Master's Thesis

Mechanisms of Permeation in Potassium Channels

Permeationsmechanismen in Kaliumkanälen

prepared by

Maximilian Alexander Vossel

from Willich, Germany

at the Max-Planck-Institute for Biophysical Chemistry
in the Department for Theoretical and Computational Biophysics
in the Computational Biomolecular Dynamics Group

Thesis Period: 7th February 2018 until 7th August 2018

First Referee: Prof. Dr. Berend L de Groot

Second Referee: Dr. Aljaz Godec

ii

Abstract

K^+ channels are proteins that facilitate the passive permeation of K^+ ions through the cell membrane. They conduct at rates close to the diffusion limit while selecting against Na^+ ions by more than a thousandfold. This transport is crucial for a variety of physiologically fundamental processes like cell homeostasis or signal transduction along nerve cells. Knowledge about K^+ channels became available through electrophysiology experiments starting in the 1950s and crystallography studies since the end of the last century. The experimental data has been combined with computational studies, which lead to a detailed understanding of the molecular mechanism of K^+ ion permeation.

It has been confirmed that the channels' narrowest constriction – the selectivity filter – forces the K^+ ions to move in single file and is thereby responsible for both high current and high fidelity in rejecting Na^+. From an evolutionary standpoint one could imagine a common ancestor of K^+ channels, as the selectivity filter is composed of highly conserved residues. During subsequent specialization the underlying permeation mechanism should not have changed too much. However, different protein backgrounds or mutations at the filters' end do induce small differences.

In this thesis, we set off for a deeper understanding of the permeation mechanism by comparing computational models of different members of the K^+ channel family against each other. Our findings are validated by analyzing the effect of different force fields on the permeation mechanism. Using data from (recently published) non–equilibrium steady state molecular dynamics simulations of permeating K^+ channels we built discrete models of the permeation mechanism on a reduced phase space. While individual trajectories of permeation do indeed differ in their microscopic details, our model allows to make humanly comprehensible statements about probabilities of how many permeations follow similar pathways.

We could show that the 'direct knock-on' mechanism, where K^+ ions are in immediate adjacency to each other, is a robust process of permeation. Although exhibiting small, measurable variations, it is conserved in its main theme through all studied channels and force fields.

Exploiting local detailed balance we estimated a lower bound for the voltage drop across the selectivity filter, finding values which are in very good agreement with those expected from the literature.

Acknowledgment

First of all, I want to thank my supervisor Prof. Bert de Groot for giving me the opportunity to participate in the research of the truly fascinating topic of K^+ channels. He always had time for answering my questions, most often event at short notice. His elegant mix of granting me the freedom to explore ideas on my own and guiding me back to the path of the main question when I was starting to get lost on sidetracks made the time of this thesis a pleasant and insightful journey.

I also want to thank:

Wojciech Kopec for sorting my thoughts when I got overwhelmed by the complexity of the biochemistry that the field of K^+ channels brings along.

David Köpfer for introducing me to the 'channel work' in his inspiring manner and for arousing my interest for that field in the first place.

Aljaz Godec and David Hartich for long and fruitful discussions about the physical principles behind stochastic machines that K+ channels represent.

Julian Tim Brennecke and Andreas Volkhardt for various discussions about both physical and technical problems where their sharp reasoning helped me finding the answers I needed.

Helmut Grubmüller for providing the infrastructure and attracting me to biophysics in an inspiring talk at the physics colloquium of the University Duisburg-Essen.

The whole department of Theoretical and Computational Biophysics for an always helpful and honest atmosphere.

Carolin Debski and my family for the unconditional support they gave me throughout the time of the thesis and the years of studying before.

vi

Contents

Contents

1. Introduction

1.1. K^+ channels

Potassium (K^+) channels are proteins that penetrate the cell membrane. They allow K^+ ions to follow their electrochemical gradient through the otherwise impermeable cell membrane. This passive transport of K^+ ions through the channel occurs at rates close to the diffusion limit [HK55]. At the same time a selectivity for K^+ against Na^+ ions at a ratio of about 1:1000 is maintained [LHM01]. Considering that the K^+ is only slightly larger ($<20\%$) [Bat01] than the Na^+ ion while both are monovalent, such a strong selectivity is remarkable.

K^+ channels have high physiological relevance as they play a crucial role in a vast variety of physiological functions. These processes cover a versatile range with examples being cell homeostasis and signal transduction along nerve cells. They belong to the largest and most diverse group of ion channels and malfunction among them leads to a variety of diseases making them a promising drug target. [SS97]

The first experiments that lead to the postulation of K^+–selective sites in the membrane were conducted by Bernstein and documented in his famous **membrane hypothesis** [BT06]. His postulations were confirmed by Hodgkin and Huxley [HH39]. Their analysis lead to the widely–known **Hodgkin-Huxley-Model** for the propagation of action potentials in neurons [Phi13]. Deeper insights into the behaviour of individual entities of these K^+–selective sites in the membrane were initiated by the **patch clamp technique** [HMN$^+$81].

1.1.1. Selectivity Filter

Ions move through the channel's narrowest constriction, the selectivity filter (SF), in single file exerting a 'knock–on' mechanism. Both single file as well as a 'knock–on' type of transport were proposed by Hodgkin and Keynes[HK55] more than 60 years ago, long before the first structures of proteins were available. Just by looking at the variation of ion fluxes with different electric fields applied they could infer that ions are constrained to move in single file with two to three sites of this chain occupied on average. This prediction has been spectacularly confirmed four decades later by MacKinnon and coworkers[DTGG11, MCZM01] when they resolved the crystal structure of KcsA, a prokaryotic member of the potassium channel family from the bacterium *streptomyces lividans.* These experiments are just the most prominent ones, a sheer abundance of experimental data on ion channels is available nowadays. Especially since the advancement in x-ray crystallography, experimental and computational studies have been fruitfully combined leading to a detailed understanding of the dynamics of the molecular mechanism of permeation[RK91, BR01, BR03, RABI04, KSG$^+$14, KKV$^+$18].

The selectivity filter has been shown to be responsible for both the selectivity and the high K^+ ion current[DCP$^+$98]. It constitutes the four 'canonical' K^+ binding sites (S1 – S4), where the single file behaviour occurs. A schematic of the SF is shown on the right, depicting one of the four subunits in liquorice representation and the binding sites S1–S4 filled with K^+ ions by violet spheres. In addition to these four binding sites there are two more, one at the SF's extracellular exit (S0) and one in the otherwise water-filled cavity at the intracellular side (S_{cav}). As K^+ channels are tetrameric channels, the SF has a four-fold symmetry, too. The binding sites are defined by carbonyl-oxygen atoms that face inward into the pore and stem from each of the subunit's backbone. It happens that one or more of these carbonyl-oxygens moves and points radially outward of the symmetry axis of the SF. We refer to these as flipped carbonyls. The group of residues these carbonyl–oxygens
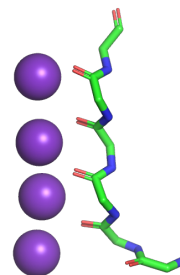
originate from is highly conserved among potassium channels. The sequence is (TxVGYG) [HLAM94, MN12], which is referred to as the 'signal sequence' of K$^+$ channels.

1.1.2. Our Selection: KcsA, MthK, Kv1.2 and NaK2K

Besides this high structural similarity, K$^+$ channels differ notably in other parts, especially in those sticking out of the membrane. This comes to no surprise, as these parts are often involved in mediating the response to external stimuli, which cover a diverse range. These differences in the rest of the protein cause changes to the hydrogen bond network in the protein background that stabilizes the filter. More prominent changes are caused by mutations of one of the residues of the SF [BM12, HLAM94]. Potential effects of these structural differences on the permeation mechanism are the subject of this thesis. Incorporating different channels into the analysis has already facilitated an understanding of the seeming contradiction of high throughput and high selectivity [KKV+18], so we expect it to help broadening the view on the permeation mechanism as well. The different K$^+$ channels we chose to analyze are introduced here briefly.
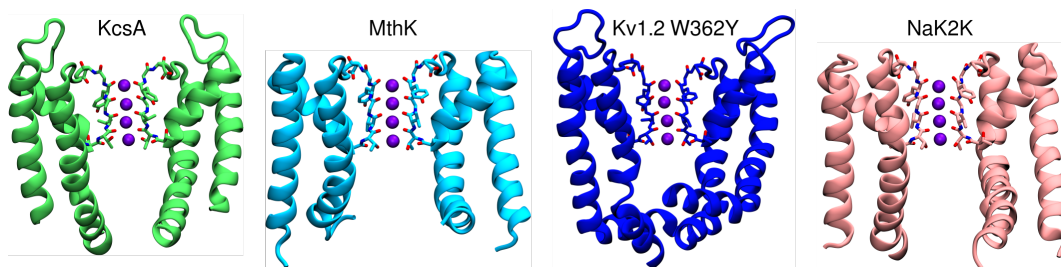

Figure 1.1.: The selection of different members of the K$^+$ channel family we chose to compare against each other. Only two subunits are shown for clarity. The similarity of the selectivity filter is immediately recognized. Image by courtesy of Wojciech Kopec

.

KcsA, a pH gated channel, has been the 'working horse' for the study of K$^+$ permeation [KSG+14]. The signal sequence reads TTVGYG.

Mutating residue 71 (and the corresponding residues in the other subunits) in KcsA from glutamate to alanine leads to a non-inactivating [1] channel named KcsA E71A. [CMT$^+$11] The signal sequence is TTVGYG, the change from E to A occurred three residues prior: ETA[..] to ATA[..]. MthK from the bacterium *Methanothermobacter thermautotrophicus* is a channel responding to intracellular Ca^{2+} binding. ATVGYG is the signal sequence. Kv1.2 is a mammalian voltage gated K$^+$ channel, whose sequence reads TTVGYG. NaK2K is an artificial channel [DSZ$^+$11] exhibiting the four canonical binding sites (S1–S4). It can be constructed by introducing mutations in the SF of the normally non-selective bacterial ion channel NaK [SYA$^+$06] that usually has only two binding sites. The corresponding sequence is TTVGYG. The SF is stabilized by the 3 residues preceding the signaling sequence as well as water molecules between these. Both the residues and the number of water molecules vary between the channels.

1.2. Computer Simulations

Problems involving the interaction of more than two bodies yield equations that are unsolvable analytically - but with the help of a computer, typically the solution can be determined up to a high accuracy. For this reason computer simulations are frequently applied in the field called many particle physics. Before computational solutions for determining properties of molecular substances were available, such properties had to be estimated using approximate theories. Examples, where these approximations lead to exact solutions, are the ideal gas or the harmonic crystal. One can say that certainly the majority of problems in nature are not as simple. Applying computer simulations to determining molecular dynamics, the only approximation one makes is on

---

[1]Inactivation is one of the two processes leading to a impermeable channel, the other mechanism is gating [Yel98]. Inactivation is caused by a conformational change of the SF while gating is a result of bundle crossing of inner helices that control the entry to the cavity S$_{cav}$. Configurations according to the gating mechanism are called open and closed. A possible coupling between gating and inactivation cannot be excluded. [IOTS10, CJC$^+$10, CJCP10]

the level of the intermolecular interaction between particles [FS02]. A more detailed description of molecular dynamics simulations follows in chapter 2.

Computer simulations often provide deeper insights into molecular processes than can be obtained from most experiments. In the ideal case, computer simulations deliver properties that can be compared to experiments directly. Especially the collaboration between scientists performing experiments and those executing simulations lead to productive and useful results, since working decoupled from each other both can independently misinterpret their findings.

More than 35 years ago the first protein, a bovine pancreatic trypsin inhibitor, was treated in a computer simulation. This simulation changed the view of proteins from rigid to dynamic structures, whose motions are crucial for its function [KM02].

