predict specific  $k_{on}$  and  $k_{off}$  values, instead it can be used for ratios and developments, i.e. if  $k_{on_1}$  is higher or lower than  $k_{on_2}$  etc.

Special care must be taken for valencies above two. While for bivalent settings, the kinITC+ method is very robust, the spectral gap analysis and the error estimate have shown the shortcomings of kinITC+ for at least a trivalent setting. To confirm and generalize this shortcoming, the method should be tested on different kinITC data of higher valencies. If again, the fitting is not robust, it is worth to

## 7. CONCLUSION

---

check for projections onto higher dimensional subspaces.

Furthermore, we learned from the data that the lower the valency number, the simpler is the overall  $k_{on}$  curve. The lack of high valency experimental data is a shortcoming. As part of future studies, the model could be challenged for kinITC data with decavalent ligands or even higher valencies.

As a last point it must be mentioned that the rebinding quantification has its limits too. There was no minimization process to determine the optimal overlap matrix  $S$ , instead  $S$  depends on the membership vectors from the clustering. In some unfavorable cases  $S$  can become the identity matrix. Then its trace equals its dimension  $m$  which usually stands for no rebinding. If  $\text{trace}(S) = m$ , one has to be cautious and should not conclude directly the absence of rebinding. In these cases, it is recommended to check how the overlap matrix looks like. If  $\text{trace}(S) = m - \epsilon, \epsilon > 0$ , the lack of rebinding is more likely.

## Outlook

The last section of the thesis is dedicated to potential future research.

As mentioned in the previous section, the model should be tested for high valency binding settings. If the valency is big, say ten or higher, the rate matrix will become exponentially big. Despite its sparse nature, the computation for the eigenvalue (reversible binding process) or the Schur (irreversible binding process) decomposition will become tedious. An alternative way to define it, would be to group the distinct states together. That means, there would be one unbound state, only one singly bound state etc., plus one fully bound state. The  $Q$  matrix for  $n$ -valent bindings would be of dimension  $\mathbb{R}^{(n+1) \times (n+1)}$ . It has to be tested if these compact rate matrices lead to different clustering results than the extensive rate matrices used above.

In this thesis, a one to one stoichiometry has been assumed for each fitting. This assumption may be an oversimplification. As a future research work, the kinITC+ fitting could be adjusted for different stoichiometries. Finally, the best



---

fitting stoichiometry could be determined for each binding setting.

Another aspect concerning higher valencies than two is that the fitting is not always robust. The error estimate study showed that the model is not stable for all trivalent bindings. Instead of projecting onto a  $2 \times 2$  matrix for PCCA+, a  $3 \times 3$  matrix might be the right choice for trivalent settings. For higher valencies it has to be tested whether a  $3 \times 3$  clustered rate matrix is enough or if the clustered matrix grows with the valency.

For kinITC+ one needs to set up the transition rate matrix. The present thesis presented one way of defining it, yet there are alternative approaches of how to obtain the transition rate matrix using square root approximation (SQRA). The state space is discretized according to a Voronoi tessellation to yield a sparse rate matrix. There is basic work already done by Donati et al. [73] and Lie et al. [74]. With this sparse  $Q$  matrix, PCCA+ is performed as usual. In this model, the states of the system are the physical locations of the bindings. So far the states were modelled according to which ligand connects to which receptor. In this potential new model, the receptors would be assumed fixed and the space would be discretized in a grid with periodic boundary conditions. If the boundary conditions are chosen to be Dirichlet or periodic, it is proven that the SQRA operator converges the Fokker-Planck operator (which is the Smoluchovski operator in conformation dynamics) [75].

The particles, i.e. receptor and ligand, move in a Lennard Jones potential  $V(r) = -\frac{A}{r^6} + \frac{B}{r^{12}}$  with  $A$  the particle attraction parameter,  $B$  the particle repulsion parameter and  $r$  the distance between ligand and receptor centers. The  $Q$  matrix would then be filled according to the SQRA. If receptor and ligand were in the same grid, we would consider them bound. For simplicity, the diffusion would be held constant in each grid. Then convection can be added to the model, which is a rate with one predominant direction. One can think of it as a flow. The potential could further be modeled with a spring constant such that  $V = k(r-r_0)^2$  with  $k$  the spring constant,  $r$  the current distance between ligand and receptor and  $r_0$  some fixed distance such as  $1nm$  or  $1\text{\AA}$ . The flux can be estimated by  $\exp(-\phi Q_{ii}(t)) = P(x_0 \in i | x_t \in i)$ .

## 7. CONCLUSION

---

Basically the state space would be discretized as adjacency matrix. For yet further advanced studies, some barriers may be included in the state space. That means that some entries in the adjacency become zero to model that a direct passage of the particle is impossible.

# Zusammenfassung

Multivalente Bindungen lassen sich hervorragend für gezieltes Wirkstoffdesign in der Medikamentenforschung nutzen. Theoretisch handelt es sich um seltene stochastische Ereignisse. Es gibt bisher umfangreiche Literatur zu Multivalenz, insbesondere zu bestimmten Aspekten der Versuchsanordnung. Jedoch gibt es bis dato kaum theoretische Studien, die sich auf  $n$ -valente Bindungsprozesse verallgemeinern lassen. Diese Dissertation zielt darauf ab, diese Lücke zu schließen, indem der Bindungsprozess als kinetische Ratenmatrix modelliert wird auf die der Clustering-Algorithmus PCCA+ angewandt werden kann. Während das Binden und Lösen der einzelnen Ligand-Rezeptor-Paare auf einer schnellen Zeitskala abläuft (Mikroperspektive), wirkt die Assoziation oder Dissoziation eines Komplexes auf einer langsamen Zeitskala (Makroperspektive). Die aktuell populäre kinITC-Methode erfasst den Wechsel zwischen Zeitskalen jedoch nicht. Daher beschreibt die in dieser Arbeit vorgeschlagene Methode eine kinITC-Erweiterung, die hier kinITC+ genannt wird.

Die Hauptideen dieser Doktorarbeit sind:

- Es ist möglich, kinetische Informationen aus thermodynamischen Daten zu gewinnen.
- Die makroskopische Bindungsrate  $k_{on}$  ist nicht konstant, sondern abhängig von der Ligandenkonzentration.
- Es gibt mindestens ein Gegenbeispiel zu der Annahme, dass der thermodynamische Beitrag des Rebinding mit der Valenz zunimmt.



# Appendix

The following appendix shows the MATLAB code used for the data preparation, PCCA+ clustering and binding coefficient fitting.

