

BAYESIAN INFERENCE FOR TOP-DOWN PROTEIN MARKOV MODELS AND THE UBIQUITY OF MICHAELIS-MENTEN KINETICS

Dissertation

for the award of the degree

'Doctor rerum naturalium'

at the Georg-August-Universität Göttingen

within the doctoral program

'Physics of Biological and Complex Systems'

of the Georg-August University School of Science (GAUSS)

submitted by

Malte Norman Schöffner

from Berlin-Schöneberg

Göttingen, 2023

Thesis advisory committee

Prof. Dr. Helmut Grubmüller

Theoretical and Computational Biophysics, Max Planck Institute for Multidisciplinary Sciences

Prof. Dr. Stefan Klumpp

Institute for the Dynamics of Complex Systems, Georg-August-Universität Göttingen

Dr. Johannes Söding

Quantitative and Computational Biology, Max Planck Institute for Multidisciplinary Sciences

Members of the examination board:

Referee:

Prof. Dr. Helmut Grubmüller

Theoretical and Computational Biophysics, Max Planck Institute for Multidisciplinary Sciences

Co-referee:

Prof. Dr. Stefan Klumpp

Institute for the Dynamics of Complex Systems, Georg-August-Universität Göttingen

Other members of the Examination Board:

Dr. Johannes Söding

Quantitative and Computational Biology, Max Planck Institute for Multidisciplinary Sciences

Prof. Dr. Timo Betz

Third Institute of Physics - Biophysics, Georg-August-Universität Göttingen

Prof. Dr. Andreas Janshoff

Institute for Physical Chemistry, Georg-August-Universität Göttingen

Prof. Dr. Marcus Müller

Institute for Theoretical Physics, Georg-August-Universität Göttingen

Date of the oral examination: November 6, 2023

Abstract

The current understanding of protein function predominantly relies on the concept of describing proteins as comprising discrete conformational and chemical states including transitions between these states. The theory of Markov processes provides the mathematical foundation for such a description in the form of Markov models. Two distinct methodologies exist for the construction of such models. In a bottom-up approach, Markov model parameters are directly determined based on molecular dynamics simulations, which typically constrains the Markov models to short timescale dynamics. In contrast, in a top-down approach, the parameters are indirectly determined by fitting observables derived from the Markov model to experimental data. This second distinctive methodology permits the integration of long timescale dynamics. However, due to insufficient experimental data, many top-down Markov models may exist that fit the experimental data equally well.

In the first thesis part, I employed Bayesian inference to address this underdetermination through a rigorous ranking of shared characteristics, such as molecular mechanisms, of well-fitting top-down Markov models in terms of posterior probability. I apply this approach to twin-ATPase ABCE1, which exhibits an unexpected kinetic asymmetry between its two ATP-binding sites that was observed in mutants. Thus far, the prevailing hypothesis to explain this asymmetry is a direct allosteric communication between the two binding sites. However, I challenge this hypothesis with successfully fitting top-down Markov models that are explicitly defined to exclude such direct allosteric communication. Instead, I demonstrate that a shift in steady-state population between reaction pathways, caused by avoiding kinetic trap states, provides an alternative explanation for the asymmetry.

During this first part, the dependence of ATP turnover rate on a logarithmic substrate concentration approximated—at least piecewise—a sigmoid curve for the majority of ABCE1 Markov models. This dependence is also observed in one of the simplest enzyme kinetics models, namely the Michaelis-Menten kinetics. Despite the perceived ubiquity of the Michaelis-Menten kinetics, the prevalence of this dependence was an unexpected outcome, given that the Markov models of ABCE1 are considerably more intricate than the Markov model that is necessary to derive the Michaelis-Menten kinetics.

In the second thesis part, I investigated this discrepancy in expectation by quan-

tifying the occurrence of Michaelis-Menten kinetics in sparsely connected Markov models. Strikingly, I showed that even for large models with up to 200 Markov states, the turnover curves are almost exclusively ($> 80\%$) comprised of such Michaelis-Menten-like pieces, as long as a fraction of concentration-dependent rates is less than 10% . As this low fraction covers the biologically relevant range, this result contributes to an explanation of the ubiquity of Michaelis-Menten kinetics. I explain this behavior based on the distance between possible Michaelis-Menten-like pieces and the combinatorial problem described by the hypergeometric distribution, which determines this distance.

Both approaches—the employment of top-down Markov models to investigate allostery and the comprehensive analysis of the intrinsic characteristics of Markov models to elucidate common motifs in proteins—have the potential for broader applications and, thus, will contribute to a more profound understanding of proteins.

Contents

1	Introduction	7
2	The Kinetic Asymmetry of ABCE1 Mutants	15
2.1	Introduction	16
2.2	Methods	21
2.3	Results	25
2.4	Conclusion	39
2.5	Supplementary Information	41
3	Unexpected Michaelis-Menten behaviour of complex Markov models	43
3.1	Introduction	44
3.2	Theory	47
3.3	Results	61
3.4	Conclusion	64
3.5	Methods	67
3.6	Supplementary Information	70
4	Conclusion	73

1 Introduction

Proteins are macromolecules that perform a variety of functions in all living organisms, for example, catalyzing reactions, transferring molecules across membranes, and assembling themselves from DNA. Each protein is a polymer with a specific sequence of amino acids, the primary sequence, with some sections of these sequences being conserved over millions of years and across species. To perform their functions, proteins fold under physiological conditions into, also protein specific, structures driven by electrostatic interactions, hydrogen bonds, salt bridges and hydrophobicity.¹

Historically, the first protein structures were determined by X-ray diffraction experiments on protein crystals of hemoglobin² and myoglobin.³ The very fact that proteins could be crystallized and their structure determined by X-ray diffraction lead to the notion that each protein has one specific structure and that its function can be explained by this particular structure.⁴

Only in the late 1970s did Frauenfelder et al. observe an unusual distribution of relaxation times for the release of carbon monoxide from myoglobin in flash photolysis experiments that can only be explained by a release process that includes multiple conformational states of myoglobin.⁵ They concluded that proteins driven by thermal fluctuations are dynamic entities adopting a magnitude of sub-states of their structures between which the proteins transition during the time span of performing their function. Today, these sub-states are referred to as conformations. The transitions between the conformations range from minor rearrangements of amino acids side chains to large displacements of whole sections of the protein structure compared to the measured structures. The transition rates between sub-states span a range from 10^{-11} s to 1 s, where rates typically correlate with the size of the conformational change.⁶

Further experiments of myoglobin suggested that the sub-states are organized in a hierarchical manner in the free energy landscape, i.e, every sub-state contains sub-sub-states separated by barriers lower than the barriers separated the sub-

states.⁷ Thus, different levels of conformations exist, where the conformations of a lower level might be described as thermal fluctuations around an average structure of a conformation of a higher level.

The conformations of the highest levels are typically quite different from each other and represent different functional modes of the protein which are controlled by external factors, e.g., binding of ligands or other proteins, or substrate and salt concentrations. For example, ion channels might have a conducting and non-conducting conformation depending on the presence of ligands⁸ whereas enzymes an active and inactive conformation depending on the presence of substrates.⁹ The structures of conformations that depend on external factors can often be determined by X-ray diffraction, nuclear magnetic resonance spectroscopy or cryo-electron microscopy of proteins exposed to these external factors. However, the transitions between these conformations and the abundance of conformations of lower levels with smaller deviations between them, which are still functionally relevant as the experiments of Frauenfelder et al. have shown, are extremely challenging to determine experimentally due to the high spatial and temporal resolution required.¹⁰

Computational methods, in particular molecular dynamics simulations, have been developed to make conformations of very similar structure, and especially the transitions between them, accessible and have provided astonishing insights into the function of proteins.¹¹ Molecular dynamics simulations describe proteins on an atomistic level and use empirical potential energy functions to model bonded and non-bonded interactions between atoms, where the set of empirical potential energy functions is referred to as a force field. Time trajectories of protein conformations are then generated by integrating the equations of motions for each atom in the force field.¹² Due to constant improvement of simulation software and the ever increasing available computing power, today simulations of large systems like the ribosome with millions of atoms¹³ and milliseconds of systems with ten-thousands of atoms are possible today.¹⁴

Molecular dynamics simulations generate high-dimensional data due to their atomistic resolution, which raises the challenge of isolating the degrees of freedom that are relevant to gain insight into a particular question. Thus, coarse-grain approaches have been developed for the analysis of molecular dynamics trajectories that remove irrelevant degrees of freedom from the data. Markov models, which use the name giving Markov chains to describe molecular dynamics, have emerged

as a theoretically sound way of coarse-graining¹⁵ and have been successfully applied to many protein systems.¹⁶⁻¹⁸

Markov chains comprise of a set of states, here protein conformations, and a set of transitions with transition rates between these states where each transition obeys the Markov property of memorylessness, i.e., the probability of a transition depends only on the current state.¹⁹ A time-continuous Markov chain is defined by its transition rate matrix, which gives the probabilities of transition between states per second.

Considering the hierarchical free energy landscape and using the terminology of Frauenfelder et al. for conformations of states and substates, it can be seen that Markov models provide a very intuitive approach to coarse-grain molecular dynamics trajectories to varying degrees by associating Markov states with states of different levels. The transition rates can then be inferred from the observed transitions between conformations in the molecular dynamics trajectories.²⁰ Thus, the construction of Markov models from molecular dynamics trajectories can be considered a data-driven bottom-up approach.

Moreover, Markov models can be thermodynamically consistent, i.e., they can satisfy detailed balance, thus providing access to the equilibrium thermodynamic and kinetic properties of the system.²¹ In particular, because transition rates can be estimated from trajectories that are only locally equilibrated in the initial state and not globally, Markov models allow access to time scales beyond those accessible in molecular dynamics simulations.¹⁸

However, despite remarkable progress in the field of molecular dynamics simulations, for many proteins not every functionally relevant transition is accessible by this method. In particular, the time scales of many protein-protein interactions, substrate or ligand binding, and protein folding are typically too large to observe even a single transition in molecular dynamics simulations. Furthermore, molecular dynamics simulations typically do not model breaking or forming bonds and cannot describe chemical reactions. Because bottom-up Markov models are inferred from molecular dynamics data, they have the same limitations. Thus, molecular dynamics simulations and Markov models derived from them are typically limited to describing conformational changes and cannot provide a complete picture of all functionally relevant states and transitions of a protein.

However, Markov models do not need to be inferred from molecular dynamic simulations and are, in general, quite suited to describe ligand exchanges and

chemical reactions. Note that for the sake of shortness, we will refer to ligand exchanges as chemical reactions too, although no covalent bounds are formed or broken. For example, Markov models were already used in the beginning of the 1900s century to describe enzyme kinetics, although they were not referred to by this term.^{22,23} Michaelis and Menten described the enzyme kinetics as a chemical reaction with the enzyme as a catalyst, and thus established that enzyme kinetics can be described like every other chemical reaction as a set of species, compounds and reactions. Setup and analysis of these chemical reaction networks were formalized in the 70s by Feinberg, Horn and Jackson.^{24,25}

There are two special cases in which chemical reaction networks can be described by a Markov model. First, if chemical reaction networks comprise only of monomolecular chemical reactions with mass action kinetics. Note that this case includes bottom-up Markov models if each conformational change is interpreted as a monomolecular reaction. Second, if chemical reaction networks describe systems that are in a quasi-steady state, all non-enzyme complexes can be assumed constant, and thus all multimolecular reactions effectively reduce to monomolecular reactions.²⁶ This assumption has become a cornerstone in the description of proteins,²⁷ where only recently the scope of its validity has been investigated.²⁸⁻³⁰

Also only recently it is widely recognized and accepted that both chemical reactions and conformational changes can be described by Markov models, allowing one to combine both within a unified theory for complex biomolecular interaction networks of 'molecular machines'. Recent examples include the use of Markov models to analyze translation,^{31,32} transcription,^{33,34} signaling pathways,³⁵ molecular motors,³⁶⁻⁴³ and complex enzymes such as the fatty acid synthesis,⁴⁴ allowing a better conceptual understanding of their functions.[†]

We will refer to this type of Markov models as top-down Markov models, because the Markov states and the connections are deliberately chosen so that the model includes all functionally relevant conformational changes and chemical reactions, as opposed to being inferred from a large data set. In addition, the determination of transition rates differs between the bottom-up and top-down Markov models. In the bottom-up Markov models all transition rates are inferred from molecular dynamics data as are the states and connections. In contrast, the top-down Markov models take advantage of the fact that Markov models are quantitative models and thus the calculation of quantities that are accessible in experiments,

[†]This paragraph is adapted from the introduction of chapter 3.

for example, turnover rates and occupations is possible. Thus, in the top-down Markov models the transition rates are inferred from experimental data by solving the backward problem of determining the Markov models that produce the measured values. In particular for increasingly complex top-down Markov models, this backward problem is underdetermined, i.e., the number of unknowns, which is equal to the number of transition rates of the Markov model, is drastically larger than the number of experimental observations.⁴⁴

To date, maximum likelihood estimation or optimization of cost functions have been used to determine Markov models that produce the experimental data used to constrain the models.^{35,44,45} These approaches have the following drawbacks: First, because of the focus on determining Markov models that produce the measured values as closely as possible, the sampled Markov models can be overfitted, i.e., they produce the measured values very well but are not able to correctly predict other experimental data. Second, in the case of an underdetermined system, many Markov models with similar high likelihood but otherwise different properties can be determined. Neither approach provides a basis for favoring or disfavoring one of these Markov models over another. Third, they do not provide a rigorous uncertainty estimate, for example, of the transition rates.⁴⁶

Here we want to address the underdetermination of top-down Markov models from a new angle with a Bayesian approach, which assigns each Markov model the probability density value that it is the one that produced the given data in the first place. Bayesian approaches use the name giving Bayes theorem, which states (ignoring the normalization by the probability of the data) that the probability of a model given the data, referred to as the posterior probability, is proportional to the probability of the data given the model multiplied by the prior probability of the model.

Bayesian approaches require the choice of a prior, which is unique to each problem and reflects the knowledge about the model without the current data. In this paper, we ask what is a good prior for transition rates in top-down Markov models.

Bayesian approaches promise to improve on the drawbacks of the above-mentioned approaches by determining the full posterior distribution rather than particular points of a likelihood or cost function. First, because the posterior probability includes experimental errors, not only are Markov models determined that produce values as close to the experimental mean as possible, but also Markov models

that match the measured values within the experimental error, thus mitigating the problem of overfitting. Second, the posterior distribution allows to calculate probabilities for subsets of Markov models that, for example, have common properties with respect to the answer to a research question. Thus, even in the case of an underdetermined problem, it is possible to make valid inferences about different subsets of Markov models, beyond the fact that individual members of the subsets explain the experimental values equally well. Third, by determining the full posterior distribution, uncertainties are naturally provided.

Together with Tampe et al. we noticed that for the ATPase ABCE1 top-down Markov models could provide insights into the unexpected asymmetry between the kinetics of two ABCE1 mutants.⁴⁷ ABCE1 has two nucleotide binding sites that are very similar in sequence and structure. Thus, an equal contribution of both binding sites to the ATP turnover rate of wild-type ABCE1 was expected. This hypothesis was apparently confirmed by measuring a twofold decrease in the ATP turnover rate of a mutant in which ATP hydrolysis was impaired in one of the two binding sites. However, the identical point mutation in the opposite binding site resulted in a tenfold increase in ATP turnover rate.

An asymmetric allostery between the two binding sites has been proposed to explain the asymmetric kinetics of the two mutants. The test of this hypothesis using top-down Markov models is presented in Chapter 2.

The top-down Markov models of ABCE1 inferred in chapter 2 are drastically more complex in terms of number of states and connections if compared to the probably first top-down Markov model of the Michaelis-Menten kinetics with only two states. However, despite this difference in complexity, we observed that the Markov models of ABCE1 unexpectedly often show a behavior that is very well described by these same Michaelis-Menten kinetics.

As mentioned above, the description of complex systems with multimolecular transitions by Markov models relies on the assumption that these systems operate in a quasi-steady state. Under this assumption all substrate, ligand, and product concentrations are assumed to be constant and steady state observables, for example, turnover rates or occupations, can be calculated from the steady state probability of each state. Importantly, the steady state probability depends on the value of the as constant assumed concentrations, i.e., the steady state probability changes if a different constant concentration is assumed. We refer to the function

of a steady state probability on a non-enzyme concentration as an occupation curve.

One example of an occupation curve is the Michaelis-Menten equation of the above-mentioned Michaelis-Menten kinetics. The equation describes how the turnover rate, which is a function of a steady state probability, of an enzyme changes with the substrate concentration if the enzyme kinetics are described by a two-state Markov model.

We noticed that the occupation curves of the Markov models of ABCE1 are unexpectedly often, at least on intervals of the substrate concentration, described by this exact Michaelis-Menten equation, i.e. are sigmoidal on a logarithmic concentration scale. In addition, and quite strikingly, we were not the first to observe Michaelis-Menten-like behavior in occupations curves of complex Markov models.^{48–50}

Consequently, already some research has been done on the cause of these unexpected occurrences of Michaelis-Menten-like behavior in complex Markov models. In particular, the case of an occupation curve to be described by one Michaelis-Menten equation over the entire concentration interval has already been investigated. Two studies derive conditions for this case based on the idea to coarse-grain the Markov models to match the two-state Markov model of the Michaelis-Menten kinetics. One is based on a distinction of states with respect to representing free enzymes or bound enzymes,⁴⁹ and one is more generally based on the structure of Markov models, i.e., the states and the connections between them.⁵⁰ The two works derive sufficient but not necessary conditions, and thus, there are still occurrences of this case of Michaelis-Menten-like behavior that remain unexplained⁵¹ including our own observations.

In particular, occurrences of occupations curves that are piecewise well described by the Michaelis-Menten equation as observed for proteins with multiple binding sites still remain unexplored.^{52–54} Thus, in Chapter 3 we asked how probable it is for occupations curves of Markov models to have Michaelis-Menten-like pieces, especially for Markov models with properties typical for the description of proteins.

2 The Kinetic Asymmetry of ABCE1 Mutants

2.1 Introduction

Protein biosynthesis is critical to all life and involves two major steps: transcription and translation. The latter is carried out by the ribosome, a macromolecular machine present in all organisms, together with species-specific translation factors. The ribosome consists of two subunits, the large and the small subunit, which must assemble anew for each synthesized protein and, at the end of translation, must be separated by translation factors to allow the translation cycle to start again.

In all eukaryotes and archaea this vital task of separating the ribosomal subunits during the termination step is fulfilled by the ATPase ABCE1, a member of the ATP-binding cassette superfamily, in coordination with further translation factors. The separation process is driven by binding two ATPs and a subsequent conformational change of ABCE1. Additionally, ABCE1 is involved in recruiting translation initiation factors to the small ribosomal subunit to restart translation.^{55,56}

ABCE1 has two sequentially and structurally almost symmetrical nucleotide binding domains (NBD).^{47,57-63} Figure 2.1 shows the structure of ABCE1 including the seven highly conserved sequence motifs in each NBD highlighted, which are required for ATP-binding and hydrolysis, namely, A-loop (Y-loop), Walker-A (P-loop), Q-loop, His-switch (H-loop), Walker-B, D-loop and the ABC-signature motif (C-loop).

The conserved loops are distributed on two structurally separated lobes including A-loop, Walker-A, Q-loop, His-switch and Walker-B in one lobe, the signature motif in the other lobe, and the D-loop between the lobes. Because of the spacial separation of the conserved loops, a single NBD is incapable of nucleotide hydrolysis.^{64,65} To allow hydrolysis, the two NBDs are oriented in an anti-parallel fashion creating two functional nucleotide binding sites (NBS) at their interface. The binding sites are referred to as NBSI and NBSII based on the order of appearance of the corresponding Walker-A motif in the primary sequence.

X-ray diffraction and cryo-electron microscopy structures show ABCE1 in an open^{47,57,58} and a closed conformation,^{62,63,66} and suggest a third intermediate conformation if ABCE1 is in pre-splitting complex with the ribosome.^{59,60} In the closed state, one nucleotide per NBSs is bound between the two NBDs.

Structural insights to ABCE1 are complemented by bulk and single-molecule ex-

periments.^{47,55,56} To investigate the functional role of each NBS, the ATP turnover rate of wild-type ABCE1 and ATP hydrolysis deficient mutants of ABCE1 were measured amongst other observables, for example, the Michaelis-Menten constant K_M .⁴⁷ In each mutant, the catalytic active Walker-B motif glutamate (one letter code E) of one NBS was changed to glutamine (one letter code Q), whereas the other was left intact, resulting in a drastically reduced ATP hydrolysis rate of the respective NBS. Mutant E238Q has the modification in the first NBS (NBSI) and mutant E485Q in the second NBS (NBSII) allowing to measure the ATP turnover rate of the opposite NBS.