1. Introduction

2. Theory

2.1. Molecular Dynamics

Molecular dynamics (MD) is a computer simulation technique for studying the motion of physical systems at the atomistic level. Systems of interest are e.g. biomolecules like proteins, where MD simulations allow useful insights in its function, as the latter is in turn often determined by the dynamics [KM02].

The exact dynamics of molecular systems is given by the time dependent Schrödinger Equation [Sch26]

$$i\hbar\partial_t\psi(\vec{r}, \vec{R}) = \hat{H}\psi(\vec{r}, \vec{R}). \tag{2.1}$$

Solving this equation is computationally prohibitively costly, even for small systems with about tens of atoms. The accessible timescales lie in the range of several hundred picoseconds nowadays, orders of magnitude below the timescales of biological interest.

A generic approach for reducing the computational effort is the Born–Oppenheimer approximation. It assumes that due to the huge difference in the mass of the light electrons ($\vec{r}_i$) and the heavy nuclei ($\vec{R}_i$) their dynamics can be addressed separately. Rephrased this means that the electrons follow the motion of the nuclei instantaneously. Due to the separability of the problem the Schrödinger equation for the electrons can be solved for stationary nuclei positions, which is done usually in the ground state, as the thermal energy in biological systems usually doesn't suffice for higher excited electronic states. The interaction between nuclei and electrons can now be described

using this solution as an effective potential energy surface $E_e(\vec{R})$, resulting in

$$\hat{H}_e \; \psi_e(\vec{r}, \vec{R}) = E_e(\vec{R}) \; \psi_e(\vec{r}, \vec{R}). \tag{2.2}$$

The second step, based on the same rationale, is to treat the nuclei classically. The nuclei's motion is now governed by that potential function, thereby we are neglecting tunneling effects and chemical bond breaking.

2.1.1. Integration Scheme

The dynamics of the point sized atoms can thus be described classically by integrating **Newton's equations** of motion,

$$m_i \, \ddot{\vec{r}}_i = \vec{F}_i, \qquad i = 1, ..., N, \tag{2.3}$$

We change the notation for the positions of the atoms from capital $\vec{R}_i$ to small $\vec{r}_i$. Further, m_i is the mass and $\vec{F}_i$ the force acting on one of the N particles. The forces are calculated using an approximation V_{ff} of the potential energy function $E(\vec{r})$ which will be described in more detail in the next section. The potential solely depends on the atoms position $\vec{r}_i$. The gradient of the potential regarding the position of a particle provides the force acting on it,

$$\vec{F} = -\frac{\partial}{\partial \vec{r}_i} V_{\text{ff}}(\vec{r_1}, \vec{r_2}, ..., \vec{r_N}). \tag{2.4}$$

These equations of motion are integrated using a numerical integrator, e. g. the 'leapfrog' algorithm, see 2.5, which is implemented in `GROMACS`. Contrary to algorithms like the Runge-Kutta integrator, the former is of symplectic nature which means that it conserves the energy of conservative systems. Conservative systems exhibit a potential function that depends on the position of the atoms solely. The 'leapfrog' algorithm calculates the velocities $\vec{v}$ at a time point between the evaluation of the positions and forces.

$$
\begin{aligned}
\vec{v}_{t+\frac{\Delta t}{2}} &= \vec{v}_{t-\frac{\Delta t}{2}} + \frac{\vec{F}}{m}\,\Delta t, \\
\vec{r}_{t+\Delta t} &= \vec{r}_t + \vec{v}_{t+\frac{\Delta t}{2}}\,\Delta t.
\end{aligned}
\tag{2.5}
$$

In order to control the numerical error associated with a finite time step Δt, it has to be chosen smaller than the oscillation period of the fastest vibration in the system. This is given by the movement of hydrogen and lies at about $1\,\mathrm{fs}$. One can fix the corresponding bond length in order to double the time step and thereby gain a better approximation of the quantum mechanical description additionally. A maximum of $\Delta t = 4\,\mathrm{fs}$ integration time step can be achieved if one applies virtual interaction sites [HKVDSL08], another constraint for the hydrogen motion concerning the angular vibration.

2.1.2. Force Fields

The approximation of the effective potential $E(\vec{r})$ by a sum of analytic functions, which are easy calculated, results in a set of parameters called the force field V_{ff}:

$$
\begin{aligned}
V_{\mathrm{ff}}(r) \;=\;\; & \sum_b \frac{1}{2}K_b(b-b_0)^2 \\
+\; & \sum_\theta \frac{1}{2}K_\theta(\theta-\theta_0)^2 \\
+\; & \sum_\zeta \frac{1}{2}K_\zeta(\zeta-\zeta_0)^2 \\
+\; & \sum_\phi K_\phi(1+\cos(n\phi-\delta)) \\
+\; & \sum_{\text{pairs: }ij} \frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}} \\
+\; & \sum_{\text{pairs: }ij} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]
\end{aligned}
\tag{2.6}
$$

Based upon the assumption of additivity, the analytic functions describing the interaction between particles can be separated into bonded and non bonded interactions. For the the bonded interactions, the first four terms in equation 2.6, the summation indices are the bonds b, the angles θ, the improper dihedral angles ζ and the dihedral angles ϕ. The last two terms in equation 2.6 are the non bonded interactions, where the sum is taken over all pairs of atoms i and j and are composed of the Coulomb and the Lennard–Jones potential. The Lennard-Jones potential approximates the effects originating directly from quantum mechanics. It consists of an attractive part representing the London dispersion forces (van der Waals forces) and a repulsive term, accounting for the Pauli exclusion principle. The angle δ is the regular value of the respective dihedral angle, r_{ij} is the distance between the two atoms i and j, ϵ_r is the dimensionless relative permittivity of the material in which the charges q_i are immersed, ϵ_0 is the electric constant, ϵ_{ij} is the depth of the Lennard-Jones potential well, σ_{ij} is the distance at which the inter-particle potential is zero.

For each atom type or possible pair of those a force field provides parameters allowing the calculation of forces for given positions. These parameters are obtained via two different methods. Quantum mechanical calculations deliver **ab initio** parameters like in the AMBER [WWC+04] and CHARMM [BBO+83] families. Another approach for determining parameters is to purely fit them to experimental values in order to reproduce thermodynamic properties correctly. In theory all force fields try to approximate the same, correct PES. In practice the force fields are optimized for different parameters and with different priorities leading to deviations between simulations using them. In the MD community an approach is emerging to minimize these errors, in the consensus force field approach the results of trajectories generated with different force fields are compared. Converging results indicate consistency whereas diverging results point towards errors.

2.1.3. Periodic Boundary Conditions

Proteins embedded in water typically yield simulation system sizes of several nanometers. Water droplets of this size hold an internal pressure of several tens of atmospheric pressures. In order to avoid such non physiological numbers it is common to make use of periodic boundary conditions, under which the system is interacting with infinite virtual copies of itself. So particles leaving the simulation box at one side re-enter the box from the opposing side.

2.1.4. Barostat and Thermostat

In order to reproduce the conditions of biological experiments, which are usually conducted under constant pressure and constant temperature (NpT), a barostat and a thermostat are applied. The thermostat compensates for the diverging kinetic energy due to the accumulation of numerical errors or the simulation not starting from an equilibrium state. It regulates the systems' temperature via rescaling of the atoms velocities; the barostat calculates the systems' pressure using virials and rescales the positions of the atoms' to the size according to the desired pressure [VLH+05].

2.1.5. Initialization

The initial positions of the atoms, i.e. the protein structures, are obtained experimentally by deploying x-ray diffraction, NMR or cryo-EM techniques. In order to prevent abnormally high forces due to too small distances between atoms, an energy minimization step is typically performed prior to the MD simulation. Small distances occur because of misplaced atoms during the refinement, or atoms that are closely packed due to crystallization in the x-ray diffraction procedure. The positions are updated according to the potential functions steepest descent, and are thus in a local minimum.

The initial velocities are generated according to a Maxwell distribution for a given temperature.

2.1.6. Limitations and Outlook

As the electrons are considered to remain in the ground state, electron transfer processes and electronically excited states cannot be treated and thus no chemical reaction can be addressed directly. Thus, molecular dynamics simulations are mainly limited to conformational changes of molecules as long as quantum chemical effects are neglected. QM/MM simulations combine both simulation types in order to take certain chemical reactions into account e. g. active sites of enzymes in catalytic reactions [Gro13].

The introduction of an additional degree of freedom, a titration coordinate λ , which allows the interpolation between protonation states, mimics the behaviour of one certain type of chemical reaction. This method allows simulations at constant pH, which influences the structure, dynamics, and function of molecules in solution [DTGG11].

At the current state of computing power the temporal range of MD simulations lies in several hundreds of nanoseconds per day for a system of several nanometers / hundreds of thousands of atoms per computing node.

2.1.7. Applied Electric Field Simulations

For realistic modeling of both electrophysiological experiments and cellular processes an electric potential difference across the membrane has to be established.

This potential difference can be generated by applying an external electric field [Rou08] over the full simulation box or by an ionic imbalance in a dual membrane setup, which is typical for the computational electrophysiology (CompEl) setup [KGDZ11]. The former seems to break the periodicity of the system by introducing different voltages for a particle crossing the box boundaries. Nevertheless the force acting on it doesn't exhibit discontinuities. The analogy of a virtual circuit with an embedded battery clarifies the setup [GKASR12].

2.2. Relevant Degrees of Freedom

Although one is usually interested just in the behaviour of a subset of atoms, for a full atomistic MD simulations the trajectory of all atoms, including all atoms of the solution, has to be calculated to reproduce thermodynamically correct properties. In other words, such a trajectory evolves in a high dimensional phase space with a subspace corresponding to the process of interest. A strong formalism to tackle this type of problem has been proposed by Mori [Mor65] and Zwanzig [Zwa61] more than half a century ago. The so called projection operator approach is based on the assumption that the effect of the environment can be modelled implicitly via a stochastic term in the otherwise explicit description of the subset of interest.

2.2.1. Stochastic Processes

The description of the system then changes from a deterministic (Newton's equation) to a stochastic one, given by the **Generalized Langevin Equation** (GLE)

$$m_i \ddot{\vec{r}}_i(t) = \vec{\mathcal{F}}_i - \int_0^t \mathrm{d}t' M_i(t - t') \, \dot{\vec{r}}_i(t') + \vec{f}_i(t), \qquad i = 1, \dots, \tilde{N}, \qquad (2.7)$$

where $\vec{\mathcal{F}}_i$ is the force acting on a particle i which depends just on the positions of particles $\vec{r}_i$. The major difference to Newton's equation of motion (2.3) is the convolution of the particles velocity $\dot{\vec{r}}_i$ with a memory function M_i, which resembles a dragging force acting opposed to the particles velocity. The stochastic term modeling the collisions with the environment is $\vec{f}_i(t)$, it is assumed to be stationary and has no preferred direction: $\langle \vec{f}_i(t) \rangle = 0$. The relationship between the memory function and the stochastic force, according to the **Fluctuation-Dissipation theorem** (FDT) [Kub66], reads

$$\langle \vec{f}_i(t) \cdot \vec{f}_i(0) \rangle = 3 k_\mathrm{B} T M_i(t). \qquad (2.8)$$

2. Theory

The memory function M_i is known to decay rapidly in a biological medium like molecular bulk [MSO82] [MO70]. For the velocity of ions in proteins the correlation time lies in the sub-picosecond range as shown explicitly in [RK91]. This fast decay in $M_i(t)$ allows to make the following approximation, $M_i(t) \simeq \gamma_i \delta(t)$, which is also known as the Markov assumption. For a process to be Markovian the following condition must be fulfilled: Only the current state determines the probability to jump to another state, i.e. no memory longer than one jump is taken into account to describe the process. While the deterministic MD trajectory given by integrating Newtons equation of motion is inherently Markovian, as the states are given by the point in phase space which determines the next point in phase space exactly, one might introduce memory by limiting the phase space to the configuration space. The derivation above allows to judge that for timescales longer than several picoseconds no memory will be introduced by neglecting the velocity of the particles. This allows for a simplification in the GLE, leading to an equation known by the name classical **Langevin Equation** (LE)

$$m_i \ddot{\vec{r}}_i(t) = \vec{\mathcal{F}}_i - \gamma_i \dot{\vec{r}}_i(t) + \vec{f}_i(t), \tag{2.9}$$

with the corresponding FDT, $\langle \vec{f}_i(t) \cdot \vec{f}_i(0) \rangle = 6 k_\mathrm{B} T \gamma_i \delta(t)$, again assuming uncorrelated, stationary, Gaussian distributed forces.