## main

```
1  % Input:
2  % concentration from ITC experiments
3  % calculated k_on rates from ITC experiments
4  % after reading the experimental excel files
5  % Output:
6  % correlation coefficient of Qc/kon and k_on-ITC
7  % best fit Qc/con
8  % maximum correlation coefficient
9  % index of the max correlation coefficient input parameters
10
11 n_valent = 3;
12 n_titration = 14;
13 n = n_titration;
14 filenames = 'Man1355LLBB25.xlsx';
15 [k_on_ITC, c0] = run_titration_mult_exp_with_output(char(filenames));
16 k_on_ITC = k_on_ITC(1:n_titration);
17 con = c0(1:n_titration,2);
18 x = 1:n; % plot range of ITC measurements
19 %%set input parameters
20 kon = [0 3];
21 koff = [0 3];
22 steps = 4;
23 B = ones(steps, n_valent);
24 B = logspace(kon(1), kon(2), steps)' .* B;
25 B = [B B];
26 kon_koff_combi = zeros(steps^size(B,2), size(B,2));
27 for k=1:size(B,2)
28     kon_koff_combi(:,k)=repmat(reshape(repmat(B(:,k)', steps^(size(B,2)-k), 1)
29     , [], 1), steps^(k-1), 1);
29 end
30 % sensitivity analysis
31 corr = zeros(length(kon_koff_combi), 1);
32 diffsq = zeros(length(kon_koff_combi), 1);
33 trS = zeros(length(kon_koff_combi), n);
34 dS = zeros(length(kon_koff_combi), n);
35 Qc_kon = zeros(length(kon_koff_combi), n);
36 Qc_koff = Qc_kon;
37 Q=create_Q_matrix(n_valent, kon_koff_combi(1,1:n_valent), kon_koff_combi(1,end-
38     n_valent+1:end), con(1)); % big Q
[1,m] = size(kon_koff_combi);
```

```

39 mse = zeros(1,1);
40 for j = 1 : l %length(kon_koff_combi)
41     k_on = kon_koff_combi(j,1:n_valent)
42     k_off = kon_koff_combi(j,end-n_valent+1:end) % if the koff are not equal
43     [ corr(j), diffsq(j), Qc_kon(j,:), Qc_koff(j,:), trS(j,:), dS(j,:)] = Qc_fit
         (Q, k_on.ITC, con, n_valent, n, k_on, k_off,j);
44     mse(j) = immse(k_on.ITC, Qc_kon(j,:));
45     if isreal(corr(j)) == false
46         corr(j) = 0;
47     end
48 end
49 kmax = sum(sort(corr)==max(corr))
50 %% Plot results
51 [max_c, ind_maxc] = maxk(corr,kmax)
52 kmin = sum(sort(diffsq)==min(diffsq))
53 [min_c, ind_mindiffsq] = mink(diffsq,kmin)
54 [min_c, ind_mindiffsq] = min(diffsq)
55 [mindiffmaxcorr, ind] = min(diffsq(ind_maxc))
56 ind = ind_maxc(ind)
57 input = kon_koff_combi(ind,:)
58 Qc = Qc_kon(ind, :)
59 kon = Qc
60 koff = Qc_koff(ind,:)
61 Kd = koff./kon
62 Qc_kon(ind)/Qc_koff(ind)
63 figure(1)
64 set(gca, 'FontSize',18)
65 subplot(1,2,1)
66 plot(Qc, 'LineWidth', 2)
67 xlabel('titration step', 'FontSize',18)
68 ylabel('k_{on} [1/(M s)]', 'FontSize',18)
69 title('computational ')
70 subplot(1,2,2)
71 plot(k_on.ITC, 'LineWidth', 2)
72 xlabel('titration step', 'FontSize',18)
73 ylabel('k_{on} [1/(M s)]', 'FontSize',18)
74 title('experimental ')
75 % plot both the experimental kon and the computed one in one diagram
76 A = [Qc', ones(n,1)]
77 sol = A\k_on.ITC'
78 Qc2 = sol(2) + sol(1).* Qc
79 figure(2)
80 set(gca, 'FontSize',18)
81 plot(Qc2, 'LineWidth', 2)
82 xlabel('titration step', 'FontSize',18)
83 ylabel('k_{on} [1/(M s)]', 'FontSize',18)
84 hold on
85 plot(k_on.ITC, '*r')
86 figure(3)
87 subplot(1,3,1)
88 plot(kon)
89 title('kon')
90 subplot(1,3,2)
91 plot(koff)
92 title('koff')
93 subplot(1,3,3)
94 plot(Kd)
95 title('Kd')
96
97 % calculate rebinding for the optimal kon and koff
98 [traceS, detS] = rebinding(n, n_valent, kon_koff_combi(ind, 1:n_valent),
    kon_koff_combi(ind, n_valent+1: end), con, Qc)

```

```

99 % plot rebinding
100 traceS = trS(ind,:)
101 detS = dS(ind,:)
102 legend('computational kon', 'experimental kon')
103 figure(4)
104 plot(traceS)
105 hold on
106 plot(detS)
107 title('Rebinding', 'FontSize',18)
108 xlabel('titration step', 'FontSize',18)
109 ylabel('trace S', 'FontSize',18)
110 figure(5)
111 plot(trS(ind_maxc,:), 'LineWidth', 2)
112 %title('Rebinding (trace of overlap matrix S)', 'FontSize',18)
113 xlabel('titration step', 'FontSize',18)
114 ylabel('trace S', 'FontSize',18)
115 set(gca, 'FontSize',18);
116 axis([1 14 0 2])

```

## $Q_C$ fit

```

1 function [ corr, diffsq, Qc_kon , Qc_koff, traceS, detS ] = Qc_fit (Q, k_on_ITC,
2     con, n_ligands , n, k_on, k_off, j)
3 Qc_kon = zeros(n,1);
4 traceS = zeros(n,1);
5 detS = zeros(n,1);
6 F = detS;
7 for i = 1:n % loop for every ITC injection
8     Q = update_Q_matrix(Q, n_ligands, k_on, k_off, con(i)); % exponentially
9     many states
10    %cluster algorithm
11    [u, ~]=eig(Q');
12    [~, ind]=min(real(u(1,:)).*real(u(end,:)));
13    u=[ones(size(u,1),1), u(:,ind)*sign(u(1,ind))];
14    chi=cluster_by_isa(u, 2);
15    [pi, one]=eigs(Q,1,'lr');
16    pi=pi/sum(pi);
17    Qc=transpose(inv(chi'*diag(pi)*chi)*(chi'*diag(pi)*Q'*chi));
18    F(i)=-trace(Qc);
19    [pi_reb, one]=eigs(Qc,1,'lr');
20    pi_reb=pi_reb/sum(pi_reb);
21    [v, ~]=eig(Qc');
22    [~, ind]=min(real(v(1,:)).*real(v(end,:)));
23    v=[ones(size(v,1),1), v(:,ind)*sign(v(1,ind))];
24    chi2=cluster_by_isa(v, 2);
25    S = Stemp./ (sum(Stemp)'); % make S stochastic
26    traceS(i) = trace(S);
27    detS(i) = det(S);
28    % test whether Qc fits the k_on values:
29    Qc_kon(i) = Qc(2,1) / con(i);
30    Qc_koff(i) = Qc(1,2);
31 end
32 corr = corrcoeff(Qc_kon, k_on_ITC);
33 corr = corr(1,2);
34 diff_kon = Qc_kon-k_on_ITC';
35 diffsq = sum(abs(diff_kon));
36 end