The high symmetry between NBSI and NBSII suggests that each binding site contributes equally to the wild-type ATP turnover rate. This expectation is fulfilled by mutant E238Q that has a twofold reduced ATP turnover rate compared to the wild-type. However, and quite astonishing, mutant E485Q does not show a lower ATP turnover rate, but instead even an tenfold increase of ATP turnover rate, thereby having a higher ATP turnover rate than the wild-type with half the active NBSs.^{47,55} This rate increase is an unprecedented behavior for a member of the ABC-superfamily, for example, P-glycoprotein (MDR3) with two consensus NBSs shows a symmetric decrease of ATP turnover rate for mutants with identical EQ point mutation in each NBS.⁶⁷

Although there are sequential and structural differences between the NBSs of ABCE1, none strikes as the sole reason for this asymmetry in ATP hydrolysis rates as they seem either too minor or were experimentally tested and rejected to be responsible for the observed asymmetry.⁶³

The asymmetric kinetics lead to the common assumption^{55,56,63} of an asymmetric allostery between both NBSs such that ATP-binding in NBSII facilitates ATP-binding or hydrolysis, or both, in NBSI. In this scenario, the asymmetric kinetics of the mutants can be explained by an increased occupation of hydrolysis-deficient NBSII by ATP and, thus, by an on average longer 'activated' NBSI.⁶³

However, the mechanism of this allostery in ABCE1 remains elusive. ABC transporters, which are homologues of ABCE1, have various interactions across both NBDs and, in particular, interactions of the D-loop of one NBS with the Walker-A motif and His-switch of the opposite binding site.^{68,69} In contrast to ABCE1, these transporters have one consensus and one degenerate side, i.e., one NBS deviates from the consensus Walker-B motif, His-Switch and ABC-

signature motif drastically reducing its ATP hydrolysis capabilities.⁶⁵ Molecular dynamic simulations of ABC transporter Sav186627 with two consensus NBSs, shows that the interaction of the D-loop across NBSs changes upon switching from ADP/ATP- to ATP/apo-occupations.⁷⁰

As the most direct experimental indicator of direct interactions between the NBSs in members of the ABC-superfamily is the measurement of cooperativity in Rad50, another homologue of ABCE1 involved in DNA double-strand break repair. After changing either the D-loop aspartate and the Walker-A asparagine of the opposite binding site, which are thought to be responsible for allostery between the NBSs, a reduced ATP-hydrolysis activity and reduced cooperativity was determined, pointing also towards the role of the D-loop in the communication between the NBSs.⁷¹ However, neither the wild-type nor the mutants of ABCE1, and in particular not the wild-type, show any cooperativity to start with.⁴⁷ Still, the current prevalent hypothesis for the mechanism of allostery in ABCE1 is a direct interaction between the two NBSs by conformational changes.^{55,56,63}

In this chapter of this work, we ask if an asymmetric allostery between the two NBSs is actually necessary in the first place, or if explanations not relying on such allostery may exist for this unexpected kinetic asymmetry. In particular, if ABCE1 without any allostery between the NBSs is not already complex enough to exhibit this kind of unexpected kinetic asymmetry.

To this end, we describe ABCE1 as a Markov model, i.e., as a set of discrete states, where the states are associated with different conformational and occupational states of ABCE1, and with memory-free transitions of conformational changes, ligand exchange and ATP catalysis between these states.¹⁹ Further, we design the Markov model of ABCE1 such that any symmetric allostery between the NBSs is excluded by having identical transition rate coefficients for binding and catalysis transitions for each NBS irrespective of the opposite NBS occupation state.

Note that we, at first, only define the 'structure' of the Markov model, i.e., its states and its connections to which we refer to as a Markov model class and not the transition rate coefficients of the Markov model. In fact, there are many transition rate coefficient matrices, referred to as instances, within this Markov model class and we aim to find the subset of this instances that agrees with the experimentally determined kinetics. We expect that a large fraction of the members of this set shares common properties that serve to explain the asymmetry.

To this end, we use Bayesian inference to determine this subset and to assign probabilities to the set that share common properties, allowing for a rigorous ranking of possible explanations. Previous approaches have used maximum likelihood approaches or optimization of a cost function to identify individual instances.^{35,36,45}

We infer instances that produce the striking asymmetry with the Michaelis-Menten kinetics parameters of wild type and both mutants and, thereby, show evidence that an direct interaction of the NBSs is not required to explain the asymmetric ATP turnover rates.

Instead, the asymmetric kinetics of the different mutants are explained as the result of a mutation-induced redistribution of the steady state probability. A preference of ATP hydrolysis in NBSI and of a conformational switch from the closed to the open state with ADP/ATP occupation over ATP hydrolysis in NBSII facilitates larger steady state probabilities of states belonging to reaction pathways with high ATP turnover rates over NBSI than of states belonging to reaction pathways with lower ATP turnover rate over NBSII. Thus, the reduction of the ATP-hydrolysis rate of NBSI in mutant E238Q results only in a minor change of the steady state probability. However, the same reduction of ATP hydrolysis in NBSII results in a major shift of steady state probability towards states belonging to high ATP turnover reaction pathways, allowing for the tenfold increased ATP-hydrolysis rate of mutant E485Q.

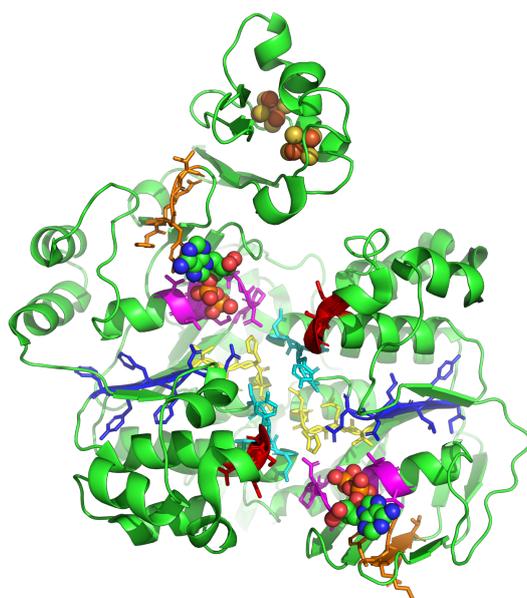


Figure 2.1 – Ribbon representation of the X-ray structure of ABCE1 in its open conformation with two bound ADPs adopted from the PDB file 3BK7.⁵⁸ The color-coded loops in stick representation are A-loop (orange), Walker-A (magenta), His-switch (yellow), Walker-B (blue), D-loop (cyan) and ABC-signature motif (red).

2.2 Methods

Markov model kinetics and occupations The probability of the model to be in a state i at a given time t is given by $X_i(t)$. Given an instance, i.e., a transition rate coefficient matrix \mathbf{Q} , of in the previous section introduced Markov model class and, given an initial distribution $\mathbf{X}(0)$, the time evolution of the probabilities $\mathbf{X}(t)$ is given by the master equation

$$\dot{\mathbf{X}} = \mathbf{Q} \cdot \mathbf{X}(t), \quad (2.1)$$

with transition rate coefficient matrix \mathbf{Q} in which $q_{tt} = -\sum_{s=1, s \neq t}^N k_{ts}$ are the diagonal elements.⁷² Because ABCE1 is assumed to operate in a quasi-steady-state during measurement of kinetics and occupations, except occupations measurement by Nürenberg-Goloub et al., it suffices to determine the steady state solution $\boldsymbol{\pi}$ of the master equation 2.1.^{47,55}

Note that \mathbf{Q} has educt and product concentration-independent entries k_{ts} and concentration-dependent entries $k'_{ts} = \frac{c}{c_0} k_{ts} = \alpha k_{ts}$ and that, thus, for each educt and product concentration, the steady state solution had to be calculated anew.

From the steady state solution, the ATP turnover rate of an instance was calculated as the sum of the net flux of all ADP-unbinding transitions with the net flux being the difference $k_{ts}\pi_s - k_{st}\pi_t$ such that the net flux is positive if ATP is hydrolyzed. To calculate the parameters of the Michaelis-Menten kinetics, the ATP turnover rate was calculated for each order of magnitude of ATP concentration between 10^{-9} to 10^2 M. Instead of fitting the Michaelis-Menten equation, the parameters were determined as follows to save computing time: The saturating ATP turnover rate $k_{v\max}$ was determined as the ATP turnover rate at a concentration of 100 M ATP. To determine the Michaelis-Menten constant, the two closest points to $\frac{k_{v\max}}{2}$ were determined and for 10 additional points, logarithmic uniformly distributed, ATP concentrations between these two values, further ATP turnover rates were calculated. The Michaelis-Menten constant was finally calculated via linear interpolation between the two new closest points to $\frac{k_{v\max}}{2}$ as the ATP concentration for which the linear interpolation has value $\frac{k_{v\max}}{2}$.

All transition rate coefficients were scaled to match the temperature, and concentration-dependent transition rate coefficients k'_{ts} were adjusted to match ATP and ADP concentrations of the respective experimental setups. In absence of an

experimental control of the ADP concentrations, the ADP concentrations were estimated for each experiment that accumulates during the experiment due to auto-hydrolysis and enzyme-based catalysis of ATP by ABCE1. All experimental setups excluded explicit addition of ADP, so that $c_{ADP}(t = 0) = 0$. In case of the ATPase assay by Barthelme et al., the ADP concentration were assumed to be $1 \mu\text{M}$ accounting for multiple ATPs per ABCE1 (0 to $10 \mu\text{M}$) being hydrolyzed. For the occupation measurement of Barthelme et al., ADP concentration were assumed to be 10^{-3} of the ATP concentration.

Occupations were directly calculated from steady state populations by multiplication of a state population with its ATP/ADP occupation number [0, 1, 2].

Bayesian inference For a given \mathbf{Q} , the steady state solution and, thus, kinetics and occupations can be calculated. To test our hypothesis that no direct interaction between the NBSs is required for the asymmetric kinetics of the mutants, we had to solve the backward problem, i.e., given the asymmetric kinetics of two mutants determine a \mathbf{Q} , defining a Markov model, having those kinetics. To this aim, we applied a Bayesian approach which additionally provides the probability for each \mathbf{Q} that this \mathbf{Q} was the one responsible for generating the experimental data in the first place. Note that the problem is underdetermined, because the Markov model class has 17 unknowns but only six kinetic values have been measured (ATP turnover rate $k_{v\max}$ and Michaelis-Menten-constant K_M for each wild-type and the two mutants).

We sample the posterior distribution

$$P(\mathbf{Q}|\mathbf{d}) = \prod_{j=1}^6 \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(d_j - d_j(\mathbf{Q}))^2}{2\sigma_j^2}\right) \times P(\mathbf{Q}) \quad (2.2)$$

of transition rates matrices \mathbf{Q} limited to those following the above defined structure with d_j and σ_j being mean and standard deviation of $k_{v\max}$ and K_M for each of wild-type, mutant E238Q with impaired NBSI and mutant E485Q with impaired NBSII, and $P(\mathbf{Q})$ the prior for which a log-uniform distribution on the rates, i.e., a uniform distribution of the log of the rates, was chosen. Further, we chose different boundaries for hydrolysis/synthesis, conformational transitions and un-/binding rates as those different kind of transitions have different ranges of typical timescales. Table 2.2 lists the boundaries for each type of transition. Thus, we will use the same label as Barthelme et al., E238Q and E485Q, to refer

to the mutants with the mutations in NBSI and NBSII, respectively.

We used the experimental results from Barthelme et al., who used mutants where the Walker-B glutamate was changed to glutamine.⁴⁷ For k_{vmax} , a second set of measurements is available showing an one-third increased k_{vmax} with mutants where the Walker-B glutamate was changed to alanine instead of glutamine.⁵⁵ It is uncertain whether the difference in k_{vmax} can be attributed to the different mutations. For ABC-transporter BmrA, no difference in ATP activity was measured between the two mutations,⁷³ and for non-ABC-superfamily member replication factor C, the opposite effect on the ATP activity was measured.⁷⁴ However, the ratio of the transition rates between the wild-type, mutant E238Q, and mutant E485Q agrees well between in the two experimental setup with different mutants, and the absolute ATP hydrolysis rates of our model can be adjusted to comply with both experimental data sets by scaling all transition rate coefficients by a constant factor.

Only Markov models that satisfy detailed balance when there is no chemical potential driving hydrolysis or synthesis, i.e, concentrations of ATP, ADP, and P_i are the respective equilibrium concentrations and the ATP dissociation energy is zero, can be considered to be thermodynamically correct. The condition of detailed balance imposes constraints on the transition rate coefficients⁷⁵ and the degrees of freedom of the Markov model class are reduced from 20 to 17.

Bayesian inference was performed using the Metropolis Monte Carlo Markov chain algorithm^{76,77} with 500 chains and 14×10^6 each. The first 1×10^6 samples were discarded to equilibrate the chains. The starting points of each chain were determined by minimization of the posterior using a differential evolution algorithm.⁷⁸

To avoid numerical errors despite the transition rate coefficients in \mathcal{Q} spanning up to 19 orders of magnitude, the steady state solution was calculated with 106-bit significant precision and all other calculations were done with double precision.

To evaluate the convergence of the Bayesian inference, the first and the second half of all chains were compared as depicted in supplementary figure 2.9. The posterior densities show convergence in most regions; however, differences between both halves suggest that the chain lengths are still insufficient or the chains have to be equilibrated longer. The discrepancies agree with the \hat{R} -value of 1.55, which measures the mixing of the chains, being larger than the recommended value of 1.05.⁷⁹ Nevertheless, we considered the convergence to be sufficient for qualita-

tive statements, whereas quantitative statements, in particular the percentage of reaction pathways, might change with more sampling of the posterior.

2.3 Results

To test the hypothesis that no direct communication between the two NBSs of ABCE1 is required to explain the asymmetric kinetics of the two mutant species, we describe the combined conformational dynamics and ATP binding, hydrolysis and product unbinding as a Markov model, which by definition does not contain such direct interaction. Therefore, if the observed peculiar and counter-intuitive kinetics can be reproduced by the Markov model, there is not necessarily any allostery involved.

Markov Model classes of ABCE1 The following assumptions determine the number of states, their association with conformational and occupational states of ABCE1, and the connections between them as shown in figure 2.2.

First, we assume that the order of binding or unbinding of ADP and P_i is irrelevant to the overall kinetics because it is determined by the rate-limiting binding or unbinding of these two ligands, whichever it is, and thus ADP and P_i binding and unbinding can be described by a combined transition with an effective transition rate coefficient. This assumption is consistent with the current model of ABCE1-driven ribosome separation, that combines ADP and P_i release in a single step.⁶³ Although cryo-EM structures for the heterodimeric ABC exporter TmrAB suggest that P_i release occurs before ADP and facilitates further opening of NBSs, no further functional relevance for P_i release was found beyond triggering NBS opening and subsequent ADP release,⁹ and thus further supports the description of ADP and P_i release as a combined transition. Note that binding and unbinding of ADP henceforth implies the binding and unbinding of ADP and P_i .

Second, we assume that two conformational states of ABCE1 are sufficient to describe ATP-hydrolysis of free ABCE1. An open state in which ATP and ADP binding are possible but hydrolysis and synthesis are not, and a closed state in which ATP and ADP binding and unbinding are impossible but hydrolysis and synthesis are possible. The existence of an open and a closed state has been established by X-ray structures of free ABCE1 occupied by ADP showing ABCE1 in an open conformation^{47,57,58} and Cryo-EM structures of ABCE1 bound to the small ribosomal subunit showing ABCE1 in a closed conformation.^{62,63,66}

Furthermore, it is well supported that in the open state primarily binding/unbinding of ATP and ADP and only background amounts of hydrolysis or synthesis take place because the signature motif and the D-loop of the opposite NBD,

which complement the NBS and coordinate the γ -phosphate, both necessary for hydrolysis, are not in contact with ATP.^{47,57,58,69,80}

By the same argument the closed state is considered to be competent for ATP hydrolysis and synthesis because all residues required for ATP hydrolysis are in contact with ATP.^{55,62,63,66,81} Additionally, when ABCE1 is arrested in the closed state by AMP-PNP or by rapid cooling of thermophilic variants of ABCE1, the bound nucleotides can not be washed out supporting that binding and unbinding of ADP and ATP are not possible in the closed state.^{47,55}

In addition to the open and closed states, Cryo-EM structures of ABCE1 bound to the ribosome show ABCE1 in an additional, intermediate or semi-open conformation.^{59,60} Förster resonance energy transfer (FRET) measurements of the distance between the NBDs of each NBS of ABCE1 support the meta-stability of this intermediate state.^{56,63} Whereas measurements of free ABCE1 in absence of nucleotides and ribosomal subunits show that the intermediate state is well populated, free ABCE1 in presence of ATP was not investigated.⁵⁶ In other members of the ABC-superfamily, an intermediate state of the NBDs is observed as well, although the coupling with trans-membrane-domains might influence the conformational space of the NBDs and reduce the transferability to ABCE1.^{9,82} In ABCE1, the intermediate state is considered necessary for regulation of ribosome separation.⁶³

However, for the purpose of testing our hypothesis that allostery is required for the asymmetric kinetics of the two mutants, i.e., to describe ATP hydrolysis of free ABCE1 in the absence of ribosomes, we believe that a third conformational state is not required. Instead, we assume that the effects of an intermediate state on the overall kinetics can be described by adopted transition rate coefficients of ligand exchange and ATP catalysis of the open and closed states in the Markov model. Consequently, the open and closed states of the Markov model can be considered as not strictly associated with the open and closed states observed in the structures. But, instead, might to be associated with a mixture of open/intermediate and intermediate/closed states with the distinction that in one state ligand exchange is possible but ATP catalysis is not, and vice versa for the other state.

Third, we assume that both binding sites can only open and close simultaneously, and that ABCE1 can close only if both binding sites are occupied by ATP and ADP. This assumption is supported by the fact that all ABCE1 structures that have been determined so far show both NBSs in identical conformations and,

if closed, have one non-hydrolyzable ATP analog in each NBS.^{47,57–60,62,63,66} Differences in FRET-efficiency between the two NBSs indicate a certain degree of independence of the NBSs,⁵⁶ but no direct measurement of the correlation between the conformational states of the two NBSs exists to date. Referring to the ABC-superfamily for support of this third assumption, we find that the prevailing model of ABC transporters assumes a symmetric conformation of both NBSs,⁸³ although a reciprocating model with asymmetric working NBSs has also been proposed.⁸⁴ Hofman et al. propose a partial opening of NBSII after phosphate release as a transition state to the open state.⁹ Taking together the above arguments, we conclude that a strict dependence between the conformational states of NBSs and an dependence of the closed state with occupation of the NBS with nucleotides is best in agreement with current the understanding of ABCE1 and the ABC-superfamily at its whole.

Fourth, we assume that no interaction between the two bindings site exists, i.e., ATP and ADP occupation of one binding site has no influence on the transition rates of binding, unbinding, hydrolysis, and synthesis of the opposite binding site. In the Markov model framework, this assumption can be realized by setting transition rate coefficients of same transition type, for example, ATP hydrolysis in NBSI, to the same value irrespective of the occupation of the other binding site. Transitions to which this applies have identical transition rate coefficients k_{ts} as depicted in figure 2.2.

The above assumptions result in a Markov model class of the wild-type with 13 states and 20 transitions with 10 unique transition rate pairs, i.e., 10 forward and 10 backward transition rate coefficients. Of these 20 transition rate coefficients, only 17 are independent due to the requirement to fulfill detailed balance. The supplementary table 2.1 lists all edge labels/transition rates together with the corresponding conformational change or ligand exchange.

The Markov model classes of the mutants are derived from the wild-type class by taking out the connections corresponding to ATP hydrolysis and synthesis of the NBS with the point mutation as illustrated by the two left panels in figure 2.2. This change corresponds to the assumption that the point mutations only influence ATP catalysis of the respective NBS. Thus, instances of wild-type and mutant classes are identical except for corresponding transition rate coefficients of ATP hydrolysis and synthesis of the impaired NBSs, i.e., k_2 and k_{12} for species E238Q and k_3 and k_{13} for species E485Q. Note, that instead of setting the corresponding

transition rate coefficients to zero to derive mutant instances from the wild-type instance, we reduced the transition rate coefficients by a factor of 100 in accordance that the mutation only impairs and not completely blocks ATP hydrolysis.^{47,67}

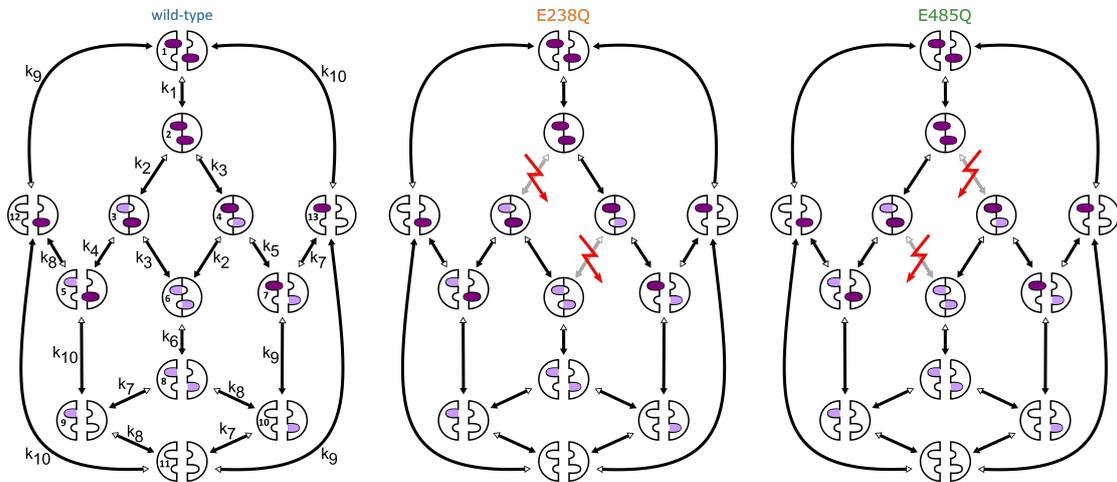


Figure 2.2 – Graph representations of the Markov model class of ABCE1. All transitions are reversible. Transition rate coefficient labels k_i belong to the transitions indicated by the solid arrow, i.e., forward direction and the labels of transition rate coefficient for the backward transitions are k_{i+10} (not shown). The two left graphs illustrate which transition rate coefficients are reduced to account for the impaired ATP hydrolysis capability by the mutation of the corresponding binding site.