In this case the process of interest is the ion permeation through the narrowest part of the channel. But not always a good intuitive choice can be found. A good rule and requirement is to include slow processes. These are processes that occur on a timescale where the fast processes have relaxed so that these can be modeled with a stochastic force. Violations of this requirement might occur if the protein undergoes changes on a longer timescale than the one on which permeation happens. These pitfalls will be addressed and discussed in chapter 4.

As a last approximation let us look at the case where the friction term with coefficient γ_i is large compared to the inertial terms $m_i \ddot{\vec{r}}_i(t)$, which therefore can be neglected. In this widely known regime of a 'low Reynolds number' [Pur77], the dynamics become overdamped. The equation of motion de-

scribing such **Brownian Dynamics** is typically written using the Einstein relation, $D = k_{\mathrm{B}}T/\gamma$ – another form of the FDT – and reads

$$\dot{\vec{r}}_i(t) = \frac{1}{\gamma}\vec{\mathcal{F}}_i + \frac{1}{\gamma}\vec{f}_i(t).\tag{2.10}$$

Using this equation it is possible to make a rough approximation of ion permeation [KSG$^+$14], [BR03]. The preceding description of Markovian dynamics on the continuum level with the Langevin equation (2.9) can be transferred onto a discrete set of states leading to a Master Equation (2.12).

2.2.2. Markov Processes

The Markov process is the simplest stochastic process exhibiting a causal relationship between different time points. Its relevance lies in the beneficial combination of its wide applicability whilst maintaining mathematical simplicity [Röp87].

The above mentioned memorylessness or Markovianity can be stated in an equation that describes the conditional probability to be in state n at timestep t given the states before,

$$p(n_t \mid n_{t_{-1}}; \ldots ; n_{t_1}) = p(n_t \mid n_{t_{-1}}).\tag{2.11}$$

Expanding the conditional probability $p(n_t \mid n_{t_{-1}})$ into a power series around $\Delta t = t - t_{-1}$ and taking the limit $\Delta t \to 0$ allows to define a transition probability w_{nm} and by that an equation of motion for the probability of being in state p_n, which is called the **Master Equation**

$$\frac{\partial}{\partial t}p_n = \sum_{m \neq n} p_m w_{mn} - p_n w_{nm}.\tag{2.12}$$

The two terms under the sum can be interpreted as influx from other states m into state n and outflux from n into these. The time dependence of $p(t)$ has been dropped for clarity, the transition rates w_{nm} are assumed not to depend

on the time t. The definition of a rate matrix W with elements

$$W_{mn} = \begin{cases} w_{mn}, & \text{if } m \neq n \\ -\sum_k w_{nk}, & \text{if } m = n \end{cases}, \tag{2.13}$$

allows to rewrite the master equation 2.12 in the following convenient way

$$\frac{\partial}{\partial t} p_n = \sum_m p_m W_{mn} \tag{2.14}$$

2.2.3. Steady States

One considers a process as **irreducible**, if the underlying network doesn't separate into distinct sets which are unconnected, i.e. from each state every other state is reachable over a series of states of that network. This implies **ergodicity**. Ergodic systems will, in the limit of $t \to \infty$, visit every state infinitely often and relax into a unique **steady-state**, where each of the p_n^s does not change over time $\frac{\partial}{\partial t} p_n^s = 0$. A consequence of ergodicity is that averaging over time and phase space Ω gives equal results. In a continuous formulation for an arbitrary observable $O(\vec{x}(t))$ that reads

$$\lim_{t \to \infty} \int_0^t dt'\, O(\vec{x}(t)) = \int_\Omega d\vec{x}\, p^s(\vec{x})\, O(\vec{x}). \tag{2.15}$$

There are two physically distinct classes of steady states, depending on whether the condition of **detailed balance** (DB) is met or not.

If the steady state obeys DB it represents the thermodynamic **equilibrium** and the stationary distribution is given by the **Boltzmann-Distribution**,

$$p_n^{\text{eq}} = Z^{-1} \exp[-E_n/k_{\text{B}}T] \tag{2.16}$$

with a potential energy E and the partition function $Z = \sum_n \exp[-E_n/k_{\text{B}}T]$. Provided that the studied system corresponds to a thermodynamic ensemble (NVT, NpT), the potential energy E is equal to the associated free energy (F: Helmholtz, G: Gibbs). DB states that at equilibrium each elementary process

occurs with the same frequency as its reverse process. This concept was first explicitly mentioned by Ludwig Boltzmann in the proof of the H-Theorem [Bol64]. There a physical justification of DB was given by microscopic reversibility, the fact that the underlying equations of motion are symmetric with respect to inversion in time. For discrete states (see figure 2.1) the concept is that for each pair of states (i.e. detailed) the forward and backward transitions balance,

$$p_n^{\text{eq}} \, w_{nm} = p_n^{\text{eq}} \, w_{mn} \qquad \forall (n, m) \, \text{pairs.} \qquad (2.17)$$
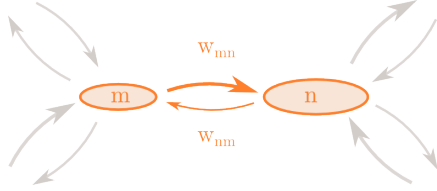
Figure 2.1.: Detailed balance

In this case there is no net probability current $j_{nm} = j_{mn} = 0$ between any pair of states m and n. Using equation 2.16 to reformulate equation 2.17 the ratio between two transition rates or stationary distributions is determined by the difference in the potential energy $E_n - E_m$ between the states n and m,

$$\ln \frac{w_{mn}}{w_{nm}} \bigg|_{\text{eq}} = \ln \frac{p_n^{\text{eq}}}{p_m^{\text{eq}}} = (E_m - E_n)/k_{\text{B}}T. \qquad (2.18)$$

When external forces or fields, for example a small uniform external electrical field, are applied detailed balance is broken and the above equation does not hold anymore. However, a weaker symmetry condition called **Local Detailed Balance** can be found,

$$\ln \frac{w_{mn}}{w_{nm}} = (E_m - E_n + \mu_{nm})/k_{\text{B}}T. \qquad (2.19)$$

In such a **non-equilibrium steady state** (NESS) the logarithmic ratio of transition rates between states m and n is shifted by μ_{nm}. In contrast to the energy difference $E_n - E_m$, which acts on the system in form of a conservative force, this μ_{nm} just acts on the link between states n and m. For example, a

particle with charge q undergoing a transition in an constant electric field E along a distance d_{nm} implies $\mu_{nm} = qEd_{nm}$.

The concept of local detailed balance was introduced in [KLS83] in a similar context, i.e. of ions moving on a lattice with an electric field applied. A deeper understanding of the constraint imposed by local detailed balance can be found in the review [Sei12]. For transitions to obey local detailed balance the dynamics inside a state should have equilibrated on the timescale the transition takes place. If this constrained equilibrium is met, thermodynamic properties like entropy, internal energy and free energy are well defined as for conventional equilibrium thermodynamics.

If detailed balance is not met, there exist non-vanishing steady state probability currents between two states n and m,

$$j_{mn}^s = p_m^s w_{mn} - p_n^s w_{nm} \neq 0. \tag{2.20}$$
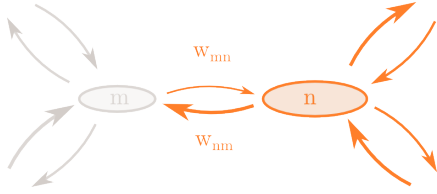
Figure 2.2.: Conservation of probability

The master equation (2.12) ensures conservation of probability (see figure 2.2), for steady states $\frac{\partial}{\partial t} p_n^s = 0$ it means that the net influx must equal the net outflux of state n, see equation 2.20. This formulation is also known as Kirchoffs' first law [Kir47]. As presented in [Sch76, Kal07, JQQ03] this conservation law induces **probability flux cycles**. The stationary probability fluxes can be decomposed using these cycles as a basis.

2.3. Markov State Models

A **Markov State Model** (MSM) approximates molecular dynamics (simulation and experiment) by a chain of Markov processes, where the states of the chain are represented by a partitioning of the phase space and the transitions are given by Markov jump processes. All the conditional transition probabili-

ties between the n distinct states of the model are stored in an $n \times n$ matrix, the transition probability matrix P. As $P = \exp[W \Delta t]$, P can be approximated from the rate matrix W by taking the first two elements of the series expansion of the matrix exponential of W. This results in

$$P_{mn} = \begin{cases} w_{mn}\Delta t, & \text{if } m \neq n \\ 1 - \sum_k w_{nk}\Delta t, & \text{if } m = n \end{cases}. \tag{2.21}$$

P is often called the propagator as it describes the propagation of probabilities through time.

If one diagonalizes P from the left,

$$P \begin{pmatrix} \vec{r}_1 & \vec{r}_2 & \cdots & \vec{r}_n \end{pmatrix} = \begin{pmatrix} \vec{r}_1 & \vec{r}_2 & \cdots & \vec{r}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

$$PR = R\Lambda$$

and from the right, respectively,

$$\begin{pmatrix} \vec{l}_1^T \\ \vec{l}_2^T \\ \vdots \\ \vec{l}_n^T \end{pmatrix} P = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} \vec{l}_1^T \\ \vec{l}_2^T \\ \vdots \\ \vec{l}_n^T \end{pmatrix}$$

$$LP = \Lambda L$$

one arrives at

$$P = R\Lambda R^{-1}$$
$$P = L^{-1}\Lambda L$$
$$P = R\Lambda L \tag{2.22}$$

with Λ being a diagonal matrix containing the eigenvalues and L,R the matrices containing the left and right eigenvectors, respectively. As the propagator P describes a stochastic process, the first of the eigenvalues $\lambda_1 = 1$ and its eigenvector corresponds to the stationary distribution. In NESS some of the eigenvalues might be complex, always appearing in complex conjugate pairs because P just contains real entries. These complex eigenvalues have in general complex eigenvectors that are related to the above mentioned cycles. For the other eigenvalues it holds that $0 < \text{Re}(\lambda_i) < 1, \quad \forall i > 1$. Each of the eigenvalues λ_i is linked to a corresponding eigenvalue μ_i of W by the same relationship linking the matrices, namely $\lambda_i = \exp[-\Delta t \mu_i]$. The reciprocal of the rates $1/\mu_i$, is a timescale t_i that describes how fast the process described by the corresponding eigenvector decays, i.e. how fast probability density is moved towards the equilibrium. These implied timescales

$$t_i[P(\Delta t)] = \frac{-\Delta t}{\ln \lambda_i(P(\Delta t)} \tag{2.23}$$

give an idea how long the memory of the initial condition persists in the system and are therefore often called correlation times.