```

## Extract titration information from experimental data sets

This function has been written by Susanna Röblitz.

```

1 function varargout = problem_titration_mult_exp(x, flag, par)
2 %general kinetics:  $sT*sR*RR = sC*CC$ 
3 %estimate single Kd and multiple kon, deltaH_V0 from all 14 experiments
4 %   Problem function.
5 %   VAR = PROBLEM(X,FLAG,par) computes the problem function f(X),
6 %   Jacobian df/dx(X), startvector for the Gauss-Newton
7 %   iteration and vector of to be fitted values.
8 % FX = PROBLEM(X, '', par) returns the right hand side f(X) for a input column
   vector X.
9 % JAC = PROBLEM(X, 'jacobian', par) returns the Jacobian df/dx(X) for a input
   column vector X.
10 switch flag
11   case ''
12     [varargout{1:2}] = f_rhs(x, par) ;
13   case 'jacobian'
14     [varargout{1:2}] = jacobian(x, par);
15   otherwise
16     error(['Unknown flag '' flag ''.']);
17 end
18 end
19 function [f, ifail] = f_rhs(x, par)
20 ifail=1;
21 %extract user parameters
22 nt=par.numt;
23 t=par.timepoints;
24 c0=par.conc_ini;
25 cend=par.conc_eq;
26 nexp=par.nexp;
27 %uncomment rows if Kd is fixed
28 Kd=par.Kd;
29 %uncomment rows if Kd is estimated
30 %Kd=x(end);
31 %uncomment if stoichiometric coefficients are also estimated
32 % ln_sT=x(end-2);
33 % ln_sR=x(end-1);
34 % ln_sC=x(end);
35 % sT=exp(ln_sT);
36 % sR=exp(ln_sR);
37 % sC=exp(ln_sC);
38 %uncomment if stoichiometric coefficients are fixed
39 s=par.coeff;
40 sT=s(1);
41 sR=s(2);
42 sC=s(3);
43 %f=zeros(nt+3,nexp);
44 f=zeros(nt, nexp);
45 for iexp=1:nexp
46   %extract injection specific user parameters
47   kon=x(2*(iexp-1)+1)
48   deltaH_V0=x(2*(iexp-1)+2)
49   TT0=c0(iexp,1);
50   RR0=c0(iexp,2);
51   CC0=c0(iexp,3);
52   koff=Kd*kon;
53   %solve the ODE system
54   options = odeset('RelTol',1e-4,'AbsTol',1e-4);

```



```

55     [~,Y] = ode15s(@(t,y) kinetic_fun(t,y,kon,Kd,sT,sR,sC),t,[TT0;RR0;CC0] ,
                    options);
56     %compute E
57     E=-deltaH_V0*(kon*sC*Y(:,1).^ (sT).*Y(:,2).^ (sR)-koff*sC*Y(:,3).^ (sC));
58     %add equilibrium concentrations to list of observations
59     f(:,iexp)=E;
60 end
61 f=reshape(f,nt*nexp,1);
62 ifail = 0;
63 end
64 function dy = kinetic_fun(t,y,kon,Kd,sT,sR,sC)
65 koff=Kd*kon;
66 dy = zeros(3,1); % a column vector
67 dy(1) = -kon*sT*y(1)^(sT)*y(2)^(sR)+koff*sT*y(3)^(sC);
68 dy(2) = -kon*sR*y(1)^(sT)*y(2)^(sR)+koff*sR*y(3)^(sC);
69 dy(3) = kon*sC*y(1)^(sT)*y(2)^(sR)-koff*sC*y(3)^(sC);
70 end
71 function [f,ifail] = jacobian(x,par)
72 ifail=1;
73 disp('Jacobian not implemented')
74 end

```

## Assessment of $k_{on}$ shapes

```

1 Qc_kon = zeros(length(kon_koff_combi), n);
2 Qc_koff = Qc_kon;
3 Q=create-Q-matrix( n_valent, kon_koff_combi(1,1:n_valent), kon_koff_combi(1,end-
    n_valent+1:end), con(1) ); % big Q
4 [l,m] = size(kon_koff_combi);
5 for j = 1: length(kon_koff_combi)
6     k_on = kon_koff_combi(j,1:n_valent);
7     k_off = kon_koff_combi(j,end-n_valent+1:end); % if the koff are not equal
8     [ corr(j), diffsq(j), Qc_kon(j,:), Qc_koff(j,:), trS(j,:), dS(j,:)] = Qc_fit
        (Q, k_on,ITC, con, n_valent, n, k_on, k_off,j);
9 end
10 unique_overall_kon = unique( Qc_kon, 'rows')
11 figure(1)
12 plot(log(unique_overall_kon'))
13 xlabel('titration step')
14 ylabel('log overall kon')
15 figure(2)
16 plot(unique_overall_kon')
17 xlabel('titration step')
18 ylabel('overall kon')

```

## $Q_C$ fit rebinding

```

1 function [ corr, Qc_kon ] = Qc_fit_rebinding(Q, k_on,ITC, con, n_ligands, n,
    k_on, k_off)
2 Qc_kon = zeros(n,1);
3 for i = 1:n % loop for every ITC injection
4     Q = create_compact-Q( n_ligands, k_on, k_off, con(i) )
5     %cluster algorithm
6     [u, ~]=eig(Q); % schur
7     [~, ind]=min(real(u(1,:)).*real(u(end,:)));
8     u=[ones(size(u,1),1),u(:,ind)*sign(u(1,ind))];

```

```

9     chi=cluster_by_isa(u, 2)
10    [pi, one]=eigs(Q,1,'lr');
11    pi=pi/sum(pi);
12    Qc=transpose(inv(chi'*diag(pi)*chi)*(chi'*diag(pi)*Q'*chi));
13    % test whether Qc fits the k_on values:
14    Qc_kon(i) = Qc(2,1) / con(i);
15  end
16  corr = corrcoef(Qc_kon, k_on_ITC);
17  corr = corr(1,2);
18  end