No direct communication required We used the Metropolis-Hasting Monte Carlo algorithm to sample the posterior probability density in equation 2.2 with 500 chains of 14 million instances each.

If sampled instances reproduce the kinetics of wild-type and mutants E238Q and E485Q within error, we conclude that no direct communication between the NBSs is required for the asymmetric kinetics because the Markov model class of the instances excludes any direct communication between the NBSs. Figure 2.3 shows histograms of k_{cat} and K_M of the sampled instances normalized to probability densities, i.e, each panel is a probability density of the posterior density. The means of all six distributions agree well with the experimental determined values and the variances are within the experimental uncertainty, and thus instances that reproduce the experimentally measured kinetics within error do exist. Thus, no direct communication between the two NBSs is required for the asymmetric

kinetics of the three species.[†]

Next, we studied the mechanism that allows the Markov models to have the asymmetric kinetics between the mutants without relying on direct communication. To this aim, we identified the sequences of states a Markov model cycles through to hydrolyze ATP, for example, the sequence states 1-2-3-5-12 that correspond to hydrolyzing ATP in NBSI with ATP bound in the NSBII. We refer to the sequences as reaction pathways, and to states of the sequence as belonging to this reaction pathway. Note that one Markov model might have multiple reaction pathways.

We considered only the dominant reaction pathways, i.e, those reaction pathways that contribute at least 10% to the overall ATP turnover rate at saturating concentrations of ATP. For two instances to be considered to have the same dominant reaction pathways, the dominant reaction pathways of the wild-type instance and of the two derived mutant instances have to be identical. We refer to instances having the same triplet of dominant reaction pathways, one per species, as being of the same Markov model type. We determined the dominant reaction pathways of each instance and calculated the probabilities of each Markov model type based on the posterior values of the instances.

In total we determined over 4000 Markov model types. Figure 2.4 shows the reaction pathways of the three most probable Markov model types.

We identified that in all three Markov model types it is essential for the asymmetric kinetics of the three species that the steady state probabilities in the wild-type are asymmetrically distributed between states belonging to reaction pathways via NBSI, that dominate the ATP turnover rate in species E485Q, and states that do not belong to these reaction pathways. The low steady state probabilities of states belonging to reaction pathways via NBSI are a prerequisite for a large shift of steady state probability towards these states in species E485Q to realize the over tenfold increase of the ATP hydrolysis rate.

Because it is the most intuitive one, we use the third most probable Markov model type with 5.4% to illustrate this principle. In this Markov model type, ABCE1 uses two reaction pathways to hydrolyze ATP in all three species. One in which ATP is hydrolyzed in NBSI (upper right reaction pathway) and one in which ATP is hydrolyzed in NSBII (upper left reaction pathway), while in both

[†]A similar result based on a maximum-likelihood approach and with a reduction of the mutant affected transition rates of ATP catalysis to zero was already previously obtained.⁸⁵

cases an ATP is bound to the opposite NBS.

In the wild-type, the reaction pathway via NBSI contributes more to the overall ATP turnover rate than the reaction pathway via NBSII. Counter-intuitively, because anti-proportional to the ATP turnover rates, the states belonging to the reaction pathway via NBSII have a lower steady state population than the states belonging to the reaction pathways via NBSI (state 6 has over 80 % steady state population).

For species E485Q, the reduction of ATP hydrolysis in NBSII results in a large shift of steady state probability from states belonging to the reaction pathways via the now impaired NBSII towards the states belonging to the reaction pathways via NBSI. The increased probability ultimately leads to an increase of the ATP turnover rate via NBSI and to the ten-times increase of overall ATP turnover rate.

In contrast, for mutant E238Q with a reduced ATP hydrolysis in NBSI, only a minor shift in the steady-state population towards states belonging to reaction pathway via NBSI is observed, resulting in a decreased ATP turnover rate in NBSI and a reduction in overall ATP turnover rate.

At the same time, the asymmetric distribution of steady state probability allows to realize the reduction of the ATP-hydrolysis rate in mutant E238Q by a minor shift of steady state probability away from states of reaction pathways via NBSI.

Two factors contribute to the asymmetric distribution of steady state probability between states belonging to reaction pathways via NBSI and other states. First, the shorter the average time is required to cycle through a reaction pathway, i.e., the higher the transition rate coefficients of the transition of this pathway, the lower the steady state probabilities have to be for the reaction pathway to have a particular ATP turnover rate. Thus because, the ATP turnover is faster in NBSI, the Markov model has to have lower steady state probability of the states belonging to this pathway to produce the same ATP turnover rate as in the experiments, than of the states belonging to reaction pathways via NBSII.

Second, some reaction pathways are more probable to be 'chosen' than others. This second factor is best understood by reference to the above discussed Markov model type that has one reaction pathway via NBSI and one via NBSII in the wild-type. Both reaction pathways include the states 1 and 2, the open and closed state with two ATPs bound. Being in state 2, the system can 'choose' which ATP is hydrolyzed first and, thereby, 'chose' between the reaction pathway via NBSI and via NBSII. This decision depends only on the ratio of the two forward transition

rate coefficients of the corresponding transitions and, in particular, is independent of the ATP turnover time of the 'chosen' reaction pathway. Thus, this state acts as a fork between both reaction pathways and therefore is termed 'fork state'. Note that, in general, a similar situation can occur at state 3, only that the choice is between a conformational switch to the open state and hydrolyzing ATP in NBSII.

Both factors, alone or combined, can result in higher steady state probabilities of states of reaction pathways with lower ATP turnover rate than of states of reaction pathways with higher ATP turnover rate.

Figure 2.5 shows the two forward transition rate coefficients at the two above mentioned fork states 2 and 3, ultimately deciding which reaction pathway will be 'chosen'. The transition rate on the x-axis determines the transition rate coefficient corresponding to a reaction pathway over NBSII, and the transition rate coefficient on the y-axis to a reaction pathway including NBSI. Indeed, larger transition rate coefficients and, thereby, a bias towards reaction pathways over NBSI are observed. This bias counteracts the long turnover times of reaction cycles over NBSII and, thus, ensures that reaction pathways over NBSI contribute to the overall ATP turnover rate. This type of correlation was only found between these two pairs of transition rate coefficients, emphasizing the exceptional role of states 2 and 3.

The first and second most probable Markov model types follow the same principle of asymmetric distribution of steady state probability but use only reaction pathways over NBSI to hydrolyze ATP in wild-type and both mutants, i.e., reaction pathways over NBSII do not contribute significantly to the overall ATP turnover rate. The usage of only reaction pathways over NBSI is particularly peculiar for mutant E238Q, where the transition rate coefficients of ATP hydrolysis of NBSI are reduced by a factor 100 and, still, ATP is only hydrolyzed by reaction pathways via NBSI and the ATP turnover rate is only reduced by a factor of 2 to 3. Also this Markov model type is possible because in the wild-type the majority of steady state probability is not bound in states belonging to reaction pathways that contribute to the ATP turnover rate. Thereby, in species E238Q the reduction of transition rate coefficients can be compensated by an increase of steady state probabilities of the start and end states of the ATP catalysis.

In general, we determined over 4000 different Markov model types characterized by their reaction pathways triplets. The large number of possible combinations of reaction pathways is responsible for the low probability of each individual Markov

model type. Also, the overall number of Markov model types and their probabilities depends severely on the net flux threshold at which a transition is considered to be part of a reaction pathway or not. However, the three most probable Markov model types illustrated in figure 2.4 belong to the most probable Markov model types over a wide range of this threshold.

Prediction of nucleotide occupations So far, we showed that Bayesian approaches help to extract information from the samples of an underdetermined system by allowing to calculate probabilities of sets of samples with shared properties. Another notorious problem with the applications of maximum-likelihood approaches to underdetermined systems is the problem of overfitting, i.e., only a small set of Markov models with highest likelihood are sampled which because of their 'specialization' might fail to explain additional data that was not included in the likelihood. Here we tested if the extensive sampling of Markov models taking the experimental error into account can reduce the problem of overfitting. To this end, we checked whether the sampled instances can correctly predict ATP and ADP occupations of wild-type and both mutant species.

Occupation data from two different experimental setups are available.^{47,55} The ATP and ADP occupations by Nürenberg-Goloub et al. were measured with a 2:1 ratio of ATP to ABCE1 concentration. ATP and ABCE1 were let to interact for 30sec before reactions were stopped by cooling and the subsequent measurement of the occupations.

Figure 2.6 shows mean and standard deviation of the measured occupations by Nürenberg-Goloub et al. as red crosses together with the occupations calculated from the Markov models under quasi-steady-state assumption.⁵⁵ In none of the Markov models, ABCE1 binds more than one nucleotide and no Markov model has occupation values within two standard deviations of the experimental values.

The most probable cause for the discrepancy is that the quasi-steady-state assumption used to calculate the occupation values of the Markov models is not valid for this experimental setup. Because only twice as many ATP as ABCE1 molecules were used, binding and hydrolysis of ATP by each ABCE1 drastically changes the ATP and ADP concentrations, that are considered to be constant under the quasi-steady-state assumption. To reproduce the experimental setups a stochastic instead of a deterministic method to determine the occupation like Gillespie's-algorithm might be better suited.⁸⁶

In contrast, Barthelme et al. used a higher ATP concentration of 500 mM together with 5 μ M ABCE1, but measured only the ATP occupations.⁴⁷ Due to the 100:1 ratio of ATP to ABCE1, we consider the quasi-steady-state assumption to be more viable for this experimental setup.

Figure 2.7 shows the ATP and ADP occupations as calculated for the conditions of the experimental setup by Barthelme et al.⁴⁷ Double nucleotide bound states with non-zero steady state probability do occur in the Markov models as occupation values above 1 are achieved, in contrast to the occupation values in figure 2.6. The experimental ATP occupations within one standard deviation of experimental error are predicted with 0.7% probability.

Note that the limitation of one standard deviation from the experimental occupation as agreement with experiments is strict because the experimental errors are large and considered up two standard deviation qualitatively changes the interpretation of occupations, for example, for wild-type ABCE1 and measurements by Nürenberg-Goloub et al. an ATP occupation of zero or one are in agreement with experiment. Additionally, because both measured ATP occupations are well within one standard deviation, despite the different experimental setups, the error of one experiment overestimates the error of the combined data.

In particular, the wild-type ATP occupation value below one is responsible that most Markov models incorrectly predict the occupations. An ATP occupation below one means that the sum of steady state probabilities of states with less than one ATP bound is higher than 50%. Thus, to achieve an ATP occupation below one, the majority of the steady state probability has to be within the states with apo/apo, apo/ADP or ADP/ADP occupation, severely limiting the 'allowed' space of occupation conform Markov models.

Consequently, almost all Markov model types that conform with the occupation data use a reaction pathway in the wild-type hydrolyzing both ATPs. The most probable reaction pathway with 59% probability according to the Bayesian posterior for kinetics and occupations is shown in figure 2.8. This reaction pathway facilitates a low ATP occupation in the wild-type by allowing states with ADP or apo occupations to have higher steady state probabilities compared to Markov model types without this reaction pathway. Importantly, this Markov model type too obeys the principle of asymmetric steady state probability in the wild-type (in this case between a reaction pathway via NBSI and one via NBI and NBSII) which is created by larger transition rate coefficients for ATP hydrolysis

in NBSI and transition to the open state with ADP/ATP than ATP hydrolysis in NBSII, and thus emphasizes its universality.

Choice of Prior Distribution We want to briefly justify the choice of a logarithmic uniform prior on the transition rate coefficients as a prior for the Markov models of ABCE1. There are three possibilities for an uninformed prior. A uniform prior on the transition rate coefficients, a uniform prior on the inverse of the transition rate coefficients, i.e., the average transition time, or a logarithmic uniform prior on the transition rate. First, only one of the transition rates and average transition times can be uniformly distributed, and we know of no evidence for or against either option. In contrast, using a logarithmic uniform prior on the transition rate coefficients implies that the timescales are also logarithmically uniformly distributed, and thus avoids this problem. Second, Markov models describing physical systems should either be in equilibrium and satisfy detailed balance or be in a steady state. In the latter case, it must be possible for the resulting Markov model to satisfy detailed balance by scaling transition rates that depend on external factors that drive the system out of equilibrium to equilibrium conditions. For example, if a Markov model has substrate-concentration-dependent transition rate coefficients, the Markov model with these transition rate coefficients scaled according to the equilibrium concentrations of substrate and product should satisfy detailed balance. However, it is rather tedious to choose transition rate coefficients such that a Markov model satisfies detailed balance. In contrast, in the free energy representation of Markov models, detailed balance is automatically satisfied by choosing an absolute free energy for each Markov model; it is more convenient to sample the posterior density in the space of free energies. A logarithmic uniform prior on the transition rate coefficients facilitates this free energy approach because it is equivalent to a uniform prior on the free energy barriers.

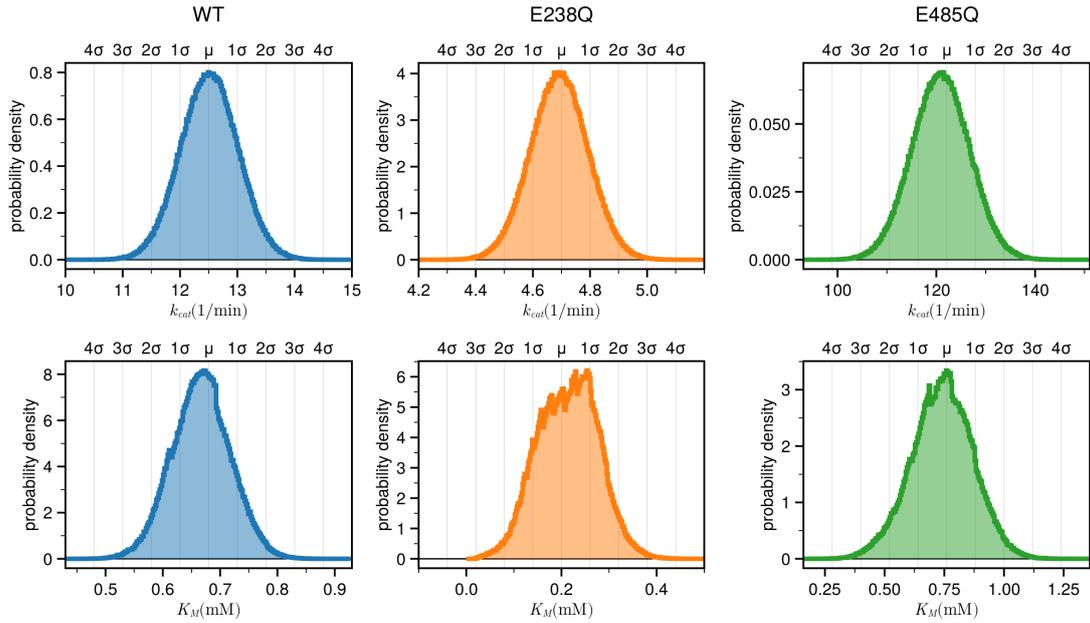


Figure 2.3 – Histograms of saturating ATP turnover rates k_{cat} (top row) and Michaelis-Menten constants K_M (bottom row) of the sampled instances normalized to probability densities. The columns show from left to right, the histograms of wild-type (blue), species E238Q (orange) and species E485Q (green). Each x-axis is centered around the corresponding experimental value and is limited to $\pm 5\sigma$ with σ being the standard deviation as reported by Barthelme et al.⁴⁷

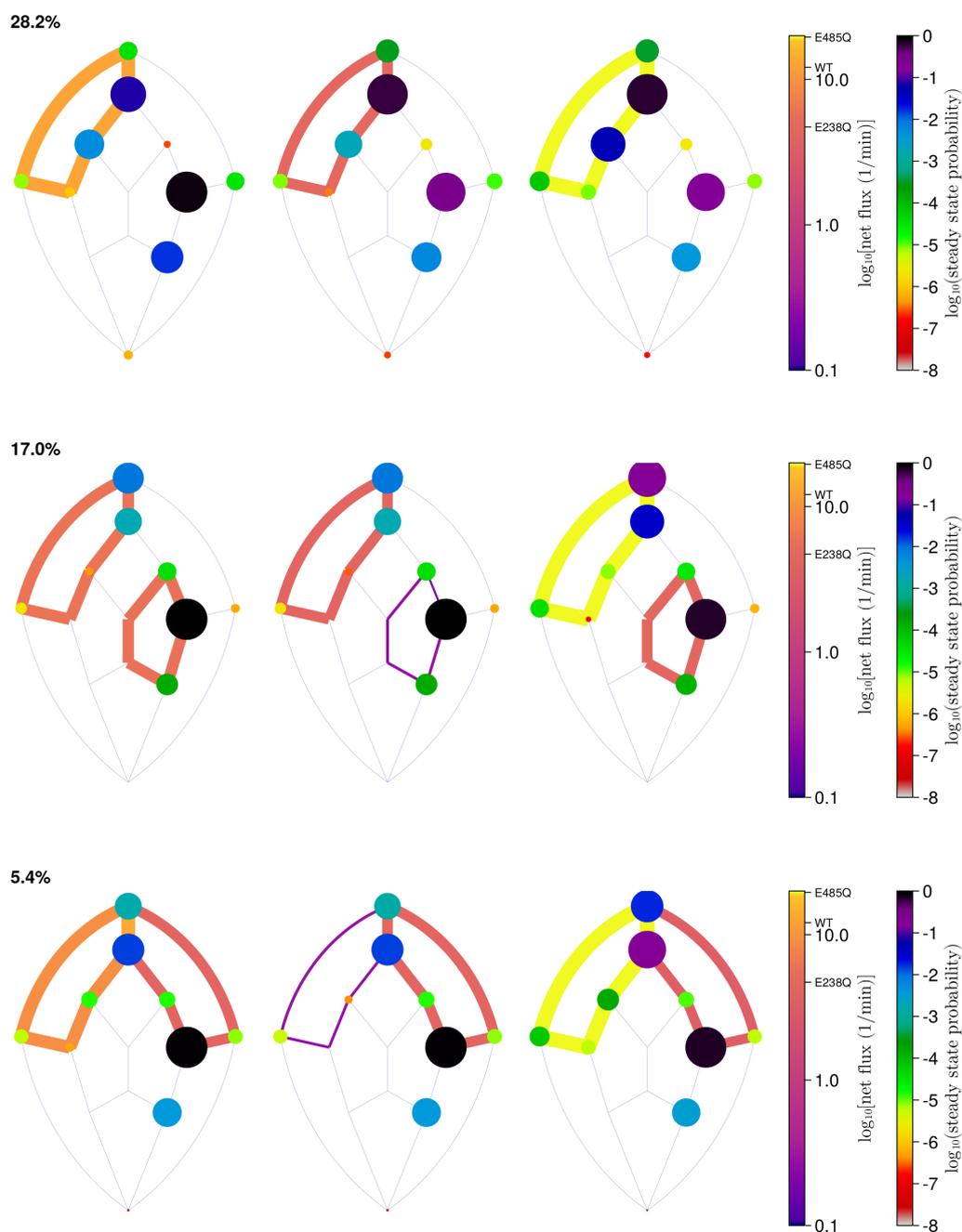


Figure 2.4 – The three most probable reaction pathways (most probable at the top to least probable at the bottom) according to the sampling of the Bayesian posterior. The line widths scale with the net flux of the transition, and the state sizes with the steady state probability of the state. Probability of each Markov model type is indicated in the top right corner.

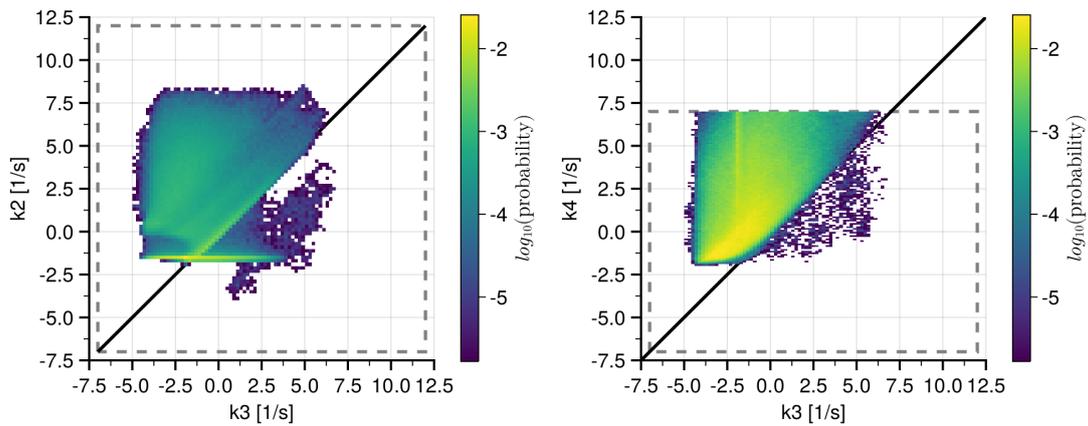


Figure 2.5 – Comparison of transition rate coefficients at the fork state 2 (left) and state 3 (right). Transitions corresponding to the transition rate coefficient on the x-axis lead towards reaction pathway via NBSII, whereas transitions corresponding to the transition rate coefficient on the y-axis lead towards reaction pathway via NBSI. Dashed gray lines indicate the boundaries given by the prior.