The partitioning of the state space of biomolecules in general relies on the experimental well justified assumption of long-lived, metastable states the molecule visits during along a trajectory [WBB$^+$10], [NGB$^+$07], [EML$^+$05]. As introduced above, the loss of Markovianity caused by neglecting the momentum part of the phase space gets recovered in the time scales biological processes take place, so that the probability to jump to a certain state in the configuration space solely depends on the previous state. One could call this the Markov property of proteins. MSMs generate a humanly comprehensible

reduction of MD trajectory data while maintaining statistical significance, a typical problem of traditional MD analysis. The view of single trajectories is abandoned and replaced by a ensemble view of the dynamics [SFHD99]. Once a transition probability matrix is estimated, individual realizations (artificial trajectories) can be generated by performing a Monte Carlo simulation of the Markov chain, as described in section 3.4.

2. Theory

3. Methods

In this thesis we use both above mentioned methods to create an intuitive understandable model of ion permeation. We reduce the degrees of freedom of the phase space first and build a Markov model on this reduced phase space assuming that all dimensions relevant for ion permeation are covered.

3.1. MD simulation

All MD simulations were carried out by Wojciech Kopec and David Köpfer. A detailed description of the computational procedure can be found in the articles [KKV$^+$18] and [KSG$^+$14] and their respective supplementary information. For the sake of completeness we give a short overview over the computational setup. The theoretical foundations for MD simulation can be found in section 2.1.

All channel proteins were embedded in a patch of a POPC membrane (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine). A transmembrane voltage U was established with the CompEL [KGDZ11] setup ($U \approx 220\,\mathrm{mV}$) or with an applied electric field [GKASR12] ($U \approx 280\,\mathrm{mV}$) as described in section 2.1.7. The MD simulations were performed using either `GROMACS` 5.0 or 5.1 [VDSLH$^+$05, HKVDSL08, PPS$^+$13, AMS$^+$15]. TIP3P was used as the water model in all simulations of this analysis [JCM$^+$83]. For each simulation of a channel protein two force field families were used, `AMBER` (version 99sb, just Kv1.2 with 99sb*ILDN) [HAO$^+$06, BH09, LLPP$^+$10] and `CHARMM` (version 36m) [HRN$^+$16, BBO$^+$83, KVF$^+$10, BZS$^+$12]. Applying virtual sites [FHB99] allowed for an integration timestep of $4\,\mathrm{fs}$, control simulations at $2\,\mathrm{fs}$ assured validity of the generated trajectories. The crystal structures of the four channel proteins were taken from the Protein Data Bank [BWF$^+$06], all in the open configuration, as introduced in section 1.1.

KcsA
- PDB ID: 3f5w [CJCP10]
- conductive SF configuration from PDB ID: 1k4c [ZMCKM01]

MthK
- PDB ID: 3ldc [YLJ10]

Kv1.2
- PDB ID: 2r9r [JJES13]
- pore domain from 'paddle-chimera' Kv1.2-Kv2.1 structure
  (remains open at positive voltages)
- W362Y mutation to prevent SF instabilities
  (equivalent mutation in Kv1.6 doesn't alter conduction and selectivity
  [SZRJ11]

NaK2K
- PDB ID: 3ouf [DSZ+11]
- F92A mutation to increase ionic current [SZC+13]

The electric field in the simulations is applied in the direction that leads, according to the natural orientation of K^+ channels in the cell membrane, to an outward current of K^+ ions. The further analysis is performed for each protein and for each force field resulting in one MSM for one of these setups.

3.2. Dimensionality Reduction

For the analysis of MD trajectories that span a time window of several µs and exhibit hundreds of permeation events, we reduce the dimensionality of the system.

Own contribution: The first filtering of the trajectory–the extraction the positions of molecules in direct neighbourhood of the SF–was done using a script written by Bert de Groot. I wrote the scripts for all steps of the further

analysis. All of those scripts are written in python [VR$^+$07] making heavy use of the `numpy` package [WCV11]. The scripts are made available at `https: //gitlab.gwdg.de/deGrootLab/permeation-pathways`.

3.2.1. Projection

We first extracted the positions of oxygen atoms of the SF and of K$^+$ ions and water molecules in the direct neighbourhood of the SF out of the full trajectory and projected these coordinates onto the z-axis defined by the SF. A typical time trace of the resulting one dimensional data is shown in figure 3.1 in the upper graph.

3.2.2. Digitization

The one dimensional binding sites in the SF are defined according to the canonical binding sites (see section 1.1). The bin edges of the SF binding sites are represented by the mean of each of the SF oxygen layers of the four subunits. Using the identity of the K$^+$ ions and water molecules we bin each of them according to the defined bin edges. This procedure is done for each time step resulting in a digitized time trace for all molecules entering or permeating the SF. An example of such digitized molecule positions is depicted in figure 3.1, middle panel. The outer parts of the SF can be occupied by several molecules at once. This results in an apparent overlap of molecules in this representation, which is an artifact. For the binding sites S1-S4 on the contrary, multiple occupation is not in agreement with the single file knock-on mechanism described in 1.1. Since in this thesis the focus lies on the potassium permeation mechanism, we treat deviations from the canonical occupation as exceptions. Double occupations in S1 and S4 are mapped to the corresponding single occupation in a way that the molecule nearer to the interior of the SF is kept. Double occupations in S2 and S3 do not comply with the single file transport, therefore parts of the trajectory, where this happens, are cut out and not taken into account for the analysis. Such deviations will be addressed in chapter 4.

3.2.3. Discretization

The lowest graph in figure 3.1 depicts the discretized trajectory. This trajectory evolves in a phase space representing the instantaneous occupation of the SF. Assuming that each of the four binding sites can be occupied by just one molecule (water or K^+ ion) at a time or be void, the resulting phase space has $3^4 = 81$ states. Most of these 81 states are energetically very unfavourable and are therefore not visited during simulations of permeation resulting in a far smaller phase space for permeation. Using the digitized time trace that we obtained for each molecule, it is easy to extract the number of molecule jumps between binding sites for each time step.

This step of the analysis converts the trajectory, a former 3N×T array,

$$X = [x_1 = x(t = 0), x_2 = x(t = \Delta t) \dots x_T = x(t = (T-1)\Delta t), \qquad (3.1)$$

that contained the coordinates of every atom for each timestep into a 1×T array of consecutively visited states

$$S = [s_1, s_2 \dots s_T] \qquad (3.2)$$

and a 2×(T-1) array for the intermediate molecule jumps

$$J = [j_1, j_2 \dots j_{T-1}] \qquad (3.3)$$

for each trajectory with the number of time steps T. J contains two numbers for each timestep: $j_i = (j_i^k, j_i^w)$ for K^+ ions and water molecules, respectively. As these arrays are not vectors in the physical sense but just tuples of numbers we refrain from using the vector notation here.
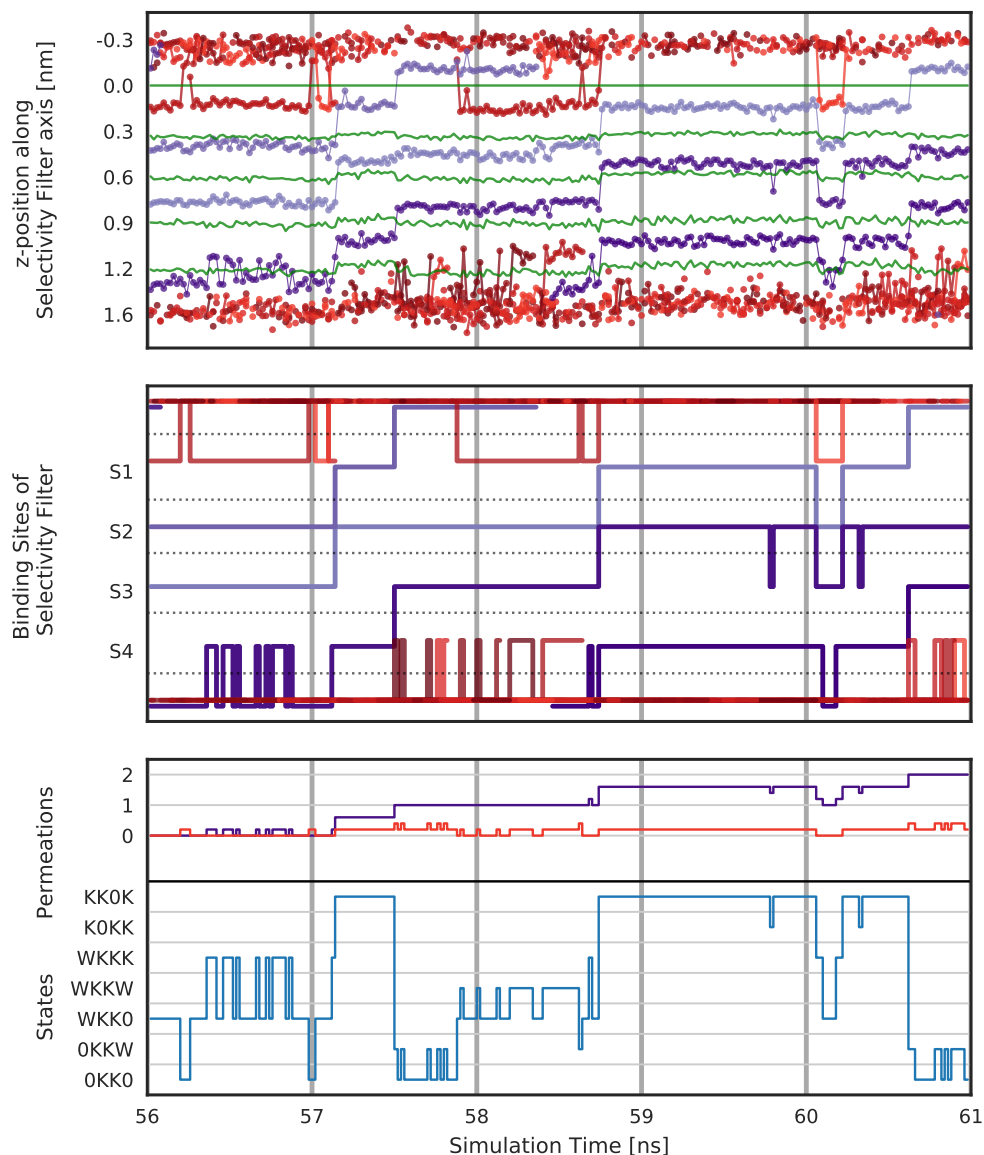
Figure 3.1.: Methods involved in the discretization of trajectories for MSM estimation. From the top to the bottom different stages of discretization are depicted. All graphs share the x-axis (simulation time). In the upper graph the z-positions of oxygen atoms of the SF (green, lines) and of K^+ ions and water molecules (violet and red lines with dots, respectively, different shades for different molecules) in the SF are given relative to the first layer of oxygen. The middle graph shows the digitized version of the data in the upper graph, all molecules entering or permeating the SF are binned into the canonical binding sites S1-S4. The graph at the bottom shows the fully discretized version of the trajectory, the discrete states represent the instantaneous occupation of the SF. The violet and red line count the jumps of K^+ ions and water molecules through the 4 binding sites, 5 jumps yield one full permeation.

3.3. Markov State Model Generation

3.3.1. Transition Count Matrix

The discrete trajectory is in a format that can be used to determine a count matrix ,

$$c_{od}^{kw}(\tau) = \sum_{n=1}^{T-m} \chi_o(s_n)\chi_d(s_{n+m})\chi^{kw}(j_n), \tag{3.4}$$

where the lag time $\tau = m\Delta t$ needs to be an integer m multiple of the time resolution Δt of the trajectory. A combination of different versions of the indicator function χ allows for discrimination of distinct transitions between states:

$$\chi_{o/d}(s_n) = \begin{cases} 1, & \text{if } s_n \in o/d \\ 0, & \text{otherwise} \end{cases} \qquad \chi^{kw}(j_n) = \begin{cases} 1, & \text{if } (j_n^k, j_n^w) = (k, w) \\ 0, & \text{otherwise} \end{cases}. \tag{3.5}$$

While these distinct transitions between the same states are rather untypical for MSMs, they are mentioned in [Sei12] as transition channels and should in principle pose no problem in the further evaluation of the count matrix. Each possible transition contains the origin state (o), destination state (d), K^+ jumps (k) and water jumps (w), requiring a four dimensional array in order to store them. We circumvented this by storing a dictionary containing origin, destination, k- and w-jumps as key and the corresponding count number as value. The counts for different trajectories are all summed up in a joint count matrix/dictionary.