```

## rebinding

```

1  function [traceS, detS] = rebinding(n, n_ligands, kon, koff, con, Qc)
2  % this function calculates an estimate for the rebinding for the optimal kon and
   koff values
3  traceS = zeros(n,1);
4  detS = zeros(n,1);
5  for i =1:n
6      Q = create_compact_Q( n_ligands, kon, koff, con(i) );
7      [u, X]=eig(Q'); % schur
8      [~, ind]=min(real(u(1,:)).*real(u(end,:)));
9      u=[ones(size(u,1),1),u(:,ind)*sign(u(1,ind))];
10     chi=cluster_by_isa(u, 2);
11     [pi, one]=eigs(Qc,1,'lr');
12     pi=pi/sum(pi);
13     X = diag(1./pi)
14     Stemp = chi'*X*chi
15     S = Stemp./ (sum(Stemp'))); % make S stochastic
16     traceS(i) = trace(S)
17     detS(i) = det(S)
18  end
19  figure(3)
20  plot(traceS(:))
21  hold on
22  plot(detS(:))
23  legend('trace S', 'det S')
24  title('Rebinding')
25  xlabel('titration step')
26  end

```

## Sensitivity analysis

```

1  function [u,v, chi, Q, compare, minQc_kon, minQc_c_kon] =
   sensitivity_analysis_SYK_ITC(Q, k_on_ITC, con, n_ligands, n)
2  % this sensitivity analysis compares how well the SYK model fits the measured
   data from ITC experiments the matrix compare gives the output for all kinds
   of different inputs: kon1, kon2, koff1 = koff2. concentration
3  % then it is calculated: Qc(2,1), Qc(2,1)/con(i), abs(Qc(2,1) / con(i) ) -
   k_on_ITC(i)
4  % compare = [ kon1, kon2, koff1, con, Qc(2,1), Qc(2,1)/con(i), abs(Qc(2,1) / con
   (i) - k_on_ITC(i))]
5  % minQc_kon: all the rows from the compare matrix that display the minimum
   overall k_on
6  % minQc_c_kon: all the rows from the compare matrix that display the minimum
   differenece between overall k_on and k_on_ITC

```

```

7 [s, ~] = size(Q);
8 u1 = zeros(s, n);
9 u2 = zeros(s, n);
10 chi_all = zeros(s, 2, n);
11 Qc_all = zeros(2, 2, n);
12 % set all input parameters
13 steps = 5;
14 n = 14;
15 k_on_start = 0.01;
16 k_on_end = 4;
17 k_off_start = 0.01;
18 k_off_end = 4;
19 k_on = [k_on_start, k_on_start+0.01];
20 k_off = [k_off_start, k_off_start];
21 eps = 0.01;
22 h = n*steps *steps*(1+steps)/2;
23 kon = zeros(n_ligands, steps);
24 for j=1:n_ligands
25     kon(j,:) = linspace(k_on_start + (j-1)*eps, k_on_end, steps);
26 end
27 kon1 = linspace(k_on_start, k_on_end, steps);
28 kon2 = linspace(k_on_start+eps, k_on_end, steps);
29 kon3 = linspace(k_on_start+2*eps, k_on_end, steps);
30 kon4 = linspace(k_on_start+3*eps, k_on_end, steps);
31 koff1 = linspace(k_off_start, k_off_end, steps);
32 kon_mat = [kon1; kon2; kon3; kon4; koff1];
33
34 % elements = {kon1, kon2, kon3, koff1}; %cell array with N vectors to combine
35 elements{1, 1} = kon_mat(end,:);
36 for j=1:n_ligands
37     elements{1, j+1} = kon_mat(j,:);
38 end
39 combinations = cell(1, numel(elements)); %set up the varargout result
40 [combinations{:}] = ndgrid(elements{:});
41 combinations = cellfun(@(x) x(:), combinations, 'uniformoutput', false); %there
42 % may be a better way to do this
43 result = [combinations{:}];
44 h = length(result) * n;
45 compare = zeros(h, 5+n_ligands);
46 h=0;
47 for i = 1:n % loop for every ITC injection
48     for j = 1:length(result)
49         h=h+1;
50         k_on = result(j, 1:end-1);
51         k_off = result(j, end) * ones(length(k_on));
52         Q = update_Q_matrix(Q, n_ligands, k_on, k_off, con(i));
53         %cluster algorithm
54         [u, v]=eig(Q');
55         [~, ind]=min(real(u(1,:)).*real(u(end,:)));
56         u=[ones(size(u,1),1), u(:,ind)*sign(u(1,ind))];
57         chi=cluster_by_isa(u, 2);
58         [pi, one]=eigs(Q,1, 'lr');
59         pi=pi/sum(pi);
60         Qc=transpose(inv(chi'*diag(pi)*chi)*(chi'*diag(pi)*Q'*chi));
61         % test whether Qc fits the k_on values:
62         p = abs(Qc(2,1) / con(i) - k_on.ITC(i));
63         compare(h,:) = [con(i), Qc(2,1), Qc(2,1)/con(i), p, result(j,:)]];
64     end
65 end
66 [M, I] = min(abs(compare(:,6)));
67 minQc_kon = compare(I,:);
68 [M, I] = min(abs(compare(:,7)));