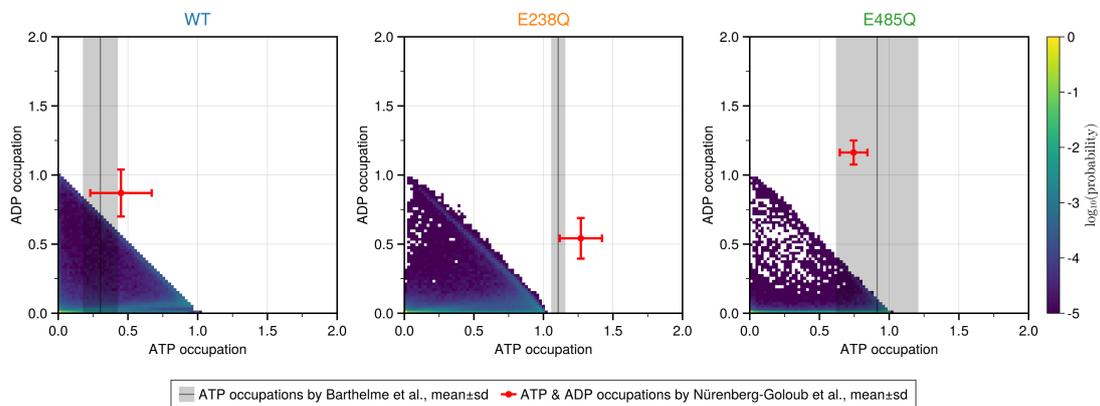


Figure 2.6 – ATP and ADP occupations as calculated from the sampled instances under experimental conditions of Nurenberg-Goloub⁵⁵ et al. and normalized to probability densities.

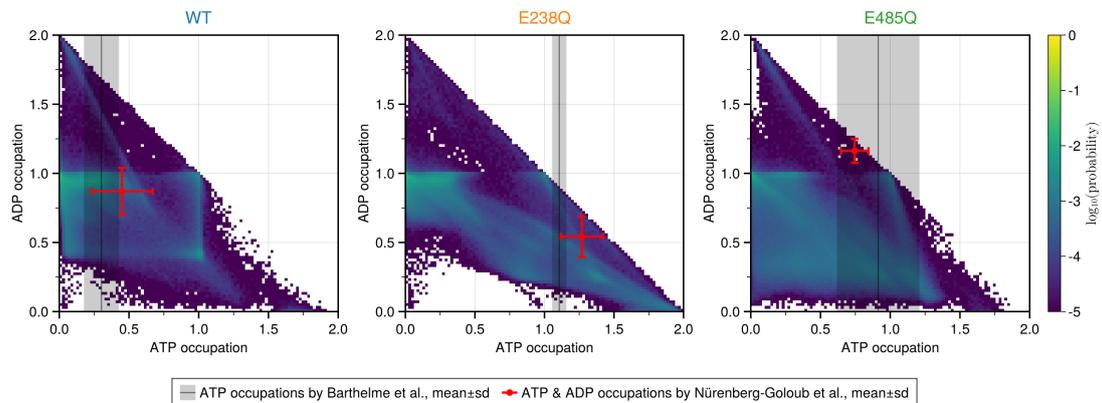


Figure 2.7 – ATP and ADP occupations as calculated from the sampled instances under experimental conditions of Barthelme et al.⁴⁷ and normalized to probability densities.

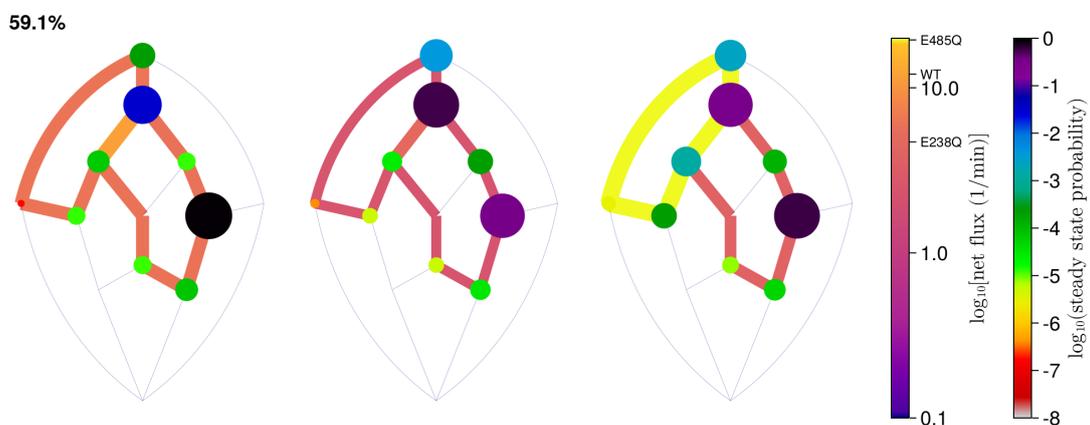


Figure 2.8 – Most probable reaction pathway after cross-validation with ATP occupations by Barthelme et al.⁴⁷ Probability of the Markov model type is indicated in the top right corner.

2.4 Conclusion

The ATPase ABCE1 has two nearly identical NBSs, both in terms of primary sequence and structure, suggesting that each NBS contributes equally to the ATP turnover rate. However, despite this near symmetry between the two NBSs, only one of the two mutants (E238Q), in which ATP hydrolysis is impaired in one of the two NBSs, has the expected twofold decrease in ATP-turnover rate, whereas the other shows a tenfold increase (E485Q).⁴⁷ We asked whether this observed kinetic asymmetry in both mutants can be explained without a direct allosteric interaction between the two NBSs of ABCE1 in the sense that the occupation of one NBS influences the kinetics of the opposite NBS, for example, by conformational rearrangements.

To this end, we described ABCE1 by Markov models which, by construction, exclude any direct communication between the two NBSs. We aimed at keeping the number of states and connections between them as low as possible while capturing all degrees of freedom required for ATP hydrolysis. We introduced the notion of a Markov model class to describe that we focused on the set of such Markov models with an equal number of states and with identical connections between them. We identified a subset of all possible Markov models of this Markov model class, referred to as instances, that agrees with the observed kinetics within experimental error, and calculated for each of these Markov models the Bayesian posterior, i.e., the probability that the Markov model is the one that originally produced the kinetics.

The main result of this chapter is that many Markov models were identified that, within experimental error, agree with the asymmetric kinetics of the two mutants. We conclude that a direct communication between the two NBSs is not required to explain the remarkable kinetics of ABCE1.

Inspection of the set of Markov models that agrees with experiment further allowed us to gain a deeper understanding of how this asymmetric kinetics was achieved without allostery. To this end, we identified subsets of instances that share predominant reaction pathways, determined the probability of these subsets, and thus obtained a rigorous ranking of possible reaction pathways. Indeed, as the most striking property shared by all investigated Markov model instances, we found that in the wild-type, the steady state probabilities are asymmetrically distributed between states belonging to reaction pathways via NBSI and NBSII.

This asymmetry is the result of an interplay between ABCE1 staying longer in reaction pathways with slower ATP turnover rate and the frequency with which reaction pathways are chosen. The later is decided in two fork-states at which the systems 'makes' a decision between reaction pathways. Preventing the system from choosing a reaction pathway with long ATP turnover time through the E458Q mutation results in a drastic redistribution of steady-state probability, leading to a tenfold increase in ATP turnover rate. In mutant E238Q, the modification results in stronger bias towards the reaction pathway with long ATP turnover time. However, because almost all steady-state probability is located in long ATP turnover time reaction pathways already in the wild-type, the additional bias results only in a minor shift of steady-state probability.

Thus ABCE1 serves as an example of allostery, that does not necessarily require direct 'allosteric pathways', but can also be brought about indirectly via redistribution of steady state probabilities as recently proposed.⁸⁷

The unexpected increase of ATP turnover rate in mutant E485Q is thus similar to the Braess paradox⁸⁸ for road networks that states that the closure of a road can improve the overall flow through the road network. The paradox is explained by drivers who choose pathways that promise the shortest travel time between their start and end, but without considering that the travel time across a road increases with more drivers on it.

The quality of the marginal probabilities depends on the convergence of the Bayesian posterior sampling, which remains challenging with Metropolis-Hastings Markov Chain Monte Carlo methods, especially for underdetermined systems with large sampling spaces, and thus advanced sampling methods, such as Hamiltonian Monte Carlo or replica exchange, may be considered to improve convergence.

One important and likely very general conclusion is that even in the absence of direct interactions between individual NBSs, the total turnover rate can not simply be understood as the sum of the turnover rates of the individual NBSs, as the tenfold increase of species E485Q strikingly demonstrates. Another example are Markov model types that hydrolyze ATP predominantly through NBSI in mutant E238Q, although even this NBS is impaired in this mutant.

The combination of extensive sampling of thermodynamically consistent Markov models with Bayesian inference offers the opportunity to not only gain insight into the functionality of ABCE1 but of other molecular machines as, for example, Myosin and the ATP synthesis complex F-ATPase.

2.5 Supplementary Information

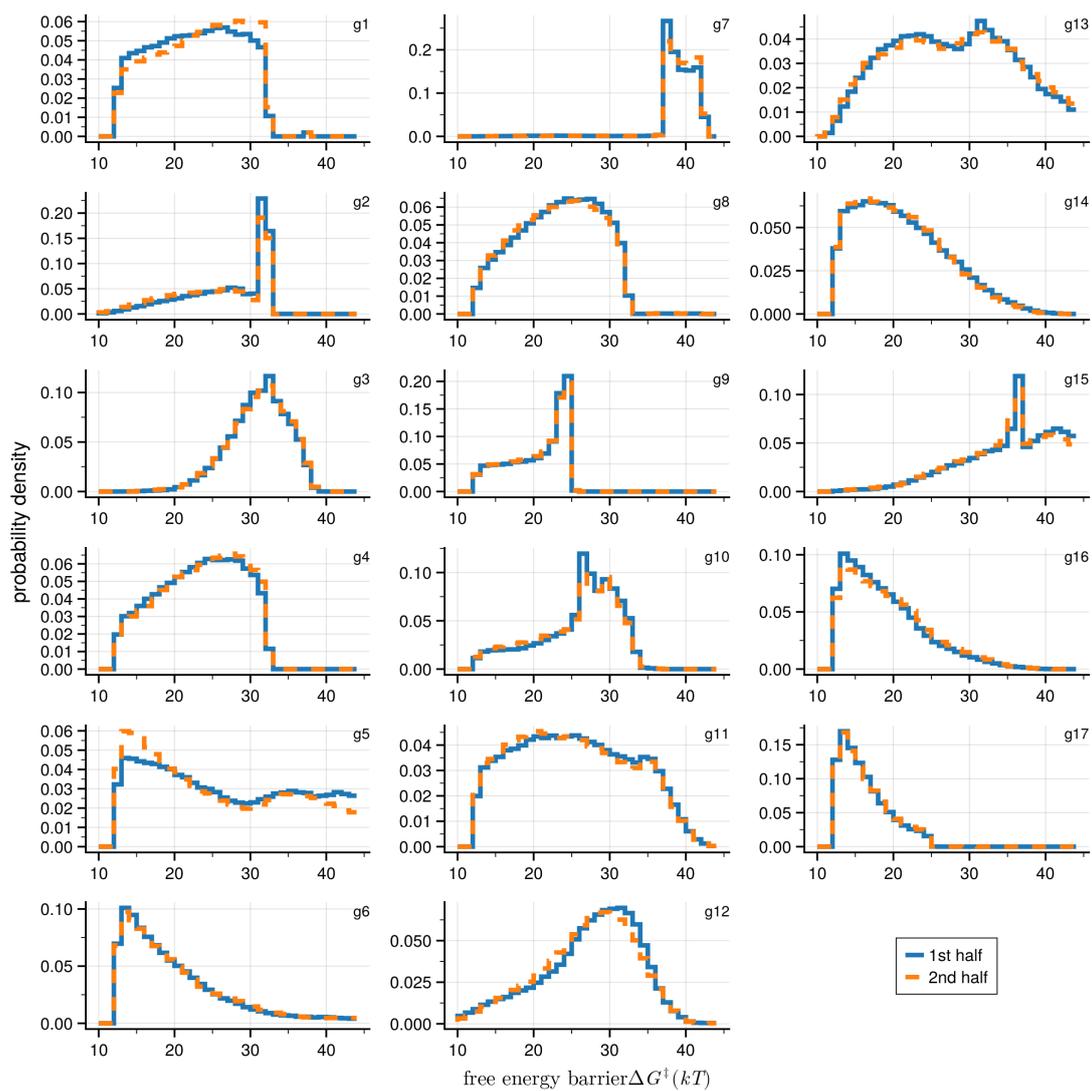
Table 2.1 – Transition rate coefficients and their correspondence to conformational changes

wild type	transition rates (forward, backward)		conformational change
	E238Q	E485Q	
k_1, k_{-1}	k_1, k_{-1}	k_1, k_{-1}	closure, opening (ATP,ATP)
k_2, k_{-2}	$10^{-2}k_2, 10^{-2}k_{-2}$	k_2, k_{-2}	ATP hydrolysis, synthesis 1st BS
k_3, k_{-3}	k_3, k_{-3}	$10^{-2}k_3, 10^{-2}k_{-3}$	ATP hydrolysis, synthesis 2st BS
k_4, k_{-4}	k_4, k_{-4}	k_4, k_{-4}	opening, closure (ADP,ATP)
k_5, k_{-5}	k_5, k_{-5}	k_5, k_{-5}	opening, closure (ADP,ADP)
k_6, k_{-6}	k_6, k_{-6}	k_6, k_{-6}	opening, closure (ADP,ADP)
k_7, k_{-7}	k_7, k_{-7}	k_7, k_{-7}	ADP unbinding, binding 1st BS
k_8, k_{-8}	k_8, k_{-8}	k_8, k_{-8}	ADP unbinding, binding 2nd BS
k_9, k_{-9}	k_9, k_{-9}	k_9, k_{-9}	ATP binding, unbinding 1st BS
k_{10}, k_{-10}	k_{10}, k_{-10}	k_{10}, k_{-10}	ATP binding, unbinding 2nd BS

Table 2.2 – Overview of boundaries of the prior distributions per type of transition

type of transition	lower bound	upper bound
ligand binding	$1 \times 10^{-6} \text{ 1/M}^2\text{s}$	$1 \times 10^7 \text{ 1/M}^2\text{s}$
ligand unbinding	$1 \times 10^{-6} \text{ 1/Ms}$	$1 \times 10^7 \text{ 1/Ms}$
catalysis	$1 \times 10^{-6} \text{ 1/Ms}$	$1 \times 10^{12} \text{ 1/Ms}$
conformational change	$1 \times 10^{-6} \text{ 1/Ms}$	$1 \times 10^6 \text{ 1/Ms}$

Figure 2.9 – First (solid blue) and second (dashed orange) halves of all 500 Monte Carlo Markov chains without burn-in. Values shown are related to transition rate coefficients in table 2.1 by $g_i = \log(k_i/10^{12})$. Good agreement is observed except for g_5 , the transition rate coefficient of transition from the open to closed conformation with ADP/ADP occupation.



3 Unexpected Michaelis-Menten behaviour of complex Markov models

3.1 Introduction

Increasingly many complex biomolecular processes are being described by Markov models, i.e., a set of discrete states with memory-free transitions between these states. Enzyme kinetics is an example, and the perhaps earliest kinetic model $E + S \rightleftharpoons ES \longrightarrow E + P$ by Michaelis and Menten²² is in fact a Markov model. Later, extensions of the Michaelis-Menten kinetics were established including inhibition²⁶ and reversibility,⁸⁹ and the increasing complexity eventually led to a formalization in terms of chemical reaction network theory.^{24,25,90}

Strictly, only chemical reaction networks comprised exclusively of first-order mass-action reactions can be described as Markov models. However, at least within the biochemical context, reaction networks with higher-order reactions can be described by Markov models as well, as long as the concentrations of all but one species can be assumed constant and, thus, the reactions effectively reduce to monomolecular first-order reactions. For biochemical processes, this reduction is known as quasi-steady-state assumption. Applied to the Michaelis-Menten kinetics, this leads to the canonical version of the Michaelis-Menten equation^{22,23}

$$v([S]) = \frac{v_{max}[S]}{K_M + [S]}, \quad (3.1)$$

where the change of the product formation rate v with substrate concentration $[S]$ depends on two parameters, the limiting rate v_{max} and the Michaelis-Menten constant K_M . Plotted on a logarithmic concentration scale, the resulting curve is sigmoidal.

The quasi-steady-state assumption turned out to be very useful and thus became a cornerstone of the analysis of enzyme kinetics, supported by extensive investigation of the particular condition for which it can be applied.²⁸⁻³⁰

More recently, and motivated by rapid advances in structural biology and atomistic simulations, internal changes of enzymes, other proteins, and biomolecular complexes shifted into focus. In particular, conformational sub-states and transitions between them were discovered⁷ and were later found to affect enzyme kinetics.⁴⁸ In contrast to chemical reactions and binding events described by chemical reaction networks, which are typically multimolecular, these conformational changes, are, by definition, monomolecular and thus, are readily described by Markov models as well.⁹⁰ In this context, Markov models provide a coarse-graining of otherwise high-dimensional atomistic descriptions such as molecular dynamics

simulations.^{16–18}

It is only recently widely recognized and accepted that the very fact that both the biochemical reaction networks and the molecular conformational transitions can be described by Markov models, allows one to combine both within a unified theory for complex biomolecular interaction networks of 'molecular machines'. Recent examples are the use of Markov models to analyze translation,^{31,32} transcription,^{33,34} signaling pathways,³⁵ molecular motors,^{36–43} and complex enzymes such as the fatty acid synthase,⁴⁴ rendering these accessible to theoretical investigations.

From a more mathematical perspective, the advantage of describing a process with Markov models is that its time-development is determined by a simple system of first-order differential equations, the so-called master equation. Notably, under very weak conditions (such as irreducibility and aperiodicity) the master equation has a unique steady state solution, which can be easily calculated via algebraic means.^{91,92} As a result, unified Markov models allow one to calculate turnover rates as a function of educt concentrations, e.g. substrate concentrations, in which case certain transition rates between Markov states become concentration dependent. We will restrict the current analysis to linear dependencies, i.e., concentration c dependent Markov transition rates between states t and s read $k'_{t,s}(c) = \frac{c}{c_0} k_{t,s}$ with normalization concentration c_0 . We will subsequently refer to concentration dependent quantities of the steady state solution as occupation curves. In the literature, these are also referred to as input-output-response curves.⁵⁰

While the original Michaelis-Menten model essentially comprised only two Markov states and three rate coefficients, increasingly complex processes and networks require up to several hundreds states and, correspondingly, several ten thousands rate coefficients.^{93,94} As a result, the system of first-order differential equations becomes quite large, and one would expect to see correspondingly complex occupation curves.

Quite to the contrary, however, we as well as others noticed in previous work that unexpectedly often, the obtained occupation curves seem to resemble simple Michaelis-Menten-like, sigmoidal, behavior.⁵⁰ For example, the ATP-driven enzyme ABCE1^{47,55} has at least 13 Markov states and 20 non-zero transition rate coefficients, six of which are concentration-dependent. Yet, the turnover rate of this system turns out to be sigmoidal most of the time.

Several studies already investigated this unexpected ubiquity of Michaelis-Men-

ten behavior. A series of studies looked at Markov models with a 2D-grid-like structure where one dimension describes the product catalysis and the other dimension interconversion between enzyme conformers. They derived conditions on the transition rate coefficients for Markov models exhibiting Michaelis-Menten-like behavior.⁹⁵⁻⁹⁸ For more general Markov models, Barel et al. found that, if the ratio of steady state probabilities of every unbound enzyme states stay constant for all substrate and product concentrations, the unbound enzyme states can be effectively treated as one state. From this observation follows further that the occupation curve are described by the reversible Michaelis-Menten equation.⁴⁹ Further, Wong et al. derived structural conditions on Markov models based on division of a Markov model into subsets divided by concentration-dependent transition rates.⁵⁰ Both conditions were found to be sufficient, but not necessary, for the occurrence of Michaelis-Menten-like behavior in the occupation curves.^{49,50}

The studies mentioned above focused on the case where an occupation curve has exactly one Michaelis-Menten-like piece, i.e., is fully described by the Michaelis-Menten equation. However, also occurrences of occupation curves with multiple pieces of Michaelis-Menten-like behavior are reported in the literature,⁵²⁻⁵⁴ and occurred also for Markov models of ABCE1.

Thus, here we extend the scope of the ubiquity of Michaelis-Menten-like behavior to occupation curves that behave Michaelis-Menten-like on pieces of the occupation curve. In particular, we ask why occupation curves of complex Markov models often have so many Michaelis-Menten-like pieces, and how probable it is for such occupation curves to have a Michaelis-Menten-like piece.