The transition counts can be estimated via a subsampling with interval $\tau = o$r with a sliding window of length τ, see figure 3.2 for a depiction. Subsampling omits a major fraction $1 - T/\tau$ of the trajectory, therefore the window count method is often preferred, as $T - m$ timesteps can be evaluated with this approach. One has to keep in mind that nearby transitions will very likely be statistically dependent. Nevertheless, the window count method asymptotically leads to the correct transition matrix [PWS+11]. Throughout this analysis the sampling time of transitions is $\tau = 40 \, \text{ps}$, the smallest com-
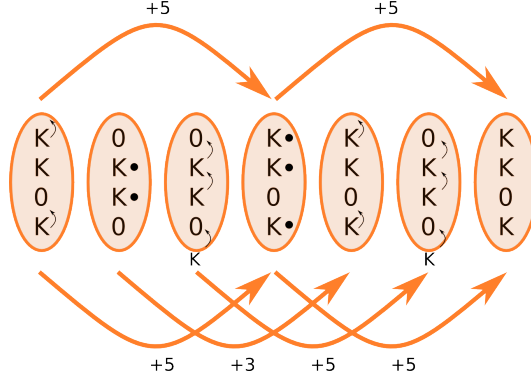
Figure 3.2.: Illustration of subsampling (top arrows) and sliding window count (lower arrows). τ is $3\Delta t$. Arrow labels are the jumps of K$^+$-ions during τ. Small arrows depict the movement of the ions in one timestep Δt, dots mean no ion move for this transition.

mon denominator of the trajectories output frequencies.

3.3.2. Transition Probability Matrix

A transition probability matrix is row-stochastic meaning the outgoing probabilities sum up to one for each state. In [PWS$^+$11] it has been shown that the maximum probability estimator for the probability matrix T_{od} is the trivial estimator

$$T_{od} = \frac{c_{od}}{c_o} \tag{3.6}$$

with $c_o = \sum\limits_{d=1}^{N} c_{od}$. The collapsed count matrix c_{od} without superscript ^{kw} contains the sum of all different transition types of between states o and d, the ratio of these is saved in an additional array.

3.4. Markov Chain Monte Carlo

As soon as a MSM with conditional probabilities has been constructed, we can perform Markov Chain Monte Carlo (MC) [VNU51, MRR$^+$53] simulations to

test our model for consistency and to judge whether a stationary distribution has been reached in the trajectories used to generate the model. Artificial realizations are generated by randomly drawing transitions according the conditional probabilities given by the MSM. The starting state is chosen the same as in the original trajectory.

The simplest possible implementation is depicted in the following algorithm. This algorithm produces an array of states of length n and $n-1$ transitions in between, each with the corresponding K^+ and water jumps (k, w).

Data : s: starting state, T_{od}: conditional probabilities, n: number of samples
Result : artificial trajectory of length n
states $=$ `emtpyarray`(n)
states$[0] = $ s
for i **in** `range`(n)**:**
    r $=$ `random`(*0,1*)
    o $=$ states$[i]$
    prob $=$ `cumsum`(T$[$o,:$]$)
    d $=$ prob$[$r$]$
    states$[i+1] = $ d
    if *multiple transitions possible***:**
        r $=$ `random`(*0,1*)
        prob $=$ `cumsum`(T$[$o,d,:$]$)
        (k,w) $=$ prob$[$r$]$

3.5. Probability Flux Cycles

There are several algorithms for decomposition of the steady state probability currents into cycles proposed in the literature [Sch76, Kal07, JQQ03, Hil05]. Each of these algorithms exhibits an ambiguity in the decomposition, i. e. the returned basis of cycles depends on the ordering of edges in the graph. Thus it is an inherent property of the steady state probability fluxes and not an artifact of the algorithms. The arbitrariness can be overcome in variety of ways,

from looking for a minimum cycle basis [HKM08] to sorting a first decomposition in all possible simple cycles by edge width first and decomposing the flux into cycles of descending occupancy. Another method for cycle decomposition is to use further information that the trajectories provide. As permeation is a reoccurring process we intuitively expect to see reoccurring patterns in the trajectory. Guiding the algorithm by this information results in an appealing cycle decomposition that furthermore uncovers the exact process of the permeation mechanism.

The following paragraph illustrates the algorithm. The discrete trajectory is first cleaned of all repetitions. Then oscillations (trivial cycles) between two states are removed if the forward and backward transition are exactly the same i. e. there are no net jumps performed after those two transitions. This preprocessed trajectory is then fed into a recursive function that decomposes it into cycles:

```
def find_cycles_recursive(dtraj):
    Data : discrete trajectory
    Result : cycle decomposition of discrete trajectory
    if no multifold states in dtraj:
        save dtraj as cycle
        exit
    else:
        s ← first of multiple occurring states
        (b,e) ← index of first and second occurrence of s
        (dtraj_1, dtraj_2, dtraj_3) ← split dtraj at (b,e)
        save dtraj_2 as cycle
        m_dtraj ← merge dtraj_1, dtraj_3
        if no multifold states in m_dtraj:
            save m_dtraj as cycle
            exit
        else:
            exit find_cycles_recursive(m_dtraj)
```

3. Methods

The algorithm is provided with a discretized version of the trajectory. All states are referenced via a unique id because of a simpler implementation of the algorithm for digits. A generic snippet of the trajectory is shown in the following. As it is more pleasant for the human eye to spot multiple occurrences in a list of digits than in a list of four letter codes which depict the selectivity filter states, we stick to the same digits as the algorithm does:

```
ntraj:    [10, 11, 9, 4, 22, 23, 11, 22, 11, 22, 10]
```

The algorithm finds all multiple occurring states in the trajectory and cuts out the cycle between the first two occurrences of the first of the multiple occurring states. This cut out part is saved as a cycle, together with the information about the transported K^+ ions. For the sake of clarity we refrain from showing the additional information here, except for the total number of jumps per cycle. In the example this first multiple times occurring state is the state with id 11 for the first and second cycle and state number 10 for the last cycle.

```
ntraj:    [10, 11, 9, 4, 22, 23, 11, 22, 11, 22, 10]
cycle:    [    11, 9, 4, 22, 23, 11               ]    K-Jumps: 5
ntraj:    [10,                 11, 22, 11, 22, 10]
cycle:    [                    11, 22, 11         ]    K-Jumps: 0
ntraj:    [10,                         11, 22, 10]
cycle:    [10,11,22,10                            ]    K-Jumps: 0
```

This trajectory snippet was decomposed in the three cycles:

```
cycle:    [11, 9, 4, 22, 23, 11]    K-Jumps: 5
cycle:    [11, 22, 11]              K-Jumps: 0
cycle:    [10,11,22,10]             K-Jumps: 0
```

32

4. Results and Discussion

4.1. Projection

We projected the positions of the oxygen atoms of the SF and of the K^+ ions and water molecules onto the z-axis defined by the SF. Histogramming these positions for the whole trajectory reveals that not all of the binding sites are occupied identically among the channels. The respective distributions are shown in figure 4.1. All positions are given relative to the first layer of oxygens above S1, its' position is marked with a green bar at z-position 0.
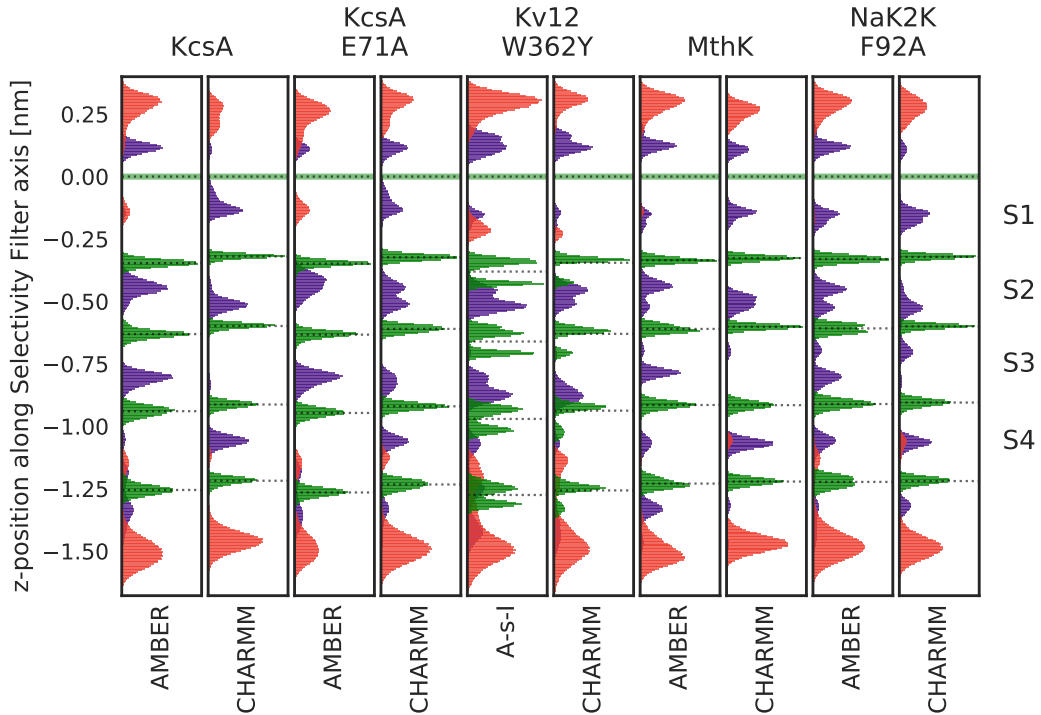
Figure 4.1.: Histograms of the z-positions of oxygen atoms of the SF (green peaks), the K^+ ions (violet peaks) and the water molecules (red peaks).

4. Results and Discussion

The overall positions of peaks in the binding sites are the same. We find violet peaks for the K^+ ions and red distributions for the water molecules between green peaks showing the oxygen positions. Both of the outer binding sites are occupied by more water than K^+. The water distribution is spread out wider and its' peak is located further outside, while the potassium is located closer to the inside of the filter. The inner binding sites S1–S4 are mainly occupied by K^+ ions, except for the `AMBER` versions of KcsA and KcsA E71A. For the latter, S1 and S4 are occupied purely by water for the former two and both versions of Kv1.2, where S1 and S4 accommodate both K^+ ions and water at slightly different positions. Some of the K^+ distributions show multiple maxima per binding site. Particularly striking is Kv1.2 where for both force field versions most of the binding sites exhibit two peaks. While a first interpretation for the multiple maxima might be that there exist multiple sub-binding sites, this interpretation can be discarded if one considers that, in Kv1.2, all the oxygen distributions exhibit two peaks as well. As all positions are given relative to the first oxygen layer and this is the only one not exhibiting multiple peaks, we can conclude that rather this one is the one moving between two positions. For the other channels' peak positions we see a shifting between two positions in the same binding site, especially strong for S2 and S3.

These differences between the channels and force fields in figure 4.1 do not result from inherently different positions in the binding sites. It is rather an effect of the averaging over all occurring selectivity filter occupations. The individual SF occupation states, like KK0K for K^+ in S1, S3 and S4 with a vacancy in S3, do not differ between the channels and force fields. The differences we see here are actually caused by different ratios of these occupancy states among the channels. In the appendix (A) in figure A.1 a breakdown of this figure into the individual states is shown, confirming that indeed the states themselves do not differ between the channels and force fields.