```

```

68 minQc_c_kon = compare(I, :);
69 end

```

## Cluster by isa

This function was written by Marcus Weber.

```

1  %% Die Funktion cluster_by_isa klassifiziert dynamische Daten anhand von
   NoOfClus Eigenvektoren (Evs) der Uebergangsmatrix. NoOfClus (default=2) legt
   dabei fest, wieviele Cluster es geben soll. Die verwendete Methode: Inner
   Simplex Algorithmus.
2  %% Restriktion: Evs muss mindestens NoOfClus Eigenvektoren beinhalten.
3  %% Beispiele:
4  %% cluster_by_isa(Evs, NoOfClus)
5  %% —> Ausgabe: Gibt einen Vektor cF aus, der zu jeder Zeile der
   Uebergangsmatrix eine Clusterzuordnung vornimmt.
6  %% [Chi, ind] = cluster_by_isa(Evs, NoOfClus)
7  %% —> Ausgabe: Gibt einen Vektor cF aus, der zu jeder Zeile der
   Uebergangsmatrix eine Clusterzuordnung vornimmt.
8  %% [cF, indic, Chi, RotMat] = cluster_by_isa(Evs, NoOfClus)
9  %% —> Ausgabe: cF ist wieder der Zuordnungsvektor. indic liefert den Indikator
   fuer die Eindeutigkeit der Zuordnung. Dieser Zahlenwert sollte ungefaehr
   Null sein. Chi gibt fuer jeden Cluster einen Vektor von
   Zugehoerigkeitsgraden im Sinne der Fuzzy-Theorie an. RotMat ist diejenige
   lineare Transformation, die Evs in Chi ueberfuehrt, sie sollte eine
   wohlkonditionierte Matrix sein. ind gibt die Eckenindizes aus
10 function [Chi, ind] = cluster_by_isa(Evs, NoOfClus)
11 %function [Chi, RotMat] = cluster_by_isa(Evs, NoOfClus)
12 % default-Zuweisung 2 Cluster
13 if nargin < 2
14     NoOfClus = 2;
15 end
16 % Sonderfaelle abfangen
17 if NoOfClus < 2
18     Chi = ones(size(Evs, 1), 1);
19     indic = 0;
20     RotMat = 1/Evs(1, 1);
21     cF = ones(size(Evs, 1), 1);
22 elseif (NoOfClus >= size(Evs, 1) & size(Evs, 2) == size(Evs, 1))
23     Chi = eye(size(Evs, 1));
24     indic = 0;
25     RotMat = inv(Evs);
26     cF = [1: size(Evs, 1)];
27 else
28     % Genuegend Eigenvektoren angegeben?
29     NoOfClus = round(NoOfClus);
30     if NoOfClus > size(Evs, 2)
31         NoOfClus = size(Evs, 2);
32     end
33     % Eigentlicher ISA-Algorithmus
34     C = Evs(:, 1:NoOfClus);
35     OrthoSys = C;
36     maxdist = 0.0;
37     % Ersten beiden Repraesentanten mit maximalen Abstand
38     for i = 1:size(Evs, 1)
39         if norm(C(i, :)) > maxdist
40             maxdist = norm(C(i, :));
41             ind(1) = i;
42         end
43 end

```

```

44     for i=1:size(Evs,1)
45         OrthoSys(i,:)=OrthoSys(i,:)-C(ind(1),:);
46     end;
47     % Weitere Repraesentanten ueber Gram-Schmidt Orthogonalisierung
48     for k = 2:NoOfClus
49         maxdist=0.0;
50         temp=OrthoSys(ind(k-1),:);
51         for i=1:size(Evs,1)
52             OrthoSys(i,:)=OrthoSys(i,:)-(temp*transpose(OrthoSys(i,:)))*temp;
53             distt=norm(OrthoSys(i,:));
54             if distt > maxdist
55                 maxdist = distt;
56                 ind(k)=i;
57             end
58         end
59         OrthoSys=OrthoSys./norm(OrthoSys(ind(k),:));
60     end
61     % Lineare Transformation der Eigenvektoren
62     RotMat= inv(C(ind,:));
63     Chi=C*RotMat;
64     % Bestimmung des Indikators
65     indic=min(min(Chi));
66     % Defuzzifizierung der Zugehoerigkeitsfunktionen
67     [minVal cF]=max(transpose(Chi));
68 end

```

## create Q matrix

```

1  function [ Q ] = create_Q_matrix( n_ligands , k_on , k_off , start_con , k )
2  % this function fills a state a matrix Q
3  % input: number of ligands or receptors , k_on and k_off vectors (each have
4  % n_ligands entries), a vector with the ligand concentration at t=0
5  % output: quadratic matrix Q with rowsum 0
6  % number of different states Z
7  temp = 0;
8  % number of different states for each conformation
9  % one default state: no bindings
10 % n_ligands conformations
11 states = ones(n_ligands+1, 1);
12 %states(2) = n_ligands * n_ligands;
13 for i=1:n_ligands-1
14     states(i+1) = n_ligands * nchoosek(n_ligands , i);
15     temp = temp+ nchoosek(n_ligands , i);
16 end
17 states(end) = n_ligands;
18 Z = 1+n_ligands+n_ligands*temp;
19 Q = eye(Z,Z);
20 % first fill first entry, first line and first column
21 Q(1,1) = -states(2)*k_on(1)+start_con(1);
22 % fill first column
23 Q(2:1+states(2),1) = k_on(1)*start_con(1);
24 % fill first line
25 Q(1,2:1+states(2)) = k_off(1);
26 % fill the off diagonal block matrices
27 for i=2:n_ligands
28     % fill the entries randomly in the smaller matrix off diagonal
29     [states(i),states(i+1)]
30     [i,1+n_ligands - i]
31     k_off(i)

```

```

32     mat = ran_mat4( [states(i),states(i+1)], [i,1+n_ligands - i], k_off(i) )
33     % put this matrix on the upper diagonal
34     Q(sum(states(1:i-1))+1 : sum(states(1:i-1))+states(i) , sum(states(1:i))+1 :
35         sum(states(1:i) + states(i+1)) = mat;
36     % transpose it, change the values and put it under the diagonal
37     if k_off(1) ~= 0
38         k_div = k_on(i)/k_off(1);
39     else
40         k_div = 0;
41     end
42     Q(sum(states(1:i))+1 : sum(states(1:i) + states(i+1), sum(states(1:i-1))+1
43         : sum(states(1:i-1))+states(i) ) = k_div*mat';
44     %Q(sum(states(1:i))+1: sum(states(1:i+1)) , states(i-1)+1:states(i-1)+states
45         (i)) = k_div*mat';
46 end
47 % fill the diagonal, the column sum must be 0
48 for i=1:Z
49     Q(i,i) = - (sum(Q(:,i)) - Q(i,i));
50 end
51 end

```

## Update Q matrix

```

1
2 function [ Q ] = update_Q_matrix( Q_old, n_ligands, k_on, k_off, start_con )
3 % this function updates an old Q matrix with new inout variables
4 % input: another Q matrix, number of ligands or receptors, k_on and k_off
5 %         vectors (each have
6 % n_ligands entries), a vector with the ligand concentration at t=0
7 % output: quadratic matrix Q with rowsum 0
8 % number of different states Z
9 temp = 0;
10 % number of different states for each conformation: 1 default state: no bindings
11 % and n_ligands bound conformations
12 states = ones(n_ligands+1, 1);
13 for i=1:n_ligands-1
14     states(i+1) = n_ligands * nchoosek(n_ligands, i);
15     temp = temp+ nchoosek(n_ligands, i);
16 end
17 states(end) = n_ligands;
18 Z = 1+n_ligands+n_ligands*temp;
19 Q = eye(Z,Z);
20 % first fill first entry, first line and first column
21 Q(1,1) = -states(2)*k_on(1)+start_con(1);
22 % fill first column
23 Q(2:1+states(2),1) = k_on(1)*start_con(1);
24 % fill first line
25 Q(1,2:1+states(2)) = k_off(1);
26 % fill the off diagonal block matrices
27 for i=2:n_ligands
28     % use the partial matrix from the previous Q matrix, just put new values on
29     % the non-zero entries
30     Q_temp = Q_old(sum(states(1:i-1))+1 : sum(states(1:i-1))+states(i) , sum(
31         states(1:i))+1 : sum(states(1:i) + states(i+1)));
32     % put this matrix on the upper diagonal
33     Q(sum(states(1:i-1))+1 : sum(states(1:i-1))+states(i) , sum(states(1:i))+1 :
34         sum(states(1:i) + states(i+1)) = Q_temp;
35     % transpose it, change the values and put it under the diagonal