3.2 Theory

Motivation Time-continuous Markov models describe a system as a set of N states and a set of $N^2 - N$ transition rate coefficients $\{k_{ts}\} \in \mathbb{R}_{\geq 0}$ describing the probability (per second) of a transition from state s to state t . Note that we define n as the number of non-zero k_{ts} . The probability of the system to be in a state i at a given time t is given by $X_i(t)$, and, given an initial distribution $\mathbf{X}(0)$, the time evolution of the probabilities $\mathbf{X}(t)$ is determined by the system of linear differential equations

$$\dot{\mathbf{X}} = \mathbf{Q} \cdot \mathbf{X}(t). \quad (3.2)$$

referred to as master equations, whose transition rate coefficient matrix \mathbf{Q} has the diagonal elements $q_{tt} = - \sum_{s=1, s \neq t}^N k_{ts}$. Note that q_{ts} are referred to as transition rates in the context of Markov models, but we refer to them as transition rate coefficients to be consistent with chemical terminology.

For irreducible and aperiodic Markov models, the probabilities $\mathbf{X}(t)$ will converge towards unique stationary or steady state probabilities $\boldsymbol{\pi}$, such that

$$\mathbf{0} = \mathbf{Q} \cdot \boldsymbol{\pi}. \quad (3.3)$$

Note that $\boldsymbol{\pi}$ can be calculated by algebraic means via the eigenvalue decomposition of \mathbf{Q} , because $\boldsymbol{\pi}$ is the eigenvector to eigenvalue 0.

The net flux of a transition is the amount of probability per time unit that is being transferred from one state to another and is calculated as the difference of the steady state probability times the transition rate coefficient between forward and backward transition.¹⁹ Given steady state probabilities $\boldsymbol{\pi}$, the net fluxes obey one of two conditions. Either, for a Markov model obeying detailed balance and describing a system in equilibrium, all net fluxes will be zero. Or, for a Markov model describing a system away from equilibrium, the net fluxes between transitions will form closed cycles, such that in- and outgoing net fluxes of one state sum to zero. In chemical reaction network theory the latter condition is referred to as complex detailed balance.^{24,25}

As an example consider a Markov model of ATPase ABCE1 with $N = 13$ and $n = 20$, whose graph representation is shown in figure 3.1A. The states correspond to ABCE1 in its open and closed conformation with each of its two binding sites

empty, ATP or ADP bound and the transitions corresponding to binding reactions, catalysis or conformational changes. It is a sparse Markov model because all states have vertex degree three, i.e., are connected to three other states, except state 8 that has vertex degree four.

The twelve transitions of ATP and ADP binding are second-order reactions, which are reduced to first-order reactions under the quasi-steady-state assumption allowing to describe the system as a Markov model. Thus, transition rate coefficient matrix \mathbf{Q} has entries of k_{ts} for concentration-independent transitions and of $k'_{ts} = \frac{c}{c_0} k_{ts} =: \alpha k_{ts}$ for concentration-dependent transitions under assumption of a linear concentration dependence. Note that we refer to k_{ts} as a concentration-independent transition rate coefficient and k'_{ts} as a concentration-dependent transition rate coefficient, and define m as the number of concentration-dependent transition rate coefficient. The explicit specification of a transition rate coefficient as concentration-independent might seem contradictory as per definition a transition rate coefficient is always concentration-dependent. However, it serves as short-hand notation for a transition rate coefficient that either depends on one of the concentrations that are considered constant due to the quasi-steady-state assumption or not, i.e., if it is a transition rate coefficient is first or higher order before application of the quasi-steady-state assumption.

The measured dependence on ATP concentration of the ATP turnover rate of ABCE1 is well described by the Michaelis-Menten kinetics.⁴⁷ In the Markov model framework, the steady state ATP turnover rate can be calculated from the net fluxes of ATP binding or ADP unbinding transitions, which in return depend on the steady state probabilities that are connected by these transitions. A change of substrate concentrations, i.e., ATP and ADP concentrations, results in a change of the corresponding transition rate coefficient k'_{ts} and, thus, the steady state probabilities, net fluxes, and turnover rates. The steady state probability π_N as function of the substrate concentrations is referred to as an occupation curve.

Selected examples of occupation curves are shown in figure 3.1B. Each occupation curve results from a Markov model of ABCE1 with states connected as shown in figure 3.1A but with different values of transition rate coefficients. Pieces of the occupations curves that resemble the Michaelis-Menten kinetics are colored green. The occupation curves are sorted from top left to bottom right according their similarity to the Michaelis-Menten kinetics, i.e, how close they are to be described by the Michaelis-Menten equation 3.1.

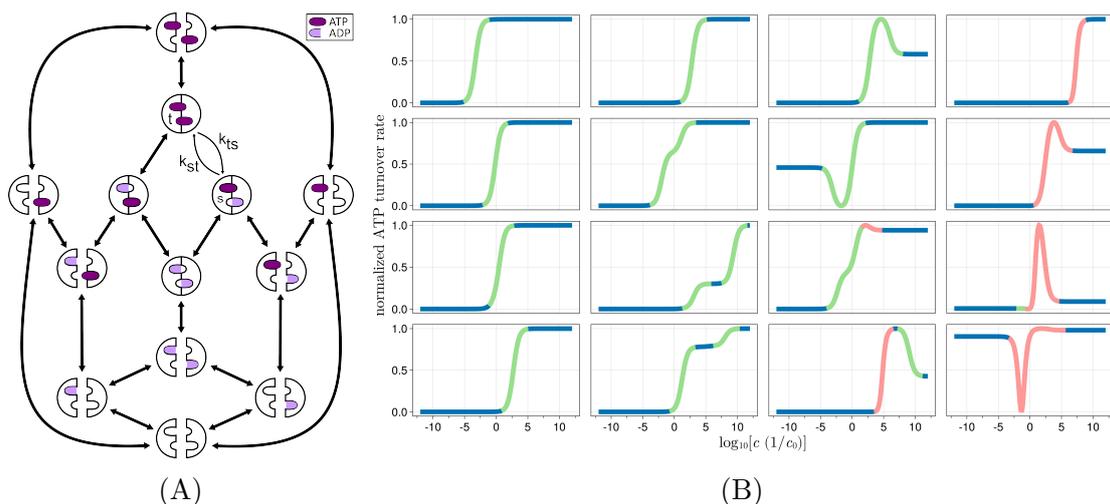


Figure 3.1 – (A) Graph representation of the Markov model of ABCE1 (B) Selected examples of occupation curves of Markov model of ABCE1. Green pieces are considered to be Michaelis-Menten-like based on square root mean error between a fit of a modified Michaelis-Menten equations with offset and negative nominator allowing for a descending sigmoid curve. Based on the same criterion, red pieces are considered to be not Michaelis-Menten-like. Examples are sorted from top left to bottom right according to their similarity to the Michaelis-Menten kinetics.

Analytical solution To explain this unexpected Michaelis-Menten-like behavior, we seek an analytical solution of the steady state probability π_i . To this end we use a version of Kirchhoff’s matrix tree theorem⁹⁹ that was derived by Tutte,¹⁰⁰ and subsequently used by Gunawardena et al. to explain common properties of Markov models in biochemical applications.²⁷

Accordingly to the matrix tree theorem the steady state probability of state i is

$$\pi_i = \frac{\sum_{r \in R^{(i)}} \left(\prod_{ts \in r} k_{ts} \right)}{\sum_{i=1}^n \sum_{r \in R^{(i)}} \left(\prod_{ts \in r} k_{ts} \right)}. \quad (3.4)$$

Here, the products iterate over all edges (directed transition between two Markov states) in a rooted spanning tree r , and the inner sums are over all rooted spanning trees r that belong to the set $R^{(i)}$ of all rooted spanning trees rooted at state i . A rooted spanning tree is a subset of $N - 1$ edges connecting all N states, such that every of the $N - 1$ edges is part of a path leading towards state i .

Next, if we assume that a subset of transition rate coefficients k_{ts} are concentration-dependent transition rate coefficients, i.e., proportional to c , equation 3.4 becomes a rational function in c by rearranging the sums with respect to power of c

$$\pi_i = \frac{p^{(i)}(c)}{q(c)} = \frac{p_0^{(i)} + p_1^{(i)}c + p_2^{(i)}c^2 + \dots + p_D^{(i)}c^D}{q_0 + q_1c + q_2c^2 + \dots + q_Dc^D} = \frac{\sum_{\kappa=0}^D p_{\kappa}^{(i)}c^{\kappa}}{\sum_{\kappa=0}^D q_{\kappa}c^{\kappa}} \quad (3.5)$$

with coefficients

$$p_{\kappa}^{(i)} = \sum_{r \in R_{\kappa}^{(i)}} \left(\frac{1}{c_0^{\kappa}} \prod_{ts \in r}^{N-1} k_{ts} \right) \quad (3.6)$$

and

$$q_{\kappa} = \sum_{i=1}^N p_{\kappa}^{(i)} = \sum_{r \in R_{\kappa}} \left(\frac{1}{c_0^{\kappa}} \prod_{ts \in r}^{N-1} k_{ts} \right) \quad (3.7)$$

where $R_{\kappa}^{(i)}$ is the set of rooted spanning trees rooted at state i with κ concentration-dependent transition rate coefficients, R_{κ} is the set of rooted spanning trees with κ concentration-dependent transition rate coefficients, r the set of transition rate coefficients k_{ts} of a rooted spanning tree, and D is the maximal number of concentration-dependent transition rate coefficients found in any rooted spanning tree. We define $\nu_{\kappa}^{(i)}$ as the number of elements in the set $R_{\kappa}^{(i)}$, and ν_{κ} as the number of elements in the set R_{κ} . To simplify notation, we will subsequently omit the index $^{(i)}$ from the coefficients $p_{\kappa}^{(i)}$. For later use note that, because $R_{\kappa}^{(i)} \subset R_{\kappa}$, it follows that $p_{\kappa} \leq q_{\kappa} \forall \kappa$. Note also that some p_{κ} or q_{κ} may be zero, such that the number Δ of non-zero terms in $p(c)$ or $q(c)$ in equation 3.5 may be smaller than D .

Piecewise approximation From the analytical solution (equation 3.4) of the steady state probability, it follows that all $\pi_i(c)$ are rational functions in concentration c . Next we will, though not with full mathematical rigor, motivate that $\pi_i(c)$ can be approximated in a piecewise manner by Michaelis-Menten-like functions if the coefficients fulfill certain conditions, which we will assume for the moment and derive subsequently.

For small enough c , such that $p_0 \gg p_b c^b$ and $q_0 \gg q_b c^b \forall b > 0$, $\pi_i(c)$ can be approximated by a constant,

$$\pi_i(c) = \frac{p_0 + p_1 c + p_2 c^2 + \dots + p_D c^D}{q_0 + q_1 c + q_2 c^2 + \dots + q_D c^D} \approx \frac{p_0}{q_0}. \quad (3.8)$$

For slightly larger c , such that $\{p_0, p_1 c\} \gg p_b c^b$ and $\{q_0, q_1 c\} \gg q_b c^b \forall b \neq \{0, 1\}$, the first-order terms in c will contribute as well, such that $\pi_i(c)$ can be approximated as

$$\pi_i(c) = \frac{p_0 + p_1 c + p_2 c^2 + \dots + p_D c^D}{q_0 + q_1 c + q_2 c^2 + \dots + q_D c^D} \approx \frac{p_0 + p_1 c}{q_0 + q_1 c}. \quad (3.9)$$

We will identify this terms with the Michaelis-Menten behaviour further below.

For further increasing c , the constant terms becomes small with respect to the terms linear in c and, similarly, for further increasing c , $\pi_i(c)$ can be approximated by the linear and quadratic terms for c obeying $\{p_1 c, p_1 c^2\} \gg p_b c^b$ and $\{q_1 c, q_1 c^2\} \gg q_b c^b \forall b \neq \{1, 2\}$.

Repeating these steps for increasing c , $\pi_i(c)$ can be approximated in a piecewise-manner by consecutive terms, i.e.,

$$\pi_i(c) = \frac{p_0 + p_1 c + p_2 c^2 + \dots + p_D c^D}{q_0 + q_1 c + q_2 c^2 + \dots + q_D c^D} \approx \frac{p_\kappa c^\kappa + p_{\kappa+1} c^{\kappa+1}}{q_\kappa c^\kappa + q_{\kappa+1} c^{\kappa+1}} = \frac{p_\kappa + p_{\kappa+1} c}{q_\kappa + q_{\kappa+1} c}, \quad (3.10)$$

for $[p_\kappa c^\kappa, p_{\kappa+1} c^{\kappa+1}] \gg p_b c^b$ and $[q_\kappa c^\kappa, q_{\kappa+1} c^{\kappa+1}] \gg q_b c^b \forall b \neq \{\kappa, \kappa + 1\}$.

Finally, for $p_D c^D \gg p_b c^b$ and $q_D c^D \gg q_b c^b \forall b \neq D$, $\pi(c)$ saturates at a constant value,

$$\pi_i(c) = \frac{p_0 + p_1 c + p_2 c^2 + \dots + p_D c^D}{q_0 + q_1 c + q_2 c^2 + \dots + q_D c^D} \approx \frac{p_D}{q_D}, \quad (3.11)$$

Note that, up to an offset, each of the above piecewise approximation represents the Michaelis-Menten function

$$\frac{p_\kappa + p_{\kappa+1} c}{q_\kappa + q_{\kappa+1} c} = \frac{v_\kappa c}{K_\kappa + c} + \frac{p_\kappa}{q_\kappa} \quad (3.12)$$

with

$$v_\kappa = \frac{p_{\kappa+1}}{q_{\kappa+1}} - \frac{p_\kappa}{q_\kappa} \quad (3.13)$$

and

$$K_\kappa = \frac{q_\kappa}{q_{\kappa+1}}. \quad (3.14)$$

This piece of the approximation provides a sigmoidal step between $\frac{p_\kappa}{q_\kappa}$ and $\frac{p_{\kappa+1}}{q_{\kappa+1}}$, which also generalizes the classical Michaelis-Menten equation to include its negative or inverted form describing, e.g., the kinetics of non-competitive inhibition.²⁶

Depending on the length of the individual concentration ranges, the adjacent Michaelis-Menten curves may merge or be well separated from each other. The later requires their positions, defined by the inflection point given by the Michaelis-Menten constant K_κ , to be separated by at least 2 orders of magnitude, where the numerical value is empirical and, thus, subject to subjectivity. This condition translates into the necessary (but not sufficient) condition

$$\varkappa := \frac{q_{\kappa+1}/q_{\kappa+2}}{q_\kappa/q_{\kappa+1}} > 10^2. \quad (3.15)$$

Note that, because this is an inverted double difference on the logarithmic scale, this conditions essentially means that the 'curvature' of the logarithmic coefficients as function of their respective degree κ must be larger than 2 which requires a concave parabola-like shape. Note that a maximum of $D-1$ Michaelis-Menten-like pieces can occur.

We note that for this piecewise approximation to work, the intervals for c given by the coefficients p and q , respectively, need to sufficiently overlap, thus,

$$\frac{p_\kappa}{p_{\kappa+1}} \approx \frac{q_\kappa}{q_{\kappa+1}}. \quad (3.16)$$

Therefore, the above condition must be additionally fulfilled for a sigmoid curve to actually occur on the respective interval.

For the sake of completeness, note that, strictly, if only one of the two nominator terms of equal degree is dominant in the respective interval, the approximation can still be cast into a Michaelis-Menten-like form

$$\frac{p_0 + p_1c + p_2c^2 + \dots + p_Dc^D}{q_0 + q_1c + q_2c^2 + \dots + q_Dc^D} \approx \frac{p_{\kappa+1}c}{q_\kappa + q_{\kappa+1}c} = \frac{p_{\kappa+1}/q_{\kappa+1} c}{q_\kappa/q_{\kappa+1} + c} \quad (3.17)$$

$$\frac{p_0 + p_1c + p_2c^2 + \dots + p_Dc^D}{q_0 + q_1c + q_2c^2 + \dots + q_Dc^D} \approx \frac{p_\kappa^{(N)}}{q_\kappa + q_{\kappa+1}c} = \frac{-p_\kappa^{(N)}/q_\kappa c}{q_\kappa/q_{\kappa+1} + c} + \frac{p_\kappa^{(N)}}{q_{\kappa+1}}, \quad (3.18)$$

whereas two consecutive dominant terms in the denominator are always required.

Symmetry Next we will estimate the fraction of all possible Markov models for which the occupation curves exhibit a particular number of sigmoidal pieces. To this aim, it will be helpful to note a symmetry between pairs of Markov models with complementary concentration dependencies defined as follows.

Given a Markov model with a transition rate matrix $\mathbf{Q}(\alpha)$ that contains n concentration-independent rates k_{ts} and m concentration-dependent rates αk_{ts} . We define its complementary Markov model with transition rate matrix $\mathbf{Q}'(\alpha)$ by mapping all above concentration-independent rates into concentration-dependent rates αk_{ts} and vice-versa.

Then the occupation curves of $\mathbf{Q}(\alpha)$ and $\mathbf{Q}'(\alpha)$ are symmetrical to each other on the logarithmic scale with respect to inversion of α , i.e.,

$$\pi_i(\mathbf{Q}(\alpha)) = \pi_i(\mathbf{Q}'(1/\alpha)). \quad (3.19)$$

To see why, note that from the above mapping of the elements of \mathbf{Q} it follows that

$$\mathbf{Q}(\alpha) = \alpha \mathbf{Q}'(1/\alpha). \quad (3.20)$$

Because π_i is the eigenvector of \mathbf{Q} , it follows that it is also the eigenvector of $\alpha \mathbf{Q}'(1/\alpha)$ and, therefore, of $\mathbf{Q}'(1/\alpha)$, from which the above symmetry follows. In particular, the number of Michaelis-Menten-like pieces is identical.

Expected number of Michaelis-Menten-like pieces Next, we estimate the coefficients p_κ and q_κ via equations 3.6 and 3.7, respectively. In both cases, these are composed of a sum of products of transition rate coefficients. Assuming the later to be log-uniform distributed on finite but large intervals, these products tend towards a log-normal distribution for a sufficiently large number of states N (see supplementary information). Moreover, because the number of factors, $N - 1$, is the same in each of the products both for the nominator and denominator, the two respective distributions are similar, and both types of coefficients tend towards sums of log-normal distributed random variables.

To our knowledge, there exists no analytical expression, nor approximation suited for the present purpose, for the mean of such a sum. Thus, we aim at

expressing the mean as a function of the number of terms, i.e., the number ν_κ of rooted spanning trees R_κ with κ concentration-dependent transition rate coefficients. Note that, even for Markov models with identical number of states N and sequence of vertex degrees $\{vd_1, \dots, vd_i, \dots, vd_N\}$, i.e., state i is with vd_i other states connected, ν_κ depends on the particular choice of which states are connected as well as the transition rate coefficients of which are concentration-dependent. We discuss in the following the means $\langle q_i \rangle$ and $\langle \nu_\kappa \rangle$ over these choices.

In the absence of analytic solutions for $\langle \nu_\kappa \rangle$, we resort to approximation by combinatorial means. To this end, note that the set of rooted spanning trees is a subset of the set of all ways of drawing $N - 1$ transition rate coefficients from a set of n transition rate coefficients. For this larger set, the number of elements is

$$\tilde{\nu} = \binom{n}{N-1}. \quad (3.21)$$

Similarly, the set R_κ is a subset of the set of all ways to draw κ concentration-dependent transition rate coefficients from a set of n transition rate coefficients of which m are concentration-dependent if a total of $N - 1$ transition rate coefficients are drawn without replacement and, thus, the number of elements of this larger set is

$$\tilde{\nu}_\kappa = \binom{m}{\kappa} \binom{n - (N - 1)}{m - \kappa}. \quad (3.22)$$

We next approximate $\langle \nu_\kappa \rangle$ by

$$\langle \nu_\kappa \rangle = \frac{\langle \nu_\kappa \rangle}{\tilde{\nu}_\kappa} \tilde{\nu}_\kappa \approx \frac{\langle \nu \rangle}{\tilde{\nu}} \tilde{\nu}_\kappa = \frac{\tilde{\nu}_\kappa}{\tilde{\nu}} \langle \nu \rangle = \widetilde{\langle \nu_\kappa \rangle}, \quad (3.23)$$

which depends on similar fractions $\langle \nu_\kappa \rangle / \tilde{\nu}_\kappa$ and $\langle \nu \rangle / \tilde{\nu}$ for which we will provide numerical evidence further below. Note that $\langle \nu \rangle = N \langle \bar{\nu} \rangle$ with $\langle \bar{\nu} \rangle$ the average number of spanning trees for which an asymptotic expression for $n \rightarrow \infty$ and sparse graphs exists.¹⁰¹

Further, note that $\tilde{\nu}_\kappa / \tilde{\nu}$ is the the probability mass function of the hypergeometric distribution \mathcal{H} and that

$$\mathcal{H} \approx \mathcal{N}(\mu_{\mathcal{H}}, \sigma_{\mathcal{H}}) \quad (3.24)$$

for $m, n \ll N - 1$ and thus, $\log(\langle \nu_\kappa \rangle)$ should approximate a parabola.¹⁰²

Let us check the quality of the approximation in equation 3.23, and thus how well our assumption is that the two fractions $\langle \nu_\kappa \rangle / \widetilde{\nu}_\kappa$ and $\langle \nu \rangle / \widetilde{\nu}$ are similar, by means of explicit enumeration of rooted spanning trees $\langle \nu_\kappa \rangle$ for Markov models with $N = 20$, which in light of the exponential increase of the number of rooted spanning trees with the number of states is about the maximum we can do.