In contrast to the averages shown in figure 4.1, figure 4.2 shifts the view onto the fraction of time each binding site is occupied by a molecule or being vacant. Throughout all channels, the water fraction of S1 is higher for `AMBER` than for `CHARMM` trajectories, meaning that the former favours states of the SF

where water is in S1. The same holds for S4 in KcsA and KcsA E71A while for Kv1.2 W362Y, MthK and NaK2K F92A no such general trend is noticeable.
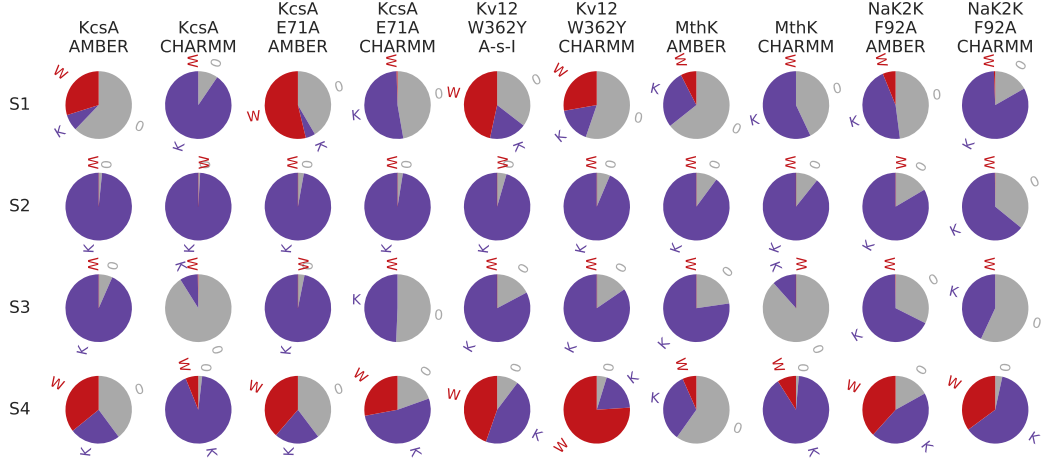
Figure 4.2.: A pie chart showing the fraction of each binding sites occupancy, either a K^+ ion (K, violet), a water molecule (W, red) or being vacant (0, grey).

It has to be emphasized here, that figures 4.1 and 4.2 just represent those parts of the trajectories, where an assignment of binding sites for each molecule was possible. These are the parts where single file transport using the canonical binding sites occurs. This limitation holds for the remaining part of the analysis, too, i.e. multiple occupations per binding site are not covered by the analysis. Double occupations in S2 and S3 occur together with other exceptions at the same time, namely flips of the carbonyl oxygens in the SF and water co-permeation. Together with anecdotal evidence showing that water does not follow the canonical binding sites, we can formulate the hypothesis that the water co-permeation mechanism is notably different from the one of pure K^+ permeation.

Figure A.4 in the appendix gives an overview of the used trajectories. The plots in this figure depict both of the above mentioned exceptions, the flips of the carbonyls as a deviation from the canonical state of the SF and the double occupations, not in agreement with the single file permeation. The fraction of trajectories that was used for the analysis is highlighted with a green shading.

Characteristic time points of the permeation of both water and K^+ ions are depicted in these overviews, too.

4.2. Markov State Models

After having the trajectories discretized as described in section 3.2.3, count and transition matrices are determined according to section 3.3.1 and 3.3.2. An example of these transition matrices is shown in figure 4.3. The transition matrices of the other channels and force field can be found in the appendix A in figure A.2.
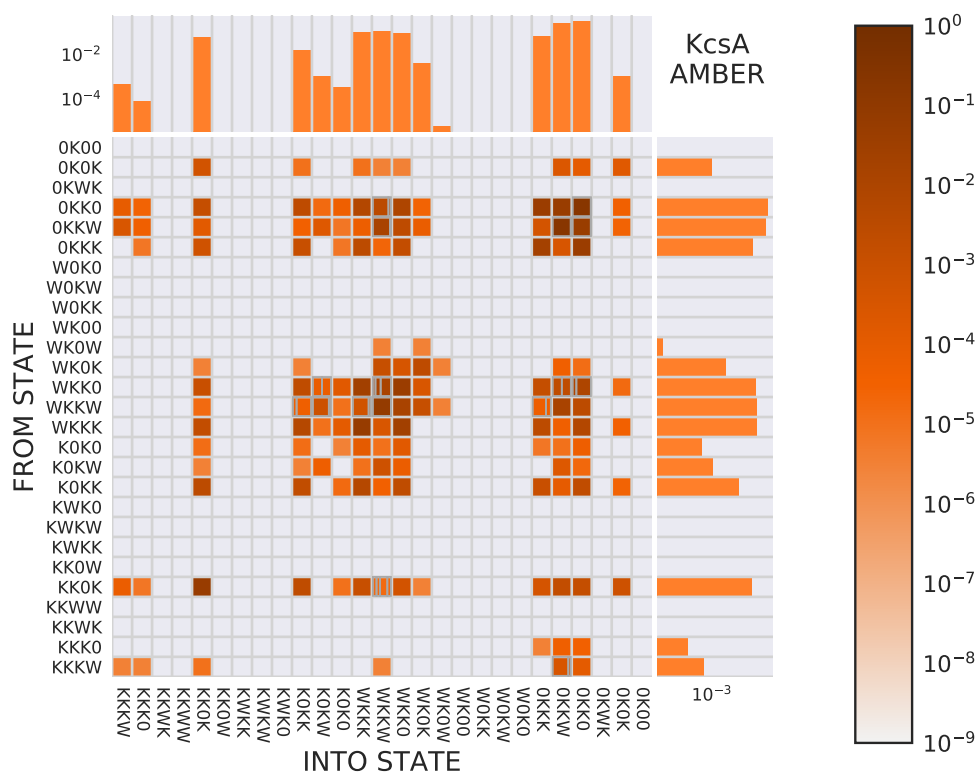
Figure 4.3.: Transition matrix for KcsA `AMBER`. The elements of the matrix are probability fluxes which equals transition probabilities times stationary distributions. Additional data is shown in the appendix A. The scale bar is the same for all later shown transition matrices in figure A.2. The units are probabilities.

On the axes of figure 4.3 the 4 letter codes denote the often visited occupation states of the SF. The elements of the transition matrix are given here as probability fluxes. That means that all transition probabilities are multiplied with their origin states' stationary probability. The scale bar attached to figure 4.3 gives the scale of these probability fluxes. The stationary distributions are determined by averaging over all trajectories' states. These are attached to the matrix on the top and on the right hand side, using the same axes as the matrix. In figure 4.3 we see that most of the elements are empty, they are included in the matrix just for the purpose of comparing these matrices among each other. On the diagonal we have the probability flux that stays in the same state, thereby denoting how long-lived this state is. The matrix possesses clusters with dark elements along and further off the diagonal in a quite symmetrical manner. These clusters depict regions where probability flows between a subset of states for a certain time before it leaves to other states again. In figure 4.3 these are especially the states 0KK0 and 0KKW in the top right corner and the states WKKW and WKK0 off diagonal. Between these states a lot of fast processes take place,which serves as a justification for clustering these states together later in section 4.4. Plotting the probability flux helps to get an intuitive picture of where most of the dynamics happens.

Wherever different transition types are possible between the same two states this is marked by a subdivision of the corresponding matrix element. The ratio of the resulting partitions gives the ratio of the transition types.

4.2.1. Timescales

The eigendecomposition of each of the transition matrices provides us with the eigenvalues shown in figure 4.4. For all matrices the first eigenvalue is 1, corresponding to the stationary distribution. Each of the other eigenvalues implies a timescale, namely the correlation time of the process the corresponding eigenvector describes. These are shown in the same figure 4.4, on the right hand side. One clearly sees that most of the processes happen on timescales between 10 ps and 10 ns, as visual inspection of the trajectories already indi-

cated. Interestingly, the correlation times show no spectral gap meaning that all the processes of permeation do occur on similar timescales.
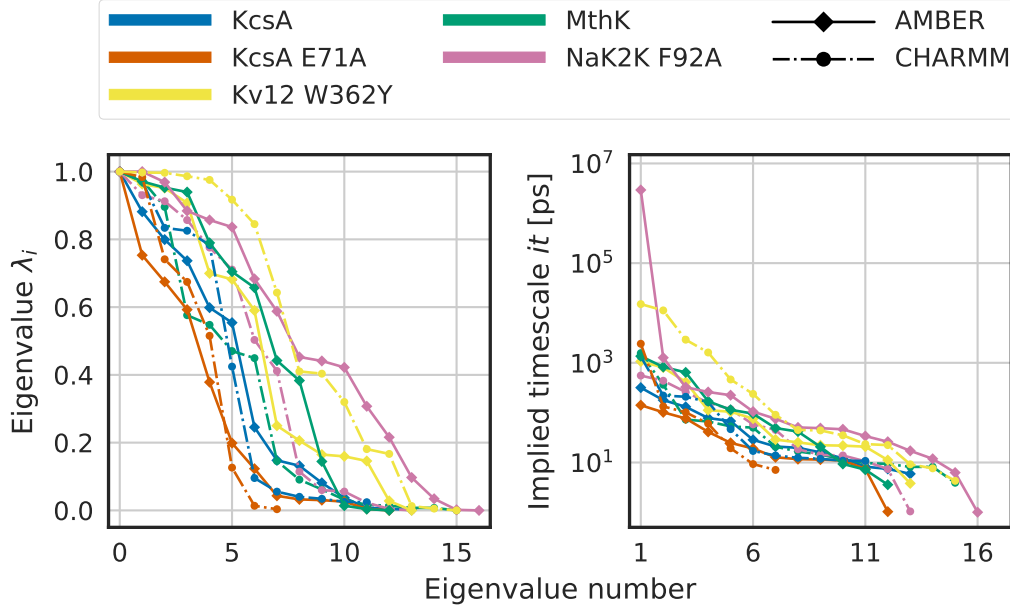
Figure 4.4.: Eigendecomposition for the transition matrices results in eigenvalues and eigenvectors. The eigenvectors describe the processes of the propagator and each eigenvalue corresponds to the correlation timescale on which the process transports probability towards the stationary distribution.

Only the first timescale of NaK2K `AMBER` deviates clearly from that pattern, corresponding to the slowest process in this channels' permeation processes. This timescale lies in the range of more than a µs, thereby exceeding the total simulation time. Hence we can conclude that the sampling for this channel – force field combination did not suffice and no steady state was reached during simulation.

4.2.2. Distribution of Dwell Times

The fortunate combination of long simulation times and fast processes of interest provides us with sufficient sampling data. This enables us to look at the distribution of waiting times the system spends in each state.

In order to conform Markovianity, the lifetimes of the states depicting the selectivity filters' occupation should be distributed according to an exponential distribution which is the only one without memory. While most of them do so, some strong deviations exist. Examples for both are shown shown in figure 4.5. On the left hand side a distribution matching an exponential distribution is shown, the distribution on the right hand side does not follow a single exponential distribution but rather a combination of two overlapping ones.
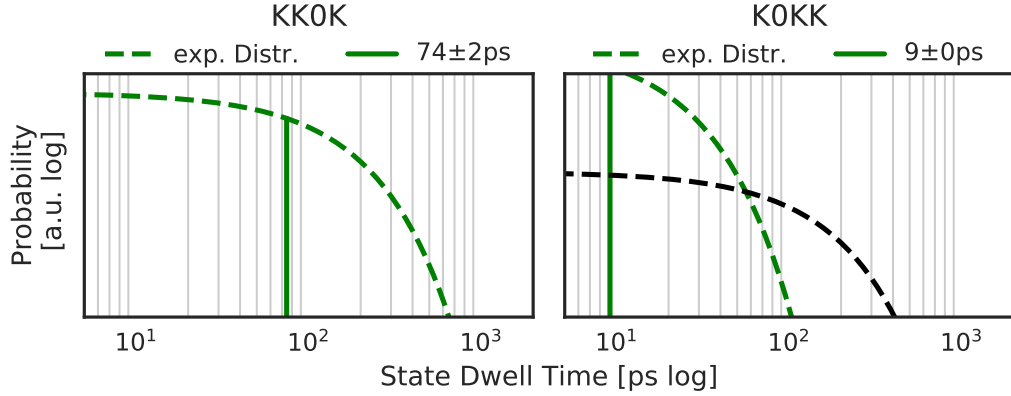
Figure 4.5.: Two generic examples for waiting time distributions, one matching the exponential distribution (l.h.s) and one not (r.h.s). The mean and fit to an exponential distribution is shown in green, the black curve shows a fit to just the waiting times longer than 100 ps. This superposition of two exponential distributions might be caused by projecting two distinct processes with identical SF occupations onto one process.