```

```

31     Q(sum(states(1:i))+1: sum(states(1:i)) + states(i+1), sum(states(1:i-1))+1 :
        sum(states(1:i-1))+states(i) ) = k_on(i)*Q_temp';
32 end
33 % fill the diagonal, the column sum must be 0
34 for i=1:Z
35     Q(i,i) = - (sum(Q(:,i)) - Q(i,i));
36 end
37 end

```





# References

- [1] “Understanding the epidemic.” <https://www.cdc.gov/opioids/basics/epidemic.html>. Accessed: 2022-11-22. 1
- [2] “Fentanyl law enforcement submissions and increases in synthetic opioid-involved overdose deaths — 27 states, 2013–2014.” <https://www.cdc.gov/mmwr/volumes/65/wr/mm6533a2.htm>. Accessed: 2022-11-22. 1
- [3] M. Klimm, *New Strategies in Conformation Dynamics: Investigation of the  $\mu$ -opioid receptor in healthy and inflamed tissue*. PhD thesis, 2018. 1
- [4] “Who coronavirus disease (covid-19) dashboard.” <https://covid19.who.int/>. Accessed: 2022-11-22. 2
- [5] T. A. Tabish and M. R. Hamblin, “Multivalent nanomedicines to treat covid-19: a slow train coming,” *Nano Today*, vol. 35, 2020. 2
- [6] V. M. Krishnamurthy, L. A. Estroff, and G. M. Whitesides, “Multivalency in ligand design,” *Fragment-based approaches in drug discovery*, vol. 34, pp. 11–53, 2006. 2, 4, 19
- [7] M. Mammen, S.-K. Choi, and G. M. Whitesides, “Polyvalent interactions in biological systems: implications for design and use of multivalent ligands and inhibitors,” *Angewandte Chemie International Edition*, vol. 37, no. 20, pp. 2754–2794, 1998. 2, 4, 11, 12, 22, 23, 25
- [8] L. L. Kiessling, J. E. Gestwicki, and L. E. Strong, “Synthetic multivalent ligands in the exploration of cell-surface interactions,” *Current opinion in chemical biology*, vol. 4, no. 6, pp. 696–703, 2000. 2, 4
- [9] J. J. Lundquist and E. J. Toone, “The cluster glycoside effect,” *Chemical Reviews*, vol. 102, no. 2, pp. 555–578, 2002. 2, 4
- [10] C. Schütte and M. Sarich, *Metastability and Markov state models in molecular dynamics: modeling, analysis, algorithmic approaches*, vol. 24. American Mathematical Soc., 2013. 2, 5, 32, 33, 34, 35, 37, 45

- [11] M. Weber and K. Fackeldey, “Computing the minimal rebinding effect included in a given kinetics,” *Multiscale Modeling & Simulation*, vol. 12, no. 1, pp. 318–334, 2014. 3, 5, 87, 88, 89, 91, 92, 93, 94, 97
- [12] M. Merchant, X. Ma, H. R. Maun, Z. Zheng, J. Peng, M. Romero, A. Huang, N.-y. Yang, M. Nishimura, J. Greve, *et al.*, “Monovalent antibody design and mechanism of action of onartuzumab, a met antagonist with anti-tumor activity as a therapeutic agent,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 32, pp. E2987–E2996, 2013. 4
- [13] S. Howorka, J. Nam, H. Bayley, and D. Kahne, “Stochastic detection of monovalent and bivalent protein–ligand interactions,” *Angewandte Chemie International Edition*, vol. 43, no. 7, pp. 842–846, 2004. 4
- [14] A. Corazza, G. Verona, C. A. Waudby, P. P. Mangione, R. Bingham, I. Uings, D. Canetti, P. Nocerino, G. W. Taylor, M. B. Pepys, *et al.*, “Binding of monovalent and bivalent ligands by transthyretin causes different short-and long-distance conformational changes,” *Journal of medicinal chemistry*, vol. 62, no. 17, pp. 8274–8283, 2019. 4
- [15] C. W. Cairo, J. E. Gestwicki, M. Kanai, and L. L. Kiessling, “Control of multivalent interactions by binding epitope density,” *Journal of the American Chemical Society*, vol. 124, no. 8, pp. 1615–1619, 2002. 5, 9
- [16] C. Fasting, C. A. Schalley, M. Weber, O. Seitz, S. Hecht, B. Koksich, J. Dervedde, C. Graf, E.-W. Knapp, and R. Haag, “Multivalency as a chemical organization and action principle,” *Angewandte Chemie International Edition*, vol. 51, no. 42, pp. 10472–10498, 2012. 5, 9, 10, 11, 18, 21, 28, 39
- [17] J. Huskens, L. J. Prins, R. Haag, and B. J. Ravoo, *Multivalency: Concepts, Research and Applications*. John Wiley & Sons, 2018. 5, 9, 10, 11, 13, 17, 18, 28, 31
- [18] P. Dumas, E. Ennifar, C. Da Veiga, G. Bec, W. Palau, C. Di Primo, A. Piñeiro, J. Sabin, E. Muñoz, and J. Rial, “Extending itc to kinetics with kinitc,” in *Methods in enzymology*, vol. 567, pp. 157–180, Elsevier, 2016. 5, 29, 54
- [19] T. Egawa, A. Tsuneshige, M. Suematsu, and T. Yonetani, “Method for determination of association and dissociation rate constants of reversible bimolecular reactions by isothermal titration calorimeters,” *Analytical Chemistry*, vol. 79, no. 7, pp. 2972–2978, 2007. 5, 54, 78
- [20] A. Nielsen, *Computation schemes for transfer operators*. PhD thesis, Freie Universität Berlin, 2016. 5, 34
- [21] M. Weber, “Meshless methods in confirmation dynamics,” 2006. 5, 46, 48, 49