Further, in the case that N is even, we consider 3-regular Markov models, i.e., all states have vertex degree three. For odd N , we consider Markov models where all states have vertex degree three except one state that has vertex degree two instead, which for simplicity we will also refer to as '3-regular'. This choice of low vertex degree is motivated by the observation that Markov models of biochemical processes are sparse. In contrast, biochemical networks differ largely in how many transitions are actually concentration dependent. Therefore, we considered the full range of m from 0 to n with n being the number of all transition rate coefficients in a Markov model. For later use, we define the fraction of concentration-dependent transition rate coefficients m/n .

The results of the above approximation $\widetilde{\langle \nu_\kappa \rangle}$ of $\langle \nu_\kappa \rangle$ (equation 3.23) are shown in figure 3.2A, and compared for reference to the exact numbers in figure 3.2B. Here, the logarithm of $\langle \nu_\kappa \rangle$ is plotted color coded as a function of the number of concentration-dependent transition rate coefficients m . These values have been obtained by averaging over 64 Markov models with random choices of which states are connected and also which transition rate coefficients are concentration-dependent. We find a good agreement between $\langle \nu_\kappa \rangle$ and $\widetilde{\langle \nu_\kappa \rangle}$. Note that $\widetilde{\langle \nu_\kappa \rangle}$ has values below one for certain combinations of m and κ , which gives the probability to observe a $\nu_\kappa = 1$ assuming a Markov model with exactly the average number of rooted spanning trees $\langle \nu \rangle$. The black line separates between values of $\widetilde{\langle \nu_\kappa \rangle}$ above and below 0.5. We conclude that $\langle \nu_\kappa \rangle$ can be well approximated by $\widetilde{\langle \nu_\kappa \rangle}$.

For later use, we assess how well the hypergeometric distribution approximates the fraction $\frac{\langle \nu_\kappa \rangle}{\langle \nu \rangle}$. To this aim, figure 3.2C shows the logarithm of $\langle \nu_\kappa \rangle$ (circles) and $\widetilde{\langle \nu_\kappa \rangle}$ (crosses) normalized by $\langle \nu \rangle$ and $\widetilde{\langle \nu \rangle}$, respectively, as function of κ for $m = \{4, 8, 12, \dots, 28, 30\}$, representing vertical cuts trough figures 3.2A and 3.2B.

As can be seen, for all shown values of m , overall, the normalized $\langle \nu_\kappa \rangle$ agree well with normalized $\widetilde{\langle \nu_\kappa \rangle}$. Exceptions are an underestimation of the normalized $\langle \nu_\kappa \rangle$ for high values of κ for $m > 4$ and an overestimation for low value of κ for $m > 24$. Further, note also that the approximation of the hypergeometric distribution by normal distribution (equation 3.24) is quite accurate, as can be

seen by comparison with the orange parabolas.

After having validated the approximation of $\langle \nu_\kappa \rangle$, we finally need to assess how $\langle q_\kappa \rangle$ depends on $\langle \nu_\kappa \rangle$ by estimation of the correlation between ν_κ and q_κ . No suitable closed form is available,¹⁰³ too, such that we will again resort to numerical estimation of the coefficients of Markov models with $N \in \{4 \dots 20\}$, vertex degree three, transition rate coefficients which are log-uniform distributed between 10^{-12} and 10^{12} , and for $n/m \in \{1/5, 2/5, 1/2\}$.

Figure 3.3B shows the relation between the logarithms of ν_κ and q_κ exemplary for Markov models with $N = 19$. We describe the correlation between both values with a proportionality and determine the constant of proportionality by linearly fitting $q_\kappa(\nu_\kappa)$ of each individual Markov model. Figure 3.3A shows the distributions of parameters of these fits as box plots for each N . For $N > 20$, the constant of proportionality was estimated by linear extrapolation from the slopes for $N \leq 20$.

Note that the average constant of proportionality is larger for Markov models with an odd number of states N than for Markov models with an even number of states, which we suspect is due to the above mentioned non-uniformity of vertex degrees of Markov models with odd N .

Combining the estimate of $\langle \nu_\kappa \rangle$ through approximation in equation 3.23 with numerical estimate of the correlation between ν_κ and q_κ allows to estimate

$$\langle q_\kappa \rangle \approx \langle \nu_\kappa \rangle ((12.2 \pm 0.25)N + (2.5 \pm 0.82)). \quad (3.25)$$

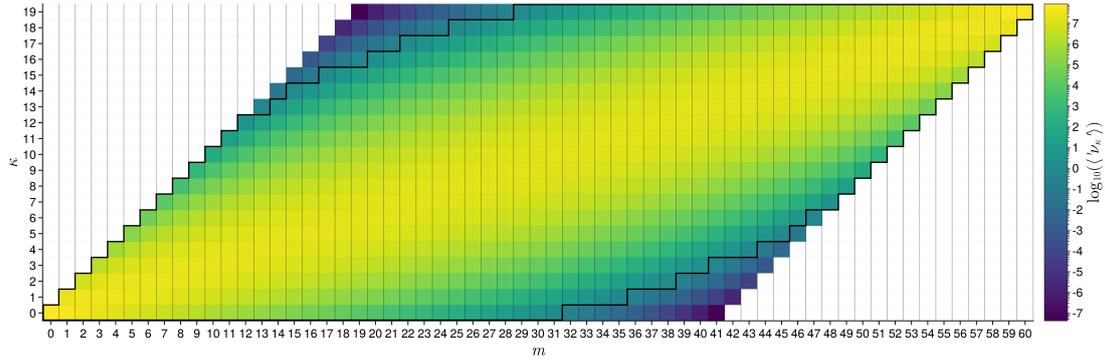
Moreover, using the approximation of the hypergeometric distribution by the normal distribution from equation 3.24, we get a simple approximation for the average double difference

$$\langle \varkappa \rangle \approx \log_{10}((12.2 \pm 0.25)N) / (\sigma_{(H)} \ln(10)). \quad (3.26)$$

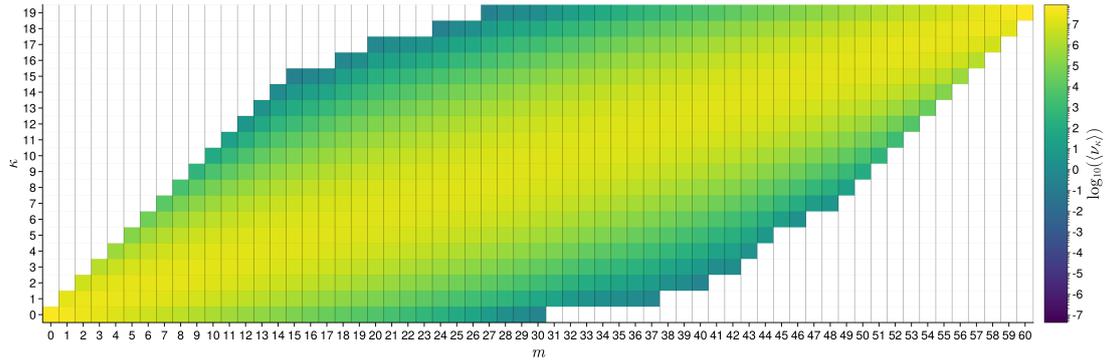
These two approximations enable us to estimate the expected number of Michaelis-Menten-like pieces. First, qualitatively based on $\langle \varkappa \rangle$ and then quantitatively based on $\langle q_\kappa \rangle$.

Figure 3.4 shows $\log_{10}(|\langle \varkappa \rangle|)$ for combinations of N and $\frac{m}{n}$. Note that $\langle \varkappa \rangle$ is always negative, symmetrical for m with respect to its maximum at $m = n/2$ and is anticorrelated with n .

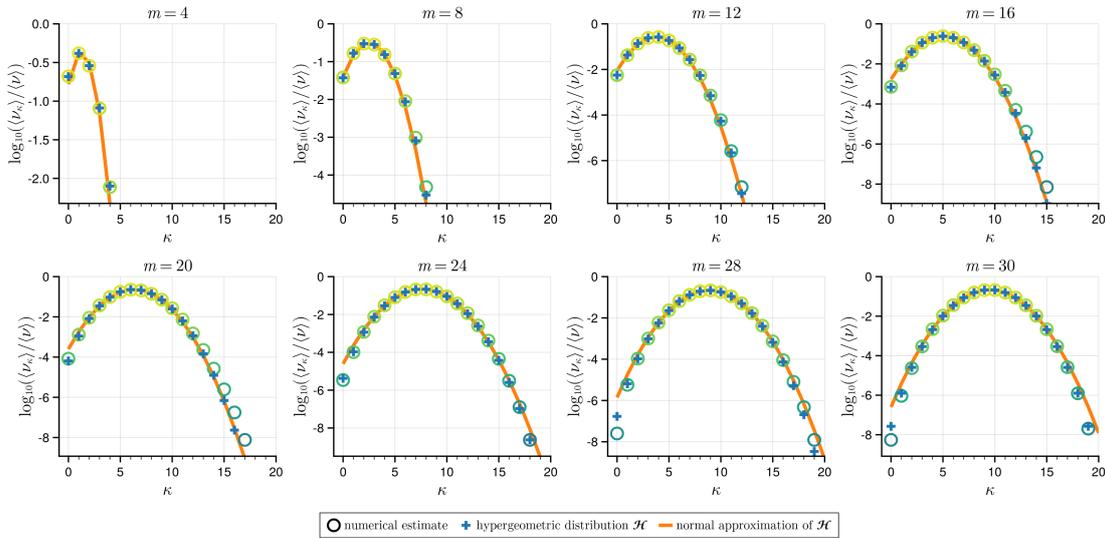
Re-interpreting the equation 3.16, which gives a condition for the coefficients



(A)



(B)



(C)

Figure 3.2 – (A) Approximation of number of rooted spanning trees with κ concentration-dependent rates ν_κ for Markov models with 20 states and vertex degree three (B) numerical estimate of (A) by equation 3.23 (C) cuts (A) and (B) along fixed m normalized by the total number of rooted spanning trees and its estimate, respectively. In orange the approximation of the hypergeometric distribution by normal distribution as given equation 3.24.

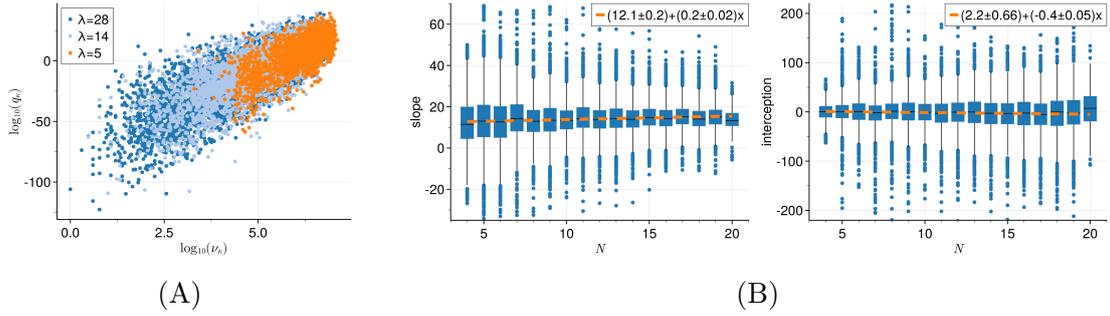


Figure 3.3 – (A) Coefficients q_κ as function of number of terms in sum in equation 3.7 on a double-log scale for $N = 19$ and $\kappa \in 5, 14, 28$ (color coded). (B) parameters of linear fit to (A) for N from 4 to 20, left slope and right y-intercept

based on the double difference on the logarithmic scale such that a rational function has only Michaelis-Menten-like pieces, in terms of averages, it is to reason that the more negative or less negative the average double difference $\langle \mathcal{Z} \rangle$ is, the higher or lower the fraction of Michaelis-Menten-like pieces is.

Obviously, the fraction of Michaelis-Menten-like pieces of a Markov model depends on the actual q_κ for which $\langle \mathcal{Z} \rangle$ provides only limited predictive power. In this context, note that for all of the investigated Markov models q_κ obeyed

$$\begin{aligned} q_\kappa &< q_{\kappa+1} \forall \kappa < \kappa' \\ q_{\kappa+1} &< q_\kappa \forall \kappa > \kappa' \end{aligned} \quad (3.27)$$

with $q'_\kappa = \max(q_\kappa)$ giving further insight into relation between different q_κ .

Note that multiplication of $\langle q_\kappa \rangle$ with a constant does not influence the double difference. Furthermore, note that according to our enumeration of rooted spanning trees, the curvature of q_d is larger than approximated by equation 3.26 for low and high κ (see deviations from the normal approximation in figure 3.2C). In particular, for the majority of m , i.e., $m \in [4, 20]$, the curvature is more negative for low and high κ . Because the q_κ approximate a parabola, the single difference determining the inflection point of Michaelis-Menten-like pieces, is larger for smaller and larger κ than for intermediate κ . Taking both of these observations together, we expect more Michaelis-Menten-like pieces as non-Michaelis-Menten-like pieces for values further away from c_0 .

Finally, using the estimate of $\langle q_\kappa \rangle$ in equation 3.25, we quantitatively estimate the number of Michaelis-Menten-like and non-Michaelis-Menten-like pieces. To this aim, we estimate q_κ by $\langle q_\kappa \rangle$ with white noise such that all q_κ obey observa-

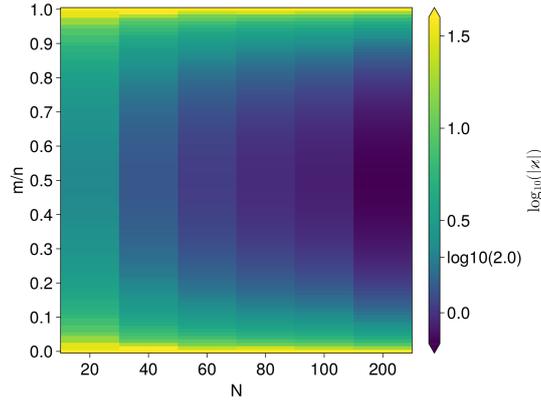


Figure 3.4 – Curvature of the normal approximation (color coded) to the hypergeometric distribution as function of the number of states N and the fraction of concentration-dependent transition rates m/n (equation 3.24)

tion 3.27. Then, we estimate number of Michaelis-Menten-like pieces and non-Michaelis-Menten-like pieces by calculating the fraction of all terms $q_\kappa c^\kappa$ over $q(c)$ for $c \in [10^{-27}, 10^{27}]$ and determining intervals on which two or more these fractions are larger than 10^{-3} .

The right most panel of figure 3.5 shows the fraction of Michaelis-Menten-like pieces among the total number of pieces in dependence on the number of states N and fraction of concentration-dependent rates m/n . For $m/n < 0.1$, occupation curves are primarily composed of Michaelis-Menten-like pieces. In general, the fraction of Michaelis-Menten-like pieces is above 0.4 for the majority of investigated combinations of N and m/n .

To better understand the overall high fraction of Michaelis-Menten-like pieces and its correlations with N and m/n , the two left panels show the number of Michaelis-Menten-like and non-Michaelis-Menten-like pieces. The number of non-Michaelis-Menten-like pieces is monomodal for fixed N and peaks at $m/n = 0.5$ and for constant m/n increases with N . The number of Michaelis-Menten-like pieces is bimodal for fixed N with its peaks at the absolute number of concentration-dependent rates $m \approx 20$ for all N , resulting its peaks to shift towards lower m/n values for larger N . In general the number of pieces remains fairly low, considering that for $N = 200$ and $m/n \approx 0.15$ the highest average number of pieces is around 15, although those Markov models can be considered very complex with 600 transition rates and 120 concentration-dependent rates, i.e., the maximum number of pieces is 199.

Note that this prediction is only based on the denominator coefficients q_κ . In general, $p_\kappa \approx \frac{q_\kappa}{N}$ and inclusion of p in the model should result in a reduction of pieces.

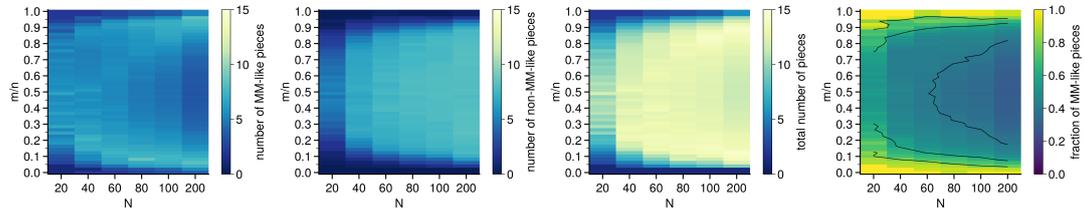


Figure 3.5 – From left to right (color coded): Estimation (by estimation of denominator coefficients) of the number of Michaelis-Menten-like pieces, of number of pieces that are neither constant nor Michaelis-Menten-like, of total number of pieces, and of the fraction of the number of Michaelis-Menten-like pieces over total number of pieces as function of the number of states N and the fraction of concentration-dependent transition rates m/n

3.3 Results

To better understand why an unexpectedly large fraction of Markov models of ABCE1 and of complex Markov-models for enzymatic reaction networks in general is observed, we have developed above a theoretical framework to predict the average fraction of Michaelis-Menten-like pieces in occupation curves for 3-regular Markov models with log-uniform distributed transition rate coefficients. In the following we test these predictions by explicit calculation of occupation curves $\pi_i(c)$ for Markov models with up to 200 states, vertex degree three, log-uniformly distributed random transition rates between 10^{-12} s^{-1} and 10^{12} s^{-1} , and 50 uniform-distributed fractions of concentration-dependent partial transitions rates m/n by algebraic means via equation 3.3.

To calculate the number of sigmoidal pieces in each $\pi_i(c)$, we dissected $\pi_i(c)$ into pieces based on its first derivative and fitted a Michaelis-Menten kinetics to the pieces. Pieces were classified to be Michaelis-Menten-like or not based on the square mean error of each fit and a threshold. The threshold was set to maximize the sensitivity, see methods. The proper functioning of dissection, fit, and classification was verified by visual inspection of random $\pi_i(c)$. The results were averaged over 600 to 2400 Markov models, the specific amount depending on N , which decreases for higher N . Note that, because of the choice of log-uniformly distribution for transition rate coefficients, the transition rate coefficients of a Markov model can span up to 24 orders of magnitude. To prevent false positive detection of pieces due to numerical inaccuracies, we only consider pieces with an amplitude of 10^{-3} of the total maximal amplitude of the corresponding occupation curve.

The main result of this numerical test is shown in figure 3.6. It depicts, color coded, the number of sigmoidal pieces in dependence on the number of states N and fraction m/n of concentration-dependent rates. The four panels show from left to right, the number of Michaelis-Menten-like pieces, the number of non-Michaelis-Menten-like pieces, the sum of both, i.e., the total number of pieces, and the fraction of Michaelis-Menten-like pieces and total number of pieces.

Overall, we observe a good qualitative agreement between our prediction in figure 3.5 and the explicit numerical determination of pieces in figure 3.6. As derived, a symmetry with respect to m/n and $1 - m/n$, i.e., $m/n = 1/2$, is seen. Moreover, the bimodality of the number of Michaelis-Menten-like pieces and the

monomodality of the number of non-Michaelis-Menten-like pieces was correctly predicted as well as the trend with N .

As expected, we overestimate the number of both types of pieces. There are two reasons for this. First, because in the numerical determination only pieces with an amplitude above a certain threshold are considered, which decreases the number of pieces observed. Second, the prediction was only based on the $\langle q_\kappa \rangle$ ignoring $\langle p_\kappa \rangle$ which results in a higher prediction of the number of pieces.

To check our predictions of the position of pieces, figure 3.7 shows the position of pieces for different N . The left panel shows the position of Michaelis-Menten-like pieces and the right panel the position of non-Michaelis-Menten-like pieces.

All distributions of position of Michaelis-Menten-like and non-Michaelis-Menten-like piece are monomodal and peak around $c = c_0$. One might consider the distributions as narrow, considering that the transition rates already span, for example for $\delta = 12, 24$ order of magnitude and might be arbitrary combined to determine the position of a piece.

Further, as predicted, the width of the distribution of position of the Michaelis-Menten-like pieces is wider than the one of non-Michaelis-Menten-like pieces.

Unexpectedly, the width of the distributions is independent on number of states N .