A possible explanation for the deviation is that the motion of the selectivity filter itself has not been taken into account so far. Different approaches for tackling this problem will be given in chapter 6. States with different configuration of the filter but the same occupation by K^+ ions can obviously exhibit different lifetime distributions. Projection of these distinct processes onto a common subspace will result in overlapping distributions that together do not follow a single exponential distribution. In figure 4.5, on the right side, an example for such a deviation is shown. The fact that a second exponential distribution fits to the distribution of times slower than 100 ps supports the above hypothesis of multiple overlapping processes. Figure A.3 in the ap-

pendix shows the dwell time distributions for the most prominent common SF occupation state of all channels. In those cases, where the distributions do not match an exponential distribution, the separation of timescales we assumed in section 2.2.2 is not fulfilled as the timescale of the protein movement is not on a completely different order of magnitude as the K$^+$ ions' movement.

4.3. Markov Chain Monte Carlo

Using the transition matrices we estimated in the previous section we performed Markov Chain Monte Carlo (MCMC) simulations. These provide a means to test our model for consistency. If the Markov model captures all relevant processes, artificial realizations generated according to the algorithm in section 3.4 should be indistinguishable from the permeation time series of the MD trajectories. Such artificial trajectories are shown in figure 4.6. The MCMC runs are started from the same initial state as the MD runs and conducted for the same time. In those cases where a part of a trajectory had to be omitted due to multiple occupations, the resulting trajectory pieces have been treated as individual trajectories. The starting time of these short trajectories has been shifted in figure 4.6. Tracking of the individual ion movements allows the estimation of currents with five consecutive K$^+$ ion / water jumps being equal to one full permeation event. The violin plots next to the permeation timelines in figure 4.6 show that both the mean and the distribution width of the currents are reproduced for the K$^+$ permeations. Furthermore, in agreement to the MD simulations, the water molecules show no net current in the MCMC timelines either.

Differences in the averaged currents between the channels and force fields can be tracked down on the single trajectory behaviour: These are switching between periods of permeating and stalling nature, whereas the current in the permeating periods is comparable between all systems. The trajectory overview in figure A.3 confirms this finding in a more detailed manner.
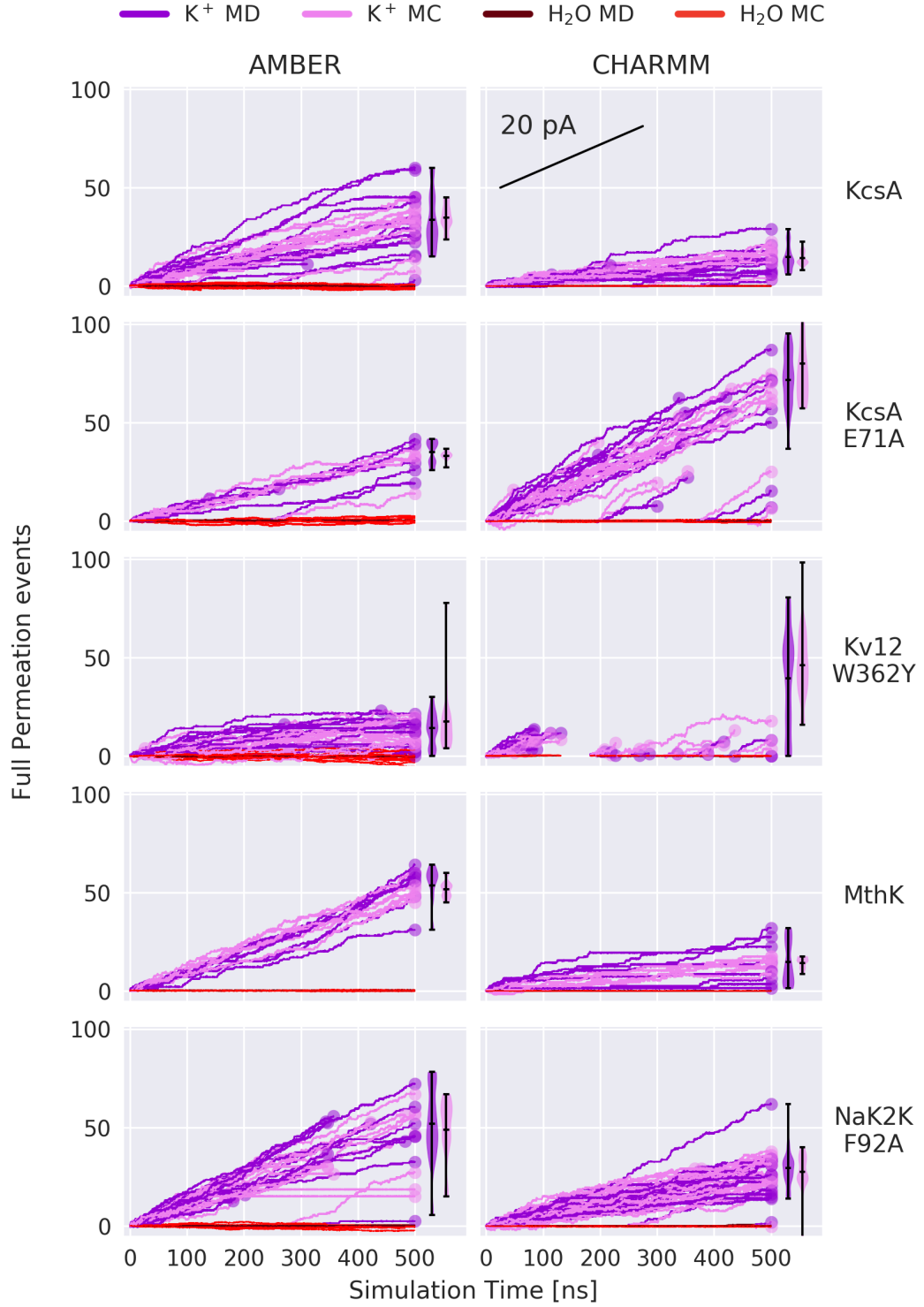
Figure 4.6.: Permeation time series for K^+ (violet) ions and water (red) molecules from both the MD (dark colors) and MC (light colors) runs. A scale bar depicting an electric current of 20 pA is shown in the upper right plot. The starting time of short trajectories has been shifted in these plots.
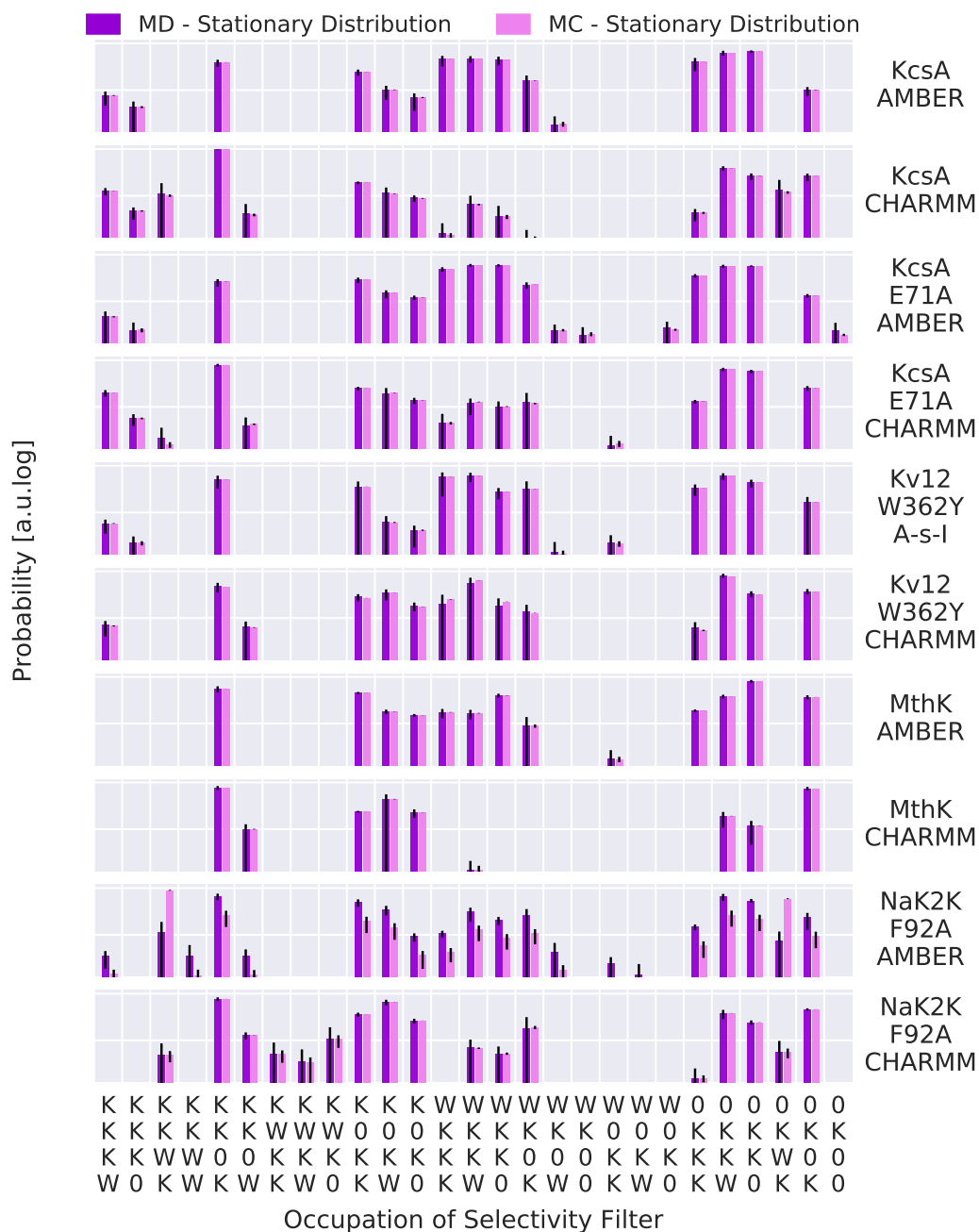
Figure 4.7.: Stationary distributions of the selectivity filters occupation states for MD trajectories (dark violet) and MC runs (light violet). A perfect agreement for all channel – force field combinations is achieved, except for NaK2K `AMBER`. This means that the MD trajectories of this system do not yet describe a steady state, which is consistent with figure 4.4.