- [22] G. Vauquelin, “Rebinding: or why drugs may act longer in vivo than expected from their in vitro target residence time,” *Expert Opinion on Drug Discovery*, vol. 5, no. 10, pp. 927–941, 2010. 5, 88
- [23] M. Weber, A. Bujotzek, and R. Haag, “Quantifying the rebinding effect in multivalent chemical ligand-receptor systems,” *The Journal of Chemical Physics*, vol. 137, no. 5, 2012. 5, 89
- [24] S. Röhl, “Computing the minimal rebinding effect for nonreversible processes,” Master’s thesis, Freie Universität Berlin, 2017. 5, 35, 38
- [25] S. Röhl and K. Fackeldey, “Computing the minimal rebinding effect for nonreversible processes,” 2017. 5
- [26] C. E. Sing, M. Olvera de la Cruz, and J. F. Marko, “Multiple-binding-site mechanism explains concentration-dependent unbinding rates of dna-binding proteins,” *Nucleic acids research*, vol. 42, no. 6, pp. 3783–3791, 2014. 6
- [27] M. Mammen, S.-K. Choi, and G. M. Whitesides, “Polyvalent interactions in biological systems: implications for design and use of multivalent ligands and inhibitors,” *Angewandte Chemie International Edition*, vol. 37, no. 20, pp. 2754–2794, 1998. 9
- [28] J. E. Gestwicki, C. W. Cairo, L. E. Strong, K. A. Oetjen, and L. L. Kiessling, “Influencing receptor- ligand binding mechanisms with multivalent ligand architecture,” *Journal of the American Chemical Society*, vol. 124, no. 50, pp. 14922–14933, 2002. 9
- [29] I. Papp, C. Sieben, A. L. Sisson, J. Kostka, C. Böttcher, K. Ludwig, A. Herrmann, and R. Haag, “Inhibition of influenza virus activity by multivalent glycoarchitectures with matched sizes,” *ChemBioChem*, vol. 12, no. 6, pp. 887–895, 2011. 9
- [30] L. M. Coussens and Z. Werb, “Inflammation and cancer,” *Nature*, vol. 420, no. 6917, p. 860, 2002. 9
- [31] R. Weissleder, K. Kelly, E. Y. Sun, T. Shtatland, and L. Josephson, “Cell-specific targeting of nanoparticles by multivalent attachment of small molecules,” *Nature biotechnology*, vol. 23, no. 11, p. 1418, 2005. 9
- [32] H. Yuan, W. Jiang, C. A. von Roemeling, Y. Qie, X. Liu, Y. Chen, Y. Wang, R. E. Wharen, K. Yun, G. Bu, *et al.*, “Multivalent bi-specific nanobioconjugate engager for targeted cancer immunotherapy,” *Nature nanotechnology*, vol. 12, no. 8, p. 763, 2017. 9
- [33] M. Shan, K. E. Carlson, A. Bujotzek, A. Wellner, R. Gust, M. Weber, J. A. Katzenellenbogen, and R. Haag, “Nonsteroidal bivalent estrogen ligands: an application of the bivalent concept to the estrogen receptor,” *ACS chemical biology*, vol. 8, no. 4, pp. 707–715, 2013. 10

- [34] A. Bujotzek, *Molecular simulation of multivalent ligand-receptor systems*. PhD thesis, 2013. 10, 11, 22, 24
- [35] C. Scheibe, A. Bujotzek, J. Dervedde, M. Weber, and O. Seitz, “Dna-programmed spatial screening of carbohydrate–lectin interactions,” *Chemical Science*, vol. 2, no. 4, pp. 770–775, 2011. 10
- [36] J. F. G. James A. Kugel, *Binding and Kinetics for Molecular Biologists*, vol. 127. CSHL Press, 2007. 11, 15, 21
- [37] T. Wiseman, S. Williston, J. Brandts, and L.-N. Lin, “Rapid measurement of binding constants and heats of binding using a new titration calorimeter,” *Analytical biochemistry*, vol. 179, no. 1, pp. 131–137, 1989. 15, 16, 26, 28
- [38] L. Kaufmann, N. L. Traulsen, A. Springer, H. V. Schröder, T. Mäkelä, K. Rissanen, and C. A. Schalley, “Evaluation of multivalency as an organization principle for the efficient synthesis of doubly and triply threaded amide rotaxanes,” *Organic Chemistry Frontiers*, vol. 1, no. 5, pp. 521–531, 2014. 18
- [39] M. I. Stefan and N. Le Novere, “Cooperative binding,” *PLoS computational biology*, vol. 9, no. 6, p. e1003106, 2013. 18, 20
- [40] M. Zumbansen, “Parameter estimation of multivalent processes on itc datasets,” Master’s thesis, Freie Universität Berlin, 2019. 20, 77
- [41] J. E. Gestwicki, C. W. Cairo, D. A. Mann, R. M. Owen, and L. L. Kiessling, “Selective immobilization of multivalent ligands for surface plasmon resonance and fluorescence microscopy,” *Analytical biochemistry*, vol. 305, no. 2, pp. 149–155, 2002. 21
- [42] G. Holdgate, S. Geschwindner, A. Breeze, G. Davies, N. Colclough, D. Temesi, and L. Ward, “Biophysical methods in drug discovery from small molecule to pharmaceutical,” *Williams, M., Daviter, T. (eds): Protein-Ligand Interactions. Methods in Molecular Biology*, vol. 1008, 2013. 21
- [43] C. P. Schaller, “Enthalpy changes in reactions.” [https://chem.libretexts.org/Bookshelves/Physical\\_and\\_Theoretical\\_Chemistry\\_Textbook\\_Maps/Supplemental\\_Modules\\_\(Physical\\_and\\_Theoretical\\_Chemistry\)/Thermodynamics/Fundamentals\\_of\\_Thermodynamics/Enthalpy\\_Changes\\_in\\_Reactions](https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Thermodynamics/Fundamentals_of_Thermodynamics/Enthalpy_Changes_in_Reactions). Accessed: 2022-05-19. 24
- [44] E. Freire, O. L. Mayorga, and M. Straume, “Isothermal titration calorimetry,” *Analytical chemistry*, vol. 62, no. 18, pp. 950A–959A, 1990. 26
- [45] M. M. Pierce, C. Raman, and B. T. Nall, “Isothermal titration calorimetry of protein–protein interactions,” *Methods*, vol. 19, no. 2, pp. 213–221, 1999. 26