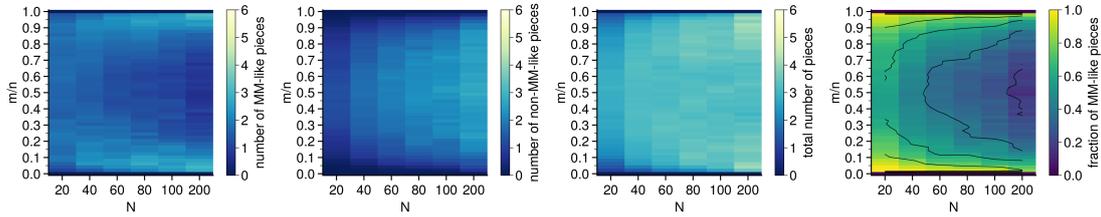


Figure 3.6 – Numerical estimates based on explicit calculation of the occupation curves of (from left to right and color coded) the number of Michaelis-Menten-like pieces, of number of pieces that are neither constant nor Michaelis-Menten-like, of total number of pieces, and of the fraction of the number of Michaelis-Menten-like pieces over total number of pieces as function of the number of states N and the fraction of concentration-dependent transition rates m/n

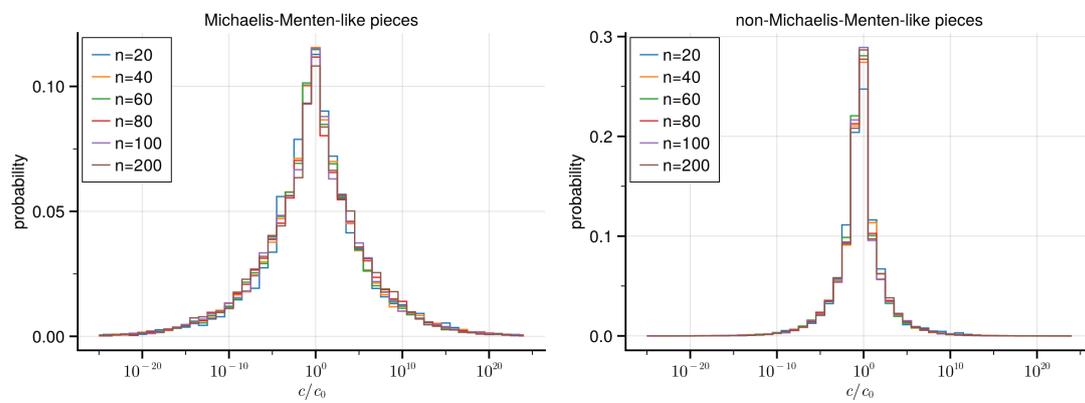


Figure 3.7 – Numerical estimate of the position in relation to the reference concentration c_0 separated by Michaelis-Menten-like (left) and non-Michaelis-Menten-like pieces (right). Color coded are the estimates for different values of N

3.4 Conclusion

The description of complex macromolecules such as proteins or protein interactions networks requires similarly complex Markov models. As reference, consider the rudimentary two-state Markov model of the Michaelis-Menten kinetics. In particular, these Markov models are drastically more complex than the, historically probably first, rudimentary two-state Markov model of a protein of the Michaelis-Menten kinetics. However, despite this increase in complexity, we observed that the occupation curves of Markov models describing ABCE1 from chapter 2 often exhibit Michaelis-Menten-like behavior, i.e., they are sigmoidal over a logarithmically scaled substrate concentration. This unexpected phenomena was also already observed by other in different Markov models. The prevalence of this phenomena raised the question to which extent the occurrence of Michaelis-Menten-like behavior in occupation curves is a general property of Markov models, and why this is the case.

To answer this question, we systematically computed occupation curves and determined the occurrences of Michaelis-Menten-like behavior of Markov models with 20 to 200 states where each state is connected to three other states, and random connections between their states, and random transition rate coefficients, and random selection of which of these are concentration-dependent.

We observed and showed that occupation curves are often indeed described in a piecewise manner, most of which are pieces of constant value and pieces that behave Michaelis-Menten-like. Further, we found that the number of these pieces is correlated with the number of concentration-dependent transition rate coefficients. Typically, the number of concentration-dependent transition rate coefficients in a Markov model is related to the number of binding sites of the described protein. This is consistent with the observation that occupation curves comprised of multiple pieces are observed for proteins with multiple binding sites, for example, F-ATPase,⁵² RecBCD,⁵⁴ and NSFs.⁵³

Note that of the occupation curves, because of experimental limitation, we only see a small interval compared to our numerical analysis. Thus, it can very well be that a process which follows the Michaelis-Menten in experiment does not necessarily require a Markov model with an occupation curve that exactly follows the Michaelis-Menten equation on the whole interval, increasing the space of Markov models that can be considered to explain the experimental observations.

Moreover, we showed that the fraction of pieces that behave Michaelis-Menten-like in each occupation curve decreases with the number of states and the number of concentration-dependent transitions. Strikingly, if less than 10% of all transition rate coefficients are concentration-dependent, which is typical for Markov models describing proteins, we observe that the occupation curves are comprised of more than 80% Michaelis-Menten-like pieces. Thus, it is not that Markov models with Michaelis-Menten-like behavior are particularly rare and thus the frequent occurrence implies no evolutionary benefit to Michaelis-Menten-like behaviour. Quite the opposite, the high occurrence of Michaelis-Menten-like behaviour is an expected outcome in a situation without any selection pressure.

Next, we asked why Michaelis-Menten-like pieces appear so frequently in the occupation curves of sparse Markov models with a low number of concentration-dependent transition rates. To this end, we used a version of the matrix tree theorem that yields an analytical expression for the occupation curves in terms of a rational function with the substrate concentration as variable and the coefficients of both polynomials, nominator and denominator, a sum over products of transition rate coefficients. We derived a condition for rational functions to only comprise of Michaelis-Menten-like pieces on the double-differences of consecutive coefficients on a logarithmic scale.

To evaluate whether this condition is met by the coefficients of Markov model occupation curves, we estimated the average coefficients for Markov models with above-mentioned random properties. Based on the resulting average double-differences of these coefficients, we successfully predicted the fraction of Michaelis-Menten-like pieces in dependence on the number of states and concentration-dependent transitions.

Furthermore, we showed that the average normalized coefficients, and thus the double differences, can be estimated from the hypergeometric distribution and a power law. The closer the average double difference is to fulfill the above mentioned conditions, the more likely it is for the double difference of Markov models to fulfill the condition. Thus, the occurrence of Michaelis-Menten-like pieces for sparse Markov models can be traced back to the combinatorial problem described by the hypergeometric distribution. In particular, the high fraction of Michaelis-Menten-like pieces with a low number of concentration-dependent transition rates can be explained by this combinatorial problem having only a few, very differently probable, outcomes resulting in an average double-difference that is close to even

fulfills above condition.

The present study was motivated by observation of occupation curves of Markov models where the transition rate coefficients depend on substrate concentrations. The mathematical framework, however, is much more general, and also applies to every linear dependency of the transition rate coefficients, for example, on salt and proton concentrations. Even photosensitive transition rate coefficients, for which a quadratic dependence would be expected, might be considered after a square root transformation of the dependency.

The limitation to 3-regular Markov models seems to be quite restrictive at first glance, but indeed the majority of states of Markov models describing proteins have a vertex degree of three, and the remaining ones typically vertex degree two or four. Therefore, it might be interesting to see whether the results also hold for Markov models with an average vertex degree between two and four to also include, for example, the 2D-grid Markov models used to describe fluctuating enzymes.⁴⁸

3.5 Methods

Generating random Markov models with given vertex degree

Algorithm 1 Generate random Markov model with given vertex degree

Require: N number of states, vd vertex degree
initiate empty adjacency matrix A
for $i \in 1, \dots, N - 1$ **do**
 determine $nDraw$, the difference between vd and the number of adjacent vertices of state i
 determine I the set of vertices with less than vd adjacent vertices
 sample $nDraw$ elements from I and update A
end for

For an even number of states N every vertex has the same vertex degree vd . For an odd number of states N and an odd vertex degree vd all but one state have vertex degree vd and the one state has any even vertex degree. Algorithm 1 assigns this one state the vertex degree $vd - 1$.

Calculating rooted spanning trees The spanning trees were calculated using the graph contraction approach by Winter,¹⁰⁴ which was determined to be the fastest from implementation perspective, although not having optimum time complexity, in an extensive study of algorithms for generating all possible spanning trees.¹⁰⁵ An own Julia implementation was developed.

Rooted spanning trees were calculating by the depth-first search algorithm.

Generating Markov models with detailed balance Given a Markov model we wanted to choose transition rate coefficients $k \in [10^{-d}, 10^d]$ such that the Markov model satisfies detailed balance. To this end, we used that the activation free energies ΔG^\ddagger can be estimated from the transition rate coefficients by the Eyring equation

$$k_{ts} = \frac{\kappa k_B T}{h} e^{-\frac{\Delta G_{ts}^\ddagger}{k_B T}} \quad (3.28)$$

with the Boltzmann constant k_B , Planck constant h and transmission coefficient κ correcting for barrier recrossing.¹⁰⁶ We set $k_B T = 1$ and $\frac{\kappa k_B T}{h} = 10^{-12}$. In the free energy framework, detailed balance is satisfied if the free energy difference along any closed cycle in the Markov model is zero. To this end, we set the absolute

free energy of one Markov model state to zero, and drew the other uniformly such that $G_i \in [0, 2d]$. In a second step, for each connection between states an absolute activation free energy was chosen according to

$$G_{ts}^\ddagger \sim \mathcal{U}(2d - \max(G_t, G_s), 2d) \quad (3.29)$$

with \mathcal{U} a uniform distribution, which ensures that $k_{ts} \in [10^{-d}, 10^d]$. Then

$$\Delta G_{ts}^\ddagger = \begin{cases} G_{ts}^\ddagger - G_s & \text{if } s > t, \\ G_{ts}^\ddagger - G_t & \text{if } s < t. \end{cases} \quad (3.30)$$

Selecting which transition rate coefficients are concentration-dependent For the lowest, non-zero fraction of concentration-dependent partial transitions rates m_0/n , m_0 transition rates were selected at random to be concentration-dependent, and, subsequently, for each larger m_{i+1}/n additional $m_{i+1} - m_i$ transition were selected at random to be concentration-dependent. Thus, the concentration-dependent rates were chosen such that the set of concentration-dependent rates for Λ_i is included in the set of concentration-dependent rates of Λ_{i+1} for $|\Lambda_i| < |\Lambda_{i+1}|$.

Counting Michaelis-Menten-like transitions We calculated $\pi(c)$ via eigenvalue decomposition using the QR-algorithm for 1000 logarithmically distributed values of $c \in [10^{-40}, 10^{60}]$. The eigenvalue decomposition of \mathbf{Q} to calculate the steady state probabilities was done with 106 bits precision, all other calculations were done with 64 bit precision. The occupation curves $p_i(c)$ were then normalized to range from 0 to 1, the first derivative $\frac{d\pi_i(c)}{d\log_{10}(c)}$ was calculated by a finite difference approach with $dc = 1$ and an cutoff to set values near zero to zero, and $\pi(c)$ dissected into pieces based on the position and width of the local extrema of $\frac{d\pi_i(c)}{d\log_{10}(c)}$. Then each piece was fit to a Michaelis-Menten kinetics with offset and negative nominator, allowing for a decreasing instead of increasing Michaelis-Menten kinetics, and the piece classified to be Michaelis-Menten like based on whether the standard mean error between fit and piece of the occupation curve exceeds a threshold. Thus, the results depends on the choice of the threshold. To this end, figure 3.8 shows how the fraction of Michaelis-Menten-like pieces averaged over states and fractions of concentration-dependent transition rate coefficients. For all plots in this study, the threshold was selected to be $5 \cdot 10^{-4}$ to ensure maximal sensibility of the analysis.

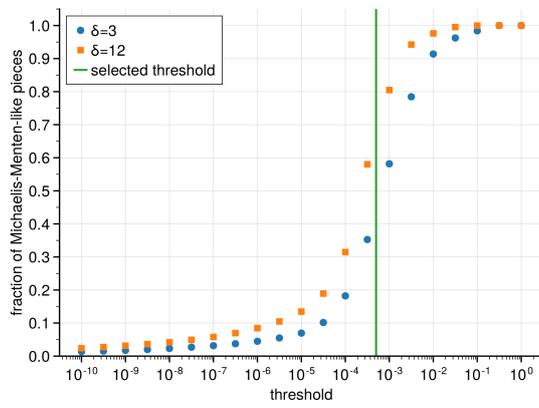


Figure 3.8 – Dependence of the fraction of Michaelis-Menten pieces on the threshold used to classify occupation curve pieces as Michaelis-Menten-like or not.

3.6 Supplementary Information

Products In this section, we will derive the distribution Ξ of the products $\xi := \prod_{j,k \in R}^{n-1} k_{jk}$. In our case the rates k_{ij} are log-uniformly distributed between $10^{-\delta}$ and 10^δ , i.e., $k_{ij} \sim F := \mathcal{LU}(-\delta, \delta)$. Each Markov model has one rate per directed edge drawn from F resulting in set \mathbf{k} with nvd elements. Even if the rates are drawn such that the Markov model obeys detailed balance, the assumption of a uniform distribution seems to hold. For each product ξ are $n - 1$ rates drawn without replacement from the set \mathbf{k} , with $n - 1$ being the number of edges in a spanning tree of graph with n edges.

To derive the distribution Ξ , we will look at the logarithm of ξ

$$\zeta := \log(\xi) = \log \left(\prod_{j,k \in R}^{n-1} k_{jk} \right) = \sum_{j,k \in R}^{n-1} \log(k_{jk}) = \sum_{j,k \in R}^{n-1} g_{jk} \quad (3.31)$$

with $g_{ij} \sim G := \mathcal{U}(-\delta, \delta)$ now uniform distributed on $[-\delta, \delta]$. The g_{ij} can be interpreted as free energy barriers.

Acknowledging that ζ is identical to the scaled sample mean $\overline{g_{ij_{n-1}}}$ of $n - 1$ free energy barriers ζ can be rewritten as

$$\zeta = \sum_{j,k \in R}^{n-1} g_{jk} = \frac{(n-1)}{(n-1)} \cdot \sum_{j,k \in R}^{n-1} g_{jk} = (n-1) \overline{g_{ij_{n-1}}}. \quad (3.32)$$

According to the central limit theorem, the sample mean is normally distributed with mean and standard error of the mean of the sampled population, which in our case is the set \mathbf{g} . Further, the family of normal distributions is closed under linear transformations, i.e., if $X \sim N(\mu, \sigma)$ and $Y = aX + b$ then $Y \sim N(a\mu + b, |a|\sigma)$. Thus

$$\overline{g_{ij_{n-1}}} \sim \mathcal{N} \left(\mu_{\mathbf{g}}, \frac{\sigma_{\mathbf{g}}}{\sqrt{n-1}} \right) \quad (3.33)$$

$$\zeta \sim Z = \mathcal{N} \left((n-1)\mu_{\mathbf{g}}, \sqrt{n-1}\sigma_{\mathbf{g}} \right) . \quad (3.34)$$

The set \mathbf{g} was a sample of $G := \mathcal{U}(-\delta, \delta)$ with

$$\mu_G = \frac{a+b}{2} = 0 \tag{3.35}$$

$$\sigma_G = \sqrt{\frac{(b-a)^2}{12}} = \frac{2\delta}{\sqrt{12}} \tag{3.36}$$

and thus

$$\mu_{\mathbf{g}} = \mathcal{N}\left(\mu_G, \frac{\sigma_G}{\sqrt{nvd}}\right) \tag{3.37}$$

$$\sigma_{\mathbf{g}} \approx \sigma_G. \tag{3.38}$$

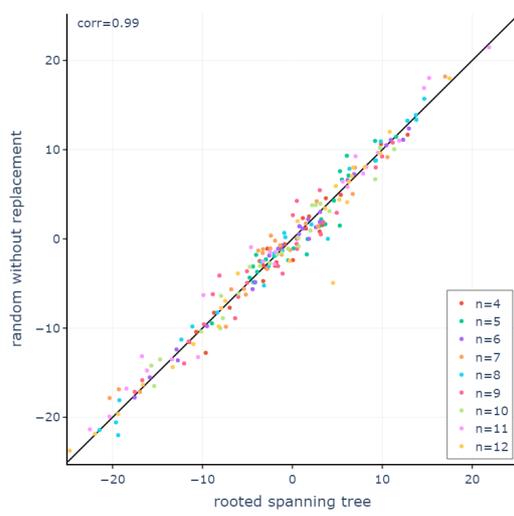
In conclusion, the products ξ are lognormal-distributed with $\Xi \sim \mathcal{LN}(\mu_Z, \sigma_Z)$

$$\mu_Z \sim \mathcal{N}\left(0, \frac{2\delta(n-1)}{\sqrt{12nvd}}\right) \tag{3.39}$$

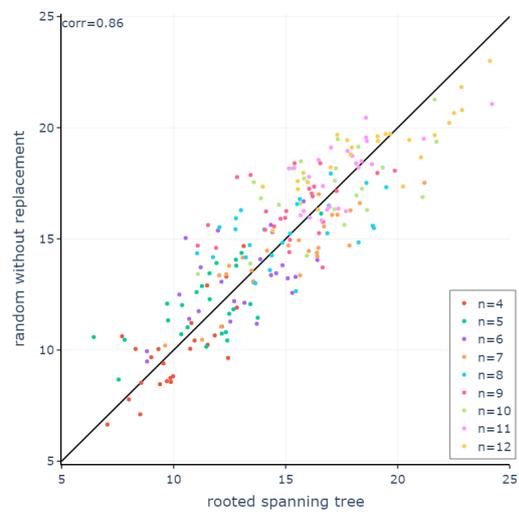
and

$$\sigma_Z = \frac{2\delta\sqrt{n-1}}{\sqrt{12}}. \tag{3.40}$$

In this derivation we assumed that the $n-1$ factors of each product are randomly drawn without replacement from of \mathbf{g} . It is not clear whether this assumption holds because correlations between the edges of different rooted spanning trees and thereby between different products might exist. To check the validity of the assumption, we calculated μ_Z and σ_Z for the same sets \mathbf{g} once with the factors randomly drawn without replacement and once according to the rooted spanning trees. Figure 3.9 shows the comparison between true product and the approximation via random draws without replacement. We consider the correlation good enough, such that the assumption is valid.



(A)



(B)

Figure 3.9 – (A) mean and (B) standard deviation of Z , diagonal black lines are references of perfect agreement

4 Conclusion

The work reported in this thesis was motivated by a puzzling observation regarding the kinetics of ABCE1, a protein that plays an important role in protein biosynthesis in eukaryotes and archaea by separating the ribosomal subunits at the end of translation.⁶³ It has two almost identical subunits with two highly symmetrical nucleotide binding sites at their interface, each of which hydrolyzes ATP. When ATP hydrolysis is partially blocked in one binding site, the ATP turnover rate is reduced twofold, as expected from the symmetry of the protein. In stark contrast, when the other binding site is partially blocked a stunning tenfold increase in ATP turnover rate is observed.⁴⁷

The prevailing explanation for this unexpected asymmetric kinetics is that, despite the symmetry between the binding sites, there is asymmetric allostery between the two binding sites, i.e., occupation of one binding site affects the other, but not vice versa. In addition, a direct communication between the two binding sites mediated by a chain of conformational changes is hypothesized.⁵⁵

One of the main results of this work is that we show here that such asymmetric allostery by direct communication between the binding sites is not required for the observed asymmetric kinetics. Therefore, we determined Markov models of ABCE1 by Bayesian inference that exclude any such allostery between the binding sites and are in agreement with the measured kinetics within the experimental error. We found that the asymmetric kinetics can be caused by a redistribution of the steady-state population with respect to the wild type. This redistribution is caused by the change in the free energy landscape of the proteins due to the partial blocking of ATP hydrolysis, i.e. an increase in the free energy barrier for ATP catalysis in the corresponding binding site.

The hypothesis of an asymmetric allostery between the binding domains is based, besides the observed asymmetric kinetics of the mutants, on contact maps of homologues of ABCE1 showing interactions between both binding sites,¹⁰⁷ which vary with the occupation of the binding sites.⁶⁸⁻⁷⁰ However, there is no evi-

dence that sheds light on the clausal relationship between the various interactions and the occupations. In the absence of such evidence for or against asymmetric allostery between the binding sites, one might rely on Occam's razor to judge the plausibility of such communication. Under this premise, the fact that Markov models without asymmetric allostery have fewer unknowns than those with asymmetric allostery and thus can be considered to provide a simpler explanation, the case of no asymmetric allostery between binding sites is to be preferred.

In a wider sense, the redistribution of steady state population can be considered a case of ensemble allostery, which makes the very general statement that all allostery can be understood by changes of the protein energy landscape⁸⁷ and need not rely on an allostery transferred by conformational changes.^{108,109} In this context, one might interpret the point mutations as ligands, which directly and non-allysterically change the ATP catalysis rates of the binding sites they bind to, but also, allysterically, change the ATP turnover rate in the other binding site. In this interpretation, the possible effects of allostery are expanded from a change of binding affinity or kinetics to include the turnover rate of the affected binding site too, as the ATP turnover rate of the other binding site in ABCE1 increases not due to a change in the first two mention allysteric effects but due to a increased population of states of this binding site.

Thus, on a more general notion, we showed that top-down Markov models are a suitable method to test whether allostery is at all required to explain a given set of experimental data. In a similar way, Markov models can be used on a more granular level to test how many and which transitions must be affected by allostery. For example, whether binding or catalysis is affected, and whether kinetics or thermodynamics or both are affected.

Because the system of Markov model transition rates and constraints given by the measured kinetics is underdetermined, there are many Markov models that are within error of the experiment and thus span a subspace in the space of all possible Markov models. Maximum likelihood approaches or optimization, which are typically used to determine individual Markov models in this subspace,^{35,44,45} naturally provide limited information about the subspace and, in particular, lack an estimate of uncertainty. We have shown that using a Bayesian approach together with extensive sampling of the subspace provides an assignment of probability density to each Markov model, allowing more rigorous statements about the subspace in terms of means and uncertainties. Note that a similar Bayesian approach,

but for transporters, is currently being developed.¹¹⁰

Any Bayesian approach relies on the subjective choice of an appropriate prior, and so does our Bayesian approach to top-down Markov models. We think that a logarithmic uniform prior on the transition rates is a reasonable choice as an uninformed prior over the two alternatives of a uniform prior on the transition rate and a uniform prior on the inverse of the transition rate, i.e., the average transition time for two reasons. First, if only one of the transition rates or average transition times can be uniformly distributed, both are equally distributed if a logarithmic uniform distribution is chosen for one, resulting in internal consistency. Second, systems defined by free energies naturally satisfy detailed equilibrium, a condition that a top-down Markov model should also be able to satisfy. The choice of a logarithmic uniform prior on the transition rate coefficients is equivalent to a uniform prior on the free energies, and thus facilitates the definition of top-down Markov models in terms of free energies.