The MCMC simulations allow to judge whether a stationary distribution has been reached in the trajectories that are used to generate the Markov model. Because both the MD and the MCMC trajectories relax into the same steady state distribution for almost all channel – force field combinations, we can conclude that the distributions are indeed stationary. In figure 4.7 both stationary distributions of the selectivity filters' occupation states for MD trajectories (dark violet) and MC runs (light violet) are shown. As expected from the implied timescales given by the eigenvalues of the transition matrices which are depicted in figure 4.4, NaK2K `AMBER` is not equilibrated yet. However, this does not implicate that the permeation of this system cannot be analyzed. This slow process might instead represent an artifact introduced by the splitting of trajectories.
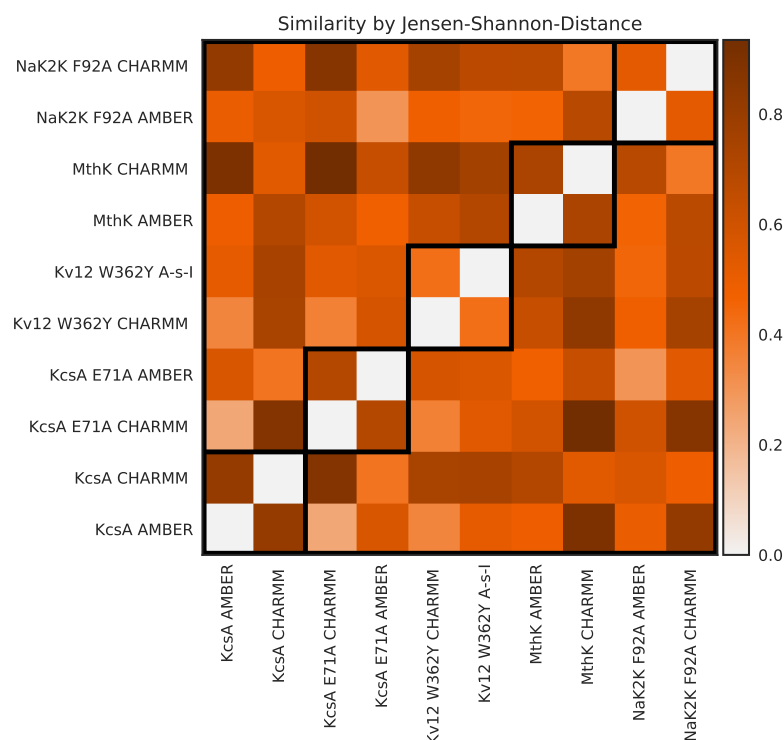
Figure 4.8.: A symmetric version of the Kullback-Leibler divergence allows to define similarity metric. This metric is applied here onto the stationary distributions of the MD trajectories, pairwise for all channels and force fields. The squares on the diagonal indicate the same protein.

A quantitative method to analyze the difference between each of these stationary distributions is the Jensen-Shannon distance, the square root of the Jensen–Shannon divergence [FT04], which again is a symmetrized version of the Kullback–Leibler divergence [Joy11]. The Kullback-Leibler divergence is a commonly used measure of the divergence of one probability distribution from another. It does not take into account the ordering of the states, which is appropriate here. We apply the Jensen-Shannon distance pairwise for all channels and force fields onto the stationary distributions of the MD trajectories. Figure 4.8 shows the result. The black squares following the diagonal indicate distances between both force field versions of the same protein. On the diagonal the distance is zero as it measures the distance between the same distributions. The tri – diagonal elements in the squares show lower values than the outer off diagonal elements, revealing that the force field has a bigger effect on the stationary distributions than the proteins have.

4.4. Permeation Cycles

As introduced in section 3.5, permeation is a reoccurring process that produces repeated sequences in the discretized trajectory. Repeated states and trivial cycles (oscillations between two states) that do not lead to net transport of ions, are not taken into account. The resulting trajectory was decomposed using the cycle decomposition algorithm from section 3.5.

The results of this decomposition are shown in figure 4.10 on the left hand side. Each cycle that occurred in the trajectories is drawn in the corresponding network with a unique color. The thickness of the arrows scales linearly with the frequency this cycle has occurred. The colors distinguish between the cycles throughout one single graph, as there are about 250 unique cycles found in total. For the sake of clarity, all these plots just show the 80% most frequent cycles. The states (drawn as orange nodes) are located at the same positions for all different systems allowing a comparison among them. The nodes shown in the graph correspond to all states that the particular system visited during all trajectories, the sizes are drawn according to the stationary distribution using a logarithmic scaling. The states are arranged in a way that the majority of

the cycles describes clockwise cycles in the networks. Similar states regarding their K^+ positions in the binding sites are arranged next to each other.

One directly recognizes that of the 27 states that are visited in total only a few participate in the permeation itself. Another thing striking the eye is the higher diversity among some graphs, almost dividing them into two groups of complicated and simple graphs. To the group of complicated graphs belong the `AMBER` versions of KcsA and MthK as well as the `CHARMM` version of NaK2K. The other channels do not exhibit such a high variety of cycles in their graphs. This higher diversity is caused by additional states that have S1 or S4 occupied by water, e. g. WKK0, 0KKW and WKKW. These states do not occur in the 'simple' graphs. The integration of these states into the network combinatorically allows for a multitude of additional cycles resulting in more complex graphs, These additional cycles do not differ noticeably from their simpler counterpart in regard of the potassium movement along their edges. The common feature is an 'in and out' movement of water in the outer binding sites S1 and S4 either before a K^+ ion moved or afterwards.

Incorporating this information into a clustering of states results in the networks on the right hand side of figure 4.10. States are clustered together if they do not differ in their K^+ ion occupation (e. g. 0KK0, WKK0, 0KKW, WKKW into xKKx) where x stands either for a vacancy or a water molecule. These are the states that build clusters in the transition matrices too, as described in section 4.2.

In the resulting graphs a simplified picture emerges, see figure 4.10, r.h.s..

We recognize that one main permeation cycle exists. It consists purely of single ion jumps and comes in two small variations, which overlap in most of the edges except for xKKK and K0Kx:

$$KK0K \uparrow xK0K \rightarrow xKKx \uparrow xKKK \rightarrow K0KK \rightarrow KK0K$$

$$KK0K \uparrow xK0K \rightarrow xKKx \rightarrow K0Kx \uparrow K0KK \rightarrow KK0K$$
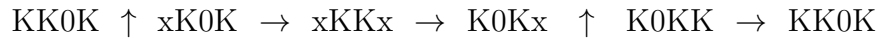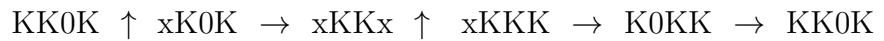
The arrow pointing to the top $\uparrow$ marks where an ion enters or leaves the SF and the arrow to the right $\rightarrow$ marks movements of K^+ ions inside the
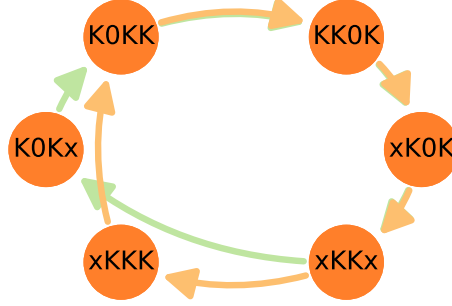
Figure 4.9.: The two variations of the most basic permeation cycle. A K$^+$ ion enters the channel along the edge of xKKx to xKKK or from K0Kx to K0KK into S4. The ion leaves S1 from state KK0K to xK0K.

SF. The difference between these two variations lies in the amount of direct contacts between K$^+$ ions. The first one allows for three K$^+$ ions in S2, S3 and S4 before the one from S2 jumps to S1 while the second one shows the jump from S2 to S1 before a new K$^+$ ion enters into S4. All of the clustered cycles shown in the following figure 4.10 are plotted in the same honeycomb shape as the basic permeation cycle in figure 4.9. Each system then prefers different shortcuts through these cycles, thereby combining several K$^+$ ion jumps in one transition. Whether the ion jumps during these shortcuts do indeed occur simultaneously or successively cannot be distinguished with the resolution of the model. In chapter 6 we discuss a method to test for that. It could very well be, that all of the studied channels do perform the same cycle and the above mentioned shortcuts with simultaneous K$^+$ ion movements disappear at higher time resolutions.

KcsA AMBER
342 of 427 Permeations (80%)
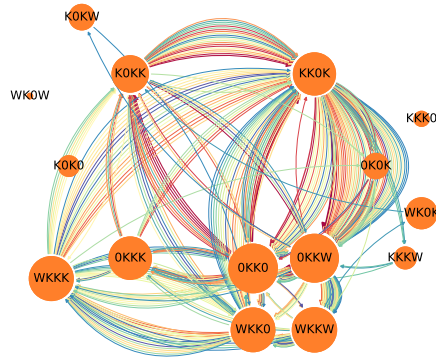
KcsA AMBER
342 of 427 Permeations (80%)
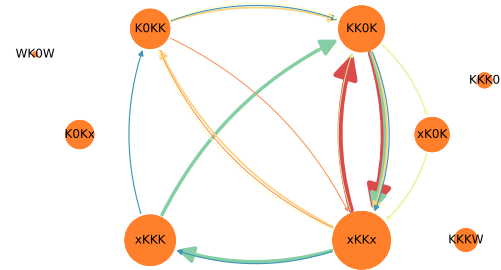
KcsA CHARMM
158 of 196 Permeations (81%)
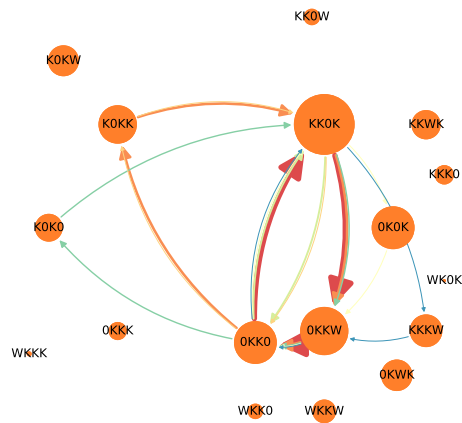
KcsA CHARMM
160 of 196 Permeations (82%)
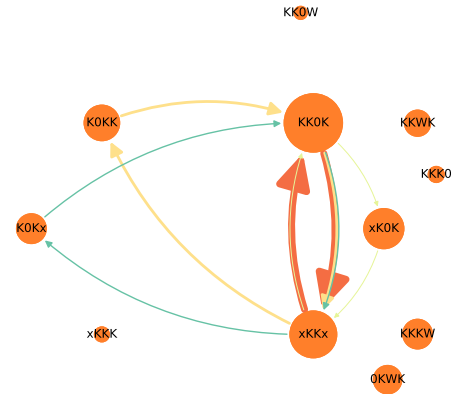
KcsA E71A AMBER
144 of 180 Permeations (80%)

KcsA E71A AMBER
144 of 180 Permeations (80%)

4. Results and Discussion

KcsA E71A CHARMM
585 of 725 Permeations (81%)

KcsA E71A CHARMM
595 of 725 Permeations (82%)

Kv12 W362Y A-s-I
165 of 205 Permeations (80%)

Kv12 W362Y A-s-I
168 of 205 Permeations (82%)

Kv12 W362Y CHARMM
104 of 130 Permeations (80%)

Kv12 W362Y CHARMM
106 of 130 Permeations (82%)
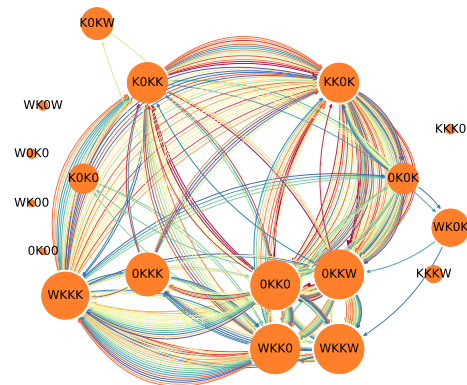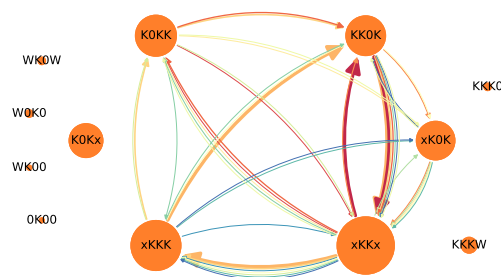