- [46] “Isothermal titration calorimetry as a tool for determining thermodynamic parameters of chemical reactions.” [https://brain.fuw.edu.pl/edu/images/6/68/Kalorymetria\\_-\\_ang.pdf](https://brain.fuw.edu.pl/edu/images/6/68/Kalorymetria_-_ang.pdf). Accessed: 2022-11-22. 26
- [47] C. Song, S. Zhang, and H. Huang, “Choosing a suitable method for the identification of replication origins in microbial genomes,” *Frontiers in MICROBIOLOGY*, vol. 6, 09 2015. 27
- [48] D. Burnouf, E. Ennifar, S. Guedich, B. Puffer, G. Hoffmann, G. Bec, F. Disdier, and P. Baltzinger, Mireille an dDumas, “kinitc: a new method for obtaining joint thermodynamic and kinetic data by isothermal titration calorimetry,” *Journal of the American Chemical Society*, vol. 134, no. 1, pp. 559–565, 2011. 28, 29
- [49] K. A. Vander Meulen and S. E. Butcher, “Characterization of the kinetic and thermodynamic landscape of rna folding using a novel application of isothermal titration calorimetry,” *Nucleic acids research*, vol. 40, no. 5, pp. 2140–2151, 2011. 29, 54, 55
- [50] J. Lamperti, *Stochastic processes: a survey of the mathematical theory*, vol. 23. Springer Science & Business Media, 2012. 31
- [51] W. Huisinga, *Metastability of Markovian systems: A transfer operator based approach in application to molecular dynamics*. PhD thesis, Freie Universität Berlin, 2001. 32, 33, 34, 35, 37
- [52] “Stochastik 2 skript 2b.” [https://www.uni-ulm.de/fileadmin/website\\_uni\\_ulm/mawi.inst.110/lehre/ws13/Stochastik\\_II/Skript\\_2B.pdf](https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.110/lehre/ws13/Stochastik_II/Skript_2B.pdf). Accessed: 2022-09-12. 33
- [53] C. Robert and G. Casella, *Monte Carlo statistical methods*. Springer Science & Business Media, 2013. 33, 49
- [54] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*. Springer Science & Business Media, 2012. 33, 36, 37
- [55] S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte, and F. Noé, “Data-driven model reduction and transfer operator approximation,” *Journal of Nonlinear Science*, vol. 28, no. 3, pp. 985–1010, 2018. 36, 38
- [56] K. Fackeldey, A. Sikorski, and M. Weber, “Spectral clustering for non-reversible markov chains,” *Computational and Applied Mathematics*, vol. 37, no. 5, pp. 6376–6391, 2018. 38
- [57] R. Syski, *Passage times for Markov chains*, vol. 1. Ios Press, 1992. 38
- [58] C. Åberg, K. E. Duderstadt, and A. M. van Oijen, “Stability versus exchange: a paradox in dna replication,” *Nucleic acids research*, vol. 44, no. 10, pp. 4846–4854, 2016. 39

- [59] P. Deuffhard and M. Weber, “Robust perron cluster analysis in conformation dynamics,” *Linear algebra and its applications*, vol. 398, pp. 161–184, 2005. 45
- [60] S. Röblitz and M. Weber, “Fuzzy spectral clustering by PCCA+: application to markov state models and data classification,” *Advances in Data Analysis and Classification*, vol. 7, no. 2, pp. 147–179, 2013. 45, 46, 48
- [61] K. Fackeldey and M. Weber, “GenPCCA – markov state models for non-equilibrium steady states,” *WIAS Report*, vol. 29, pp. 70–80, 2017. 49
- [62] M. Berberan-Santos and J. Martinho, “The integration of kinetic rate equations by matrix methods,” *Journal of Chemical Education*, vol. 67, no. 5, p. 375, 1990. 50
- [63] F. Erlekam, S. Ige, S. Röblitz, L. Hartmann, and M. Weber, “Modelling of multivalent ligand-receptor binding measured by kinitc,” *Computation*, 2019. 53, 54, 55, 56, 58, 59
- [64] S. Igde, S. Röblitz, A. Müller, K. Kolbe, S. Boden, C. Fessele, T. K. Lindhorst, M. Weber, and L. Hartmann, “Linear precision glycomacromolecules with varying interligand spacing and linker functionalities binding to concanavalin a and the bacterial lectin fimh,” *Macromolecular Bioscience*, vol. 17, no. 12, 2017. 53, 54, 55, 56, 57, 58, 59, 75, 78, 94, 95, 96, 97
- [65] U. Nowak and L. Weimann, “Numerical solution of nonlinear (nl) least squares (s) problems with nonlinear constraints (con), especially designed for numerically sensitive problems.” [http://elib.zib.de/pub/elib/codelib/nlscon\\_m/nlscon.m](http://elib.zib.de/pub/elib/codelib/nlscon_m/nlscon.m), 2017. 55
- [66] S. Igde, *Assessing the Influence of Different Structural Features on Multivalent Thermodynamics and Kinetics of Precision Glycomacromolecules*. PhD thesis, Freie Universität Berlin, 2016. 55
- [67] F. Erlekam, M. Zumbansen, and M. Weber, “Parameter estimation on multivalent itc data sets,” *Scientific Reports*, 2022. 75
- [68] T. Wiseman, S. H. Williston, J. F. Brandts, and L. N. Lin, “Rapid measurement of binding constants and heats of binding using a new titration calorimeter.,” *Analytical biochemistry*, vol. 179 1, pp. 131–7, 1989. 76
- [69] W. Swope, J. Pitner, F. Suits, M. Pitman, and M. Eleftheriou, “Describing protein folding kinetics by molecular dynamics simulations: 2. Example applications to alanine dipeptide and beta-hairpin peptide,” *Journal of Physical Chemistry B*, vol. 108, pp. 6582–6594, 2004. 83

- [70] S. Röhl, M. Weber, and K. Fackeldey, “Computing the minimal rebinding effect for nonreversible processes,” *Multiscale Modeling & Simulation*, vol. 19, no. 1, pp. 460–477, 2021. 87, 89, 93, 94
- [71] G. Vauquelin, “Impact of target binding kinetics on in vivo drug efficacy: koff , kon and rebinding,” *British journal of pharmacology*, vol. 173, 04 2016. 88
- [72] R. Zhang, “Pharmacodynamics: Which trails are your drugs taking?,” *Nature chemical biology*, vol. 11, no. 6, p. 382, 2015. 88
- [73] L. Donati, M. Heida, B. G. Keller, and M. Weber, “Estimation of the infinitesimal generator by square-root approximation,” *Journal of Physics: Condensed Matter*, vol. 30, no. 42, p. 425201, 2018. 105
- [74] H. C. Lie, K. Fackeldey, and M. Weber, “A square root approximation of transition rates for a markov state model,” *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 2, pp. 738–756, 2013. 105
- [75] M. Heida, “Convergences of the squareroot approximation scheme to the fokker–planck operator,” *Mathematical Models and Methods in Applied Sciences*, vol. 28, no. 13, pp. 2599–2635, 2018. 105

# Declaration of authorship

Name: Erlekam

First name: Franziska

I hereby declare that I have completed the submitted dissertation independently and without the use of sources and aids other than those indicated. I have marked as such all statements that are taken literally or in content from other writings. This dissertation has not yet been presented to any other examination authority in the same or a similar form and has not yet been published.

Berlin, 08-14-2023