Finally, by classifying the Markov models of this subspace into groups, e.g., Markov models in favor or against different hypothesized mechanisms, the Bayesian approach allows to rank the different groups by the probability that the corresponding mechanism is the one responsible for the given data in the first place. This can provide unbiased evidence for one hypothesis or the other and, given limited experimental resources, help decide which experiments to perform to gain the most knowledge.

The top-down Markov models of ABCE1 in chapter 2 are just one example in a series of increasingly complex Markov models describing proteins. However, the results of the probable first description of protein function with a two-state Markov model leading to the Michaelis-Menten equation are still very relevant today.^{22,26}

Unexpectedly, the sigmoid curve of the Michaelis-Menten equation occurs quite frequently in the occupation curves (the dependence of steady-state probabilities on substrate concentrations) of Markov models of ABCE1 and other complex Markov models describing proteins.⁵⁰ Other studies investigating the occurrence of Michaelis-Menten-like behavior have focused on cases where the occupation curves are described by a single sigmoid curve, and have used purely analytical means to derive the conditions under which this type of behavior occurs.^{49,50} In chapter 3, we extend this view and offer a new perspective on the ubiquity

of Michaelis-Menten behavior in occupation-curves by not only investigating the case where the occupation-curve is described by a single sigmoid curve, but also including cases where only pieces of the occupation-curve are sigmoid curves.

To this end, we calculated the occupation curves and quantified the occurrence of Michaelis-Menten-like behavior for Markov models as a function of the number of states and substrate-dependent transitions. We showed that Markov models with parameters typical for top-down Markov models of proteins have a high probability of having occupation curves that exhibit Michaelis-Menten-like behavior at least on pieces. In particular, low-state Markov models with few substrate-dependent transitions are prone to show Michaelis-Menten-like behavior in the occupation curves.

One might ask whether the frequent observation of Michaelis-Menten-like behavior in top-down Markov models of proteins indicates a selection pressure toward Michaelis-Menten-like behavior of the described proteins.^{50,52–54} However, we have shown that Michaelis-Menten-like behavior is by no means rare when the protein is well described by a Markov model with few substrate-dependent transitions. On the contrary, Michaelis-Menten-like behavior is to be expected.

Using a semi-numerical analysis of a version of the matrix tree theorem, we traced the occurrence of Michaelis-Menten-like behavior as an average over many Markov models back to properties of average Markov models, where the averages are over Markov models with random transition rates, state numbers and connections, and where a priori it was not evident that such commutation of averages would yield any insights. In this way, we showed that the Michaelis-Menten-like behavior arises from the combinatorial problem that is described by the hypergeometric distribution, i.e., to draw κ special elements if a total of $N - 1$ elements drawn from a set of n with m many special elements are drawn without replacement. The occurrence of Michaelis-Menten-like behavior in an occupation curve depends then on the curvature on the logarithmic scale of the hypergeometric distribution, which becomes smaller for m as the distribution gets stretched as more κ become possible. Thus, the occurrence of Michaelis-Menten-like behavior was traced back to very fundamental combinatorial properties of Markov models.

We investigated properties averaged over Markov models with a low level of connection between states as commonly used to describe proteins. This approach is in stark contrast to analyzing the properties of a particular Markov model, or a finite set of very similar Markov models, when one is interested in the function

of a particular protein. We see the opportunity that a broader analysis across various Markov models might offer insights above the level of individual proteins by identification of common properties.

Bibliography

- [1] Whitford, D. *Protein structure & function: Structure and Function*; Wiley: New York and Chichester, 2003.
- [2] Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Murihead, H.; Will, G.; North, A. C. *Nature* **1960**, *185*, 416–422.
- [3] Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. *Nature* **1958**, *181*, 662–666.
- [4] Mannige, R. V. *Proteomes* **2014**, *2*, 128–153.
- [5] Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. *Biochemistry* **1975**, *14*, 5355–5373.
- [6] McCammon, J. A. *Reports on Progress in Physics* **1984**, *47*, 1–46.
- [7] Ansari, A.; Berendzen, J.; Bowne, S. F.; Frauenfelder, H.; Iben, I. E.; Sauke, T. B.; Shyamsunder, E.; Young, R. D. *Proceedings of the National Academy of Sciences* **1985**, *82*, 5000–5004.
- [8] Hucho, F.; Weise, C. *Angewandte Chemie International Edition* **2001**, *40*, 3100–3116.
- [9] Hofmann, S.; Janulienė, D.; Mehdipour, A. R.; Thomas, C.; Stefan, E.; Brüchert, S.; Kuhn, B. T.; Geertsma, E. R.; Hummer, G.; Tampé, R.; Moeller, A. *Nature* **2019**, *571*, 580–583.
- [10] Gray, A.; Harlen, O. G.; Harris, S. A.; Khalid, S.; Leung, Y. M.; Lonsdale, R.; Mulholland, A. J.; Pearson, A. R.; Read, D. J.; Richardson, R. A. *Acta crystallographica. Section D, Biological crystallography* **2015**, *71*, 162–172.
- [11] Karplus, M.; McCammon, J. A. *Nature structural biology* **2002**, *9*, 646–652.

- [12] Hollingsworth, S. A.; Dror, R. O. *Neuron* **2018**, *99*, 1129–1143.
- [13] Brandman, R.; Brandman, Y.; Pande, V. S. *PloS one* **2012**, *7*, e29377.
- [14] Kutzner, C.; Páll, S.; Fechner, M.; Esztermann, A.; de Groot, B. L.; Grubmüller, H. *Journal of computational chemistry* **2019**, *40*, 2418–2431.
- [15] Nüske, F.; Keller, B. G.; Pérez-Hernández, G.; Mey, A. S. J. S.; Noé, F. *Journal of chemical theory and computation* **2014**, *10*, 1739–1752.
- [16] de Groot, B. L. et al, *Journal of Molecular Biology* **2001**, *309*, 299–313.
- [17] Pande, V. S.; Beauchamp, K.; Bowman, G. R. *Methods* **2010**, *52*, 99–105.
- [18] Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. *The Journal of chemical physics* **2011**, *134*, 174105.
- [19] Meyn, S. P.; Tweedie, R. L. *Markov chains and stochastic stability*, 2nd ed.; Cambridge University Press: Cambridge, 2009.
- [20] Noé, F. *The Journal of chemical physics* **2008**, *128*, 244103.
- [21] Chodera, J. D.; Noé, F. *Current Opinion in Structural Biology* **2014**, *25*, 135–144.
- [22] Michaelis, L.; Menten, M. L. *Biochemische Zeitschrift* **1913**, *49*, 333–369.
- [23] Briggs, G. E.; Haldane, J. B. S. *The Biochemical Journal* **1925**, *19*, 338–339.
- [24] Horn, F.; Jackson, R. *Archive for Rational Mechanics and Analysis* **1972**, *47*, 81–116.
- [25] Feinberg, M. *Archive for Rational Mechanics and Analysis* **1972**, *49*, 187–194.
- [26] Haldane, J. B. S. *Enzymes*; Longmans, Green and Co.: UK, 1930.
- [27] Gunawardena, J. *The FEBS journal* **2014**, *281*, 473–488.
- [28] Segel, L. A.; Slemrod, M. *SIAM Review* **1989**, *31*, 446–477.
- [29] Schnell, S.; Maini, P. K. *Comments on Theoretical Biology* **2003**, *8*, 169–187.

- [30] Lee, C. H.; Othmer, H. G. *Journal of mathematical biology* **2010**, *60*, 387–450.
- [31] Sharma, A. K.; Chowdhury, D. *Physical review. E, Statistical, nonlinear, and soft matter physics* **2010**, *82*, 031912.
- [32] Dutta, A.; Chowdhury, D. *Bulletin of mathematical biology* **2017**, *79*, 1005–1027.
- [33] Ong, K. M.; Blackford, J. A.; Kagan, B. L.; Simons, S. S.; Chow, C. C. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 7107–7112.
- [34] Dangkulwanich, M.; Ishibashi, T.; Liu, S.; Kireeva, M. L.; Lubkowska, L.; Kashlev, M.; Bustamante, C. J.; Zhuang, X. *eLIFE* **2013**, *2*, e00971.
- [35] Bernardo A. Mello, Wenlin Pan, Gerald L. Hazelbauer, Yuhai Tu, *PLoS computational biology* **2018**, 1–16.
- [36] Maes, C.; van Wieren, M. H. *Journal of Statistical Physics* **2003**, *112*, 329–355.
- [37] Greulich, P.; Garai, A.; Nishinari, K.; Schadschneider, A.; Chowdhury, D. *Physical review. E, Statistical, nonlinear, and soft matter physics* **2007**, *75*, 041905.
- [38] Chemla, Y. R.; Moffitt, J. R.; Bustamante, C. *The journal of physical chemistry. B* **2008**, *112*, 6025–6044.
- [39] Garai, A.; Chowdhury, D. *EPL (Europhysics Letters)* **2011**, *93*, 58004.
- [40] Bierbaum, V.; Lipowsky, R. *Biophysical journal* **2011**, *100*, 1747–1755.
- [41] Gaspard, P.; Gerritsma, E. *Journal of Theoretical Biology* **2007**, *247*, 672–686.
- [42] Shu, Y.-G.; Lai, P.-Y. *The journal of physical chemistry. B* **2008**, *112*, 13453–13459.
- [43] Gerritsma, E.; Gaspard, P. *Biophysical Reviews and Letters* **2010**, *5*, 163–208.

- [44] Gajewski, J.; Buelens, F.; Serdjukow, S.; Janßen, M.; Cortina, N.; Grubmüller, H.; Grininger, M. *Nature chemical biology* **2017**, *13*, 363–365.
- [45] Mescam, M.; Vinnakota, K. C.; Beard, D. A. *The Journal of biological chemistry* **2011**, *286*, 21100–21109.
- [46] Johnson, K. A. *Methods in enzymology* **2009**, *467*, 601–626.
- [47] Barthelme, D.; Dinkelaker, S.; Albers, S.-V.; Londei, P.; Ermler, U.; Tampé, R. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 3228–3233.
- [48] English, B. P.; Min, W.; van Oijen, A. M.; Lee, K. T.; Luo, G.; Sun, H.; Cherayil, B. J.; Kou, S. C.; Xie, X. S. *Nature chemical biology* **2006**, *2*, 87–94.
- [49] Barel, I.; Brown, F. L. H. *The Journal of chemical physics* **2017**, *146*, 014101.
- [50] Wong, F.; Dutta, A.; Chowdhury, D.; Gunawardena, J. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, *115*, 9738–9743.
- [51] Wu, J.; Cao, J. In *Single-molecule biophysics*; Komatsuzaki, T., Ed.; Advances in Chemical Physics; Wiley: Hoboken, N.J., 2012; pp 329–365.
- [52] Weber, J.; Wilke-Mounts, S.; Lee, R. S.; Grell, E.; Senior, A. E. *Journal of Biological Chemistry* **1993**, *268*, 20126–20133.
- [53] Matveeva, E. A.; He, P.; Whiteheart, S. W. *Journal of Biological Chemistry* **1997**, *272*, 26413–26418.
- [54] Zananiri, R.; Mangapuram Venkata, S.; Gaydar, V.; Yahalom, D.; Malik, O.; Rudnizky, S.; Kleifeld, O.; Kaplan, A.; Henn, A. *Nature communications* **2022**, *13*, 1806.
- [55] Nürenberg-Goloub, E.; Heinemann, H.; Gerovac, M.; Tampé, R. *Life Science Alliance* **2018**, *1*, 1–12.
- [56] Gouridis, G.; Hetzert, B.; Kiosze-Becker, K.; de Boer, M.; Heinemann, H.; Nürenberg-Goloub, E.; Cordes, T.; Tampé, R. *Cell reports* **2019**, *28*, 723–734.e6.

- [57] Karcher, A.; Büttner, K.; Märten, B.; Jansen, R.-P.; Hopfner, K.-P. *Structure (London, England : 1993)* **2005**, *13*, 649–659.
- [58] Karcher, A.; Schele, A.; Hopfner, K.-P. *The Journal of biological chemistry* **2008**, *283*, 7962–7971.
- [59] Becker, T.; Franckenberg, S.; Wickles, S.; Shoemaker, C. J.; Anger, A. M.; Armache, J.-P.; Sieber, H.; Ungewickell, C.; Berninghausen, O.; Daberkow, I.; Karcher, A.; Thomm, M.; Hopfner, K.-P.; Green, R.; Beckmann, R. *Nature* **2012**, *482*, 501–506.
- [60] Preis, A.; Heuer, A.; Barrio-Garcia, C.; Hauser, A.; Eyler, D. E.; Berninghausen, O.; Green, R.; Becker, T.; Beckmann, R. *Cell reports* **2014**, *8*, 59–65.
- [61] Kiosze-Becker, K.; Ori, A.; Gerovac, M.; Heuer, A.; Nürenberg-Goloub, E.; Rashid, U. J.; Becker, T.; Beckmann, R.; Beck, M.; Tampé, R. *Nature communications* **2016**, *7*, 13248.
- [62] Heuer, A.; Gerovac, M.; Schmidt, C.; Trowitzsch, S.; Preis, A.; Kötter, P.; Berninghausen, O.; Becker, T.; Beckmann, R.; Tampé, R. *Nature structural & molecular biology* **2017**, *24*, 453–460.
- [63] Nürenberg-Goloub, E.; Kratzat, H.; Heinemann, H.; Heuer, A.; Kötter, P.; Berninghausen, O.; Becker, T.; Tampé, R.; Beckmann, R. *The EMBO journal* **2020**, *39*, e103788.
- [64] Oldham, M. L.; Chen, J. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 15152–15156.
- [65] Stockner, T.; Gradisch, R.; Schmitt, L. *FEBS letters* **2020**, *594*, 3815–3838.
- [66] Mancera-Martínez, E.; Brito Querido, J.; Valasek, L. S.; Simonetti, A.; Hashem, Y. *RNA biology* **2017**, 1–7.
- [67] Urbatsch, I. L.; Julien, M.; Carrier, I.; Rousseau, M. E.; Cayrol, R.; Gros, P. *Biochemistry* **2000**, *39*, 14138–14149.
- [68] Grossmann, N.; Vakkasoglu, A. S.; Hulpke, S.; Abele, R.; Gaudet, R.; Tampé, R. *Nature communications* **2014**, *5*.

- [69] Hohl, M.; Hürlimann, L. M.; Böhm, S.; Schöppe, J.; Grütter, M. G.; Bordignon, E.; Seeger, M. A. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, 11025–11030.
- [70] Jones, P. M.; George, A. M. *The journal of physical chemistry. A* **2012**, *116*, 3004–3013.
- [71] de La Rosa, M. B.; Nelson, S. W. *The Journal of biological chemistry* **2011**, *286*, 26258–26266.
- [72] Schöffner, M.; Grubmüller, H.
- [73] Orelle, C.; Dalmas, O.; Gros, P.; Pietro, A. D.; Jault, J.-M. *The Journal of biological chemistry* **2003**, *278*, 47002–47008.
- [74] Chiraniya, A.; Finkelstein, J.; O'Donnell, M.; Bloom, L. B. *Genes* **2013**, *4*, 134–151.
- [75] Feinberg, M. **1989**, *44*, 1819–1827.
- [76] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *The Journal of chemical physics* **1953**, *21*, 1087–1092.
- [77] Hastings, W. K. *Biometrika* **1970**, *57*, 97–109.
- [78] Mallipeddi, R.; Suganthan, P. N.; Pan, Q. K.; Tasgetiren, M. F. *Applied Soft Computing* **2011**, *11*, 1679–1696.
- [79] Vehtari, A.; Gelman, A.; Simpson, D.; Carpenter, B.; Bürkner, P.-C. *Bayesian Analysis* **2020**,
- [80] Fetsch, E. E.; Davidson, L. A. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 9685–9690.
- [81] Prieß, M.; Göddeke, H.; Groenhof, G.; Schäfer, L. V. *ACS central science* **2018**, *4*, 1334–1343.
- [82] Oldham, M. L.; Chen, J. *Science* **2011**, *332*, 1202–1205.
- [83] Higgins, C. F.; Linton, K. J. *Nature structural & molecular biology* **2004**, *11*, 918–926.

- [84] Jones, P. M.; George, A. M. *Quarterly reviews of biophysics* **2014**, *47*, 189–220.
- [85] Schäffner, M.; Grubmüller, H. Asymmetric turnover rates of ABCE1 explained by Markov model. master thesis.
- [86] Gillespie, D. T. *Annual review of physical chemistry* **2007**, *58*, 35–55.
- [87] Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J. *Nature* **2014**, *508*, 331–339.
- [88] Braess, D. *Unternehmensforschung Operations Research - Recherche Opérationnelle* **1968**, *12*, 258–268.
- [89] Miller, W. G.; Alberty, R. A. *Journal of the American Chemical Society* **1958**, *80*, 5146–5151.
- [90] Anderson, D. F.; Kurtz, T. G. In *Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology*; Koepl, H., Setti, G., Di Bernardo, M., Densmore, D., Eds.; Springer New York: New York, NY, 2011; pp 3–42.
- [91] Altman, C.; King, E. L. *The Journal of chemical physics* **1956**, *60*, 1375–1378.
- [92] Cleland, W. W. *Biochemistry* **1975**, *14*, 3220–3224.
- [93] Gu, S.; Silva, D.-A.; Meng, L.; Yue, A.; Huang, X. *PLoS computational biology* **2014**, *10*, e1003767.
- [94] Plattner, N.; Noé, F. *Nature communications* **2015**, *6*, 7653.
- [95] Kolomeisky, A. B. *The Journal of chemical physics* **2011**, *134*, 155101.
- [96] Cao, J. *The journal of physical chemistry. B* **2011**, *115*, 5493–5498.
- [97] Min, W.; Jiang, L.; Xie, X. S. *Chemistry, an Asian journal* **2010**, *5*, 1129–1138.
- [98] Piephoff, D. E.; Wu, J.; Cao, J. *The journal of physical chemistry letters* **2017**, *8*, 3619–3623.
- [99] Kirchhoff, G. R. *Annalen der Physik* **1847**, *148*, 497–508.

- [100] Tutte, W. T. *Mathematical Proceedings of the Cambridge Philosophical Society* **1948**, *44*, 463–482.
- [101] Greenhill, C.; Kwan, M.; Wind, D. On the number of spanning trees in random regular graphs. <https://arxiv.org/pdf/1309.6710>.
- [102] Nicholson, W. L. *The Annals of Mathematical Statistics* **1956**, *27*, 471–483.
- [103] Wu, J.; Mehta, N. B.; Zhang, J. Flexible lognormal sum approximation method. GLOBECOM '05. IEEE Global Telecommunications Conference, 2005. 2005; pp 3413–3417.
- [104] Winter, P. *BIT* **1986**, *26*, 44–62.
- [105] Chakraborty, M.; Chowdhury, S.; Chakraborty, J.; Mehera, R.; Pal, R. K. *Complex & Intelligent Systems* **2019**, *5*, 265–281.
- [106] Eyring, H. *The Journal of Chemical Physics* **1935**, *3*, 107–115.
- [107] Zaitseva, J.; Jenewein, S.; Jumpertz, T.; Holland, I. B.; Schmitt, L. *The EMBO journal* **2005**, *24*, 1901–1910.
- [108] Perutz, M. F. *Biochimie* **1972**, *54*, 587–588.
- [109] Perutz, M. F.; Wilkinson, A. J.; Paoli, M.; Dodson, G. G. *Annual review of biophysics and biomolecular structure* **1998**, *27*, 1–34.
- [110] George, A.; Zuckerman, D. M. *Biophysical journal* **2023**, *122*, 284a.

Acknowledgments

First and foremost, I would like to thank my supervisor, Prof. Dr. Helmut Grubmüller, for encouraging me to pursue this thesis and for his unwavering support throughout it. The discussions with him were extremely insightful and left me highly motivated each time. I have learned a great deal from him, both scientifically and otherwise, and I am confident that this experience will serve me well in my future endeavors.

I would also like to express my gratitude to Prof. Dr. Stefan Klumpp and Dr. Johannes Söding for their continued interest in this work over the years and for their constructive feedback during our meetings.

My special thanks go to Eveline Heinemann, Stefanie Teichmann, Petra Kellers, Frauke Bergmann, and Antje Erdmann as well as Martin Fechner and Ansgar Esztermann-Kirchner. They provided invaluable assistance in all organizational and technical matters. I want to especially thank Petra Kellers for proofreading the introduction and the second chapter.

I am particularly grateful to Andreas, Nicolai and Steffen not only for the scientific discussions, but also for their camaraderie. I am certain that our frequent trips to the vending machine are the only reason it still exists.

For the countless opportunities for scientific and non-scientific exchange and for creating a wonderful work environment, I have to thank not only these three colleagues, but also the entire Department of Theoretical and Computational Biophysics. I can't imagine a better place to have done my PhD.

Finally, I would like to thank my parents for their support over the years and for having my back when I needed it. I am also grateful to my friends who gave me much needed breaks from work with their company. And most importantly, I deeply thank Miriam for always believing in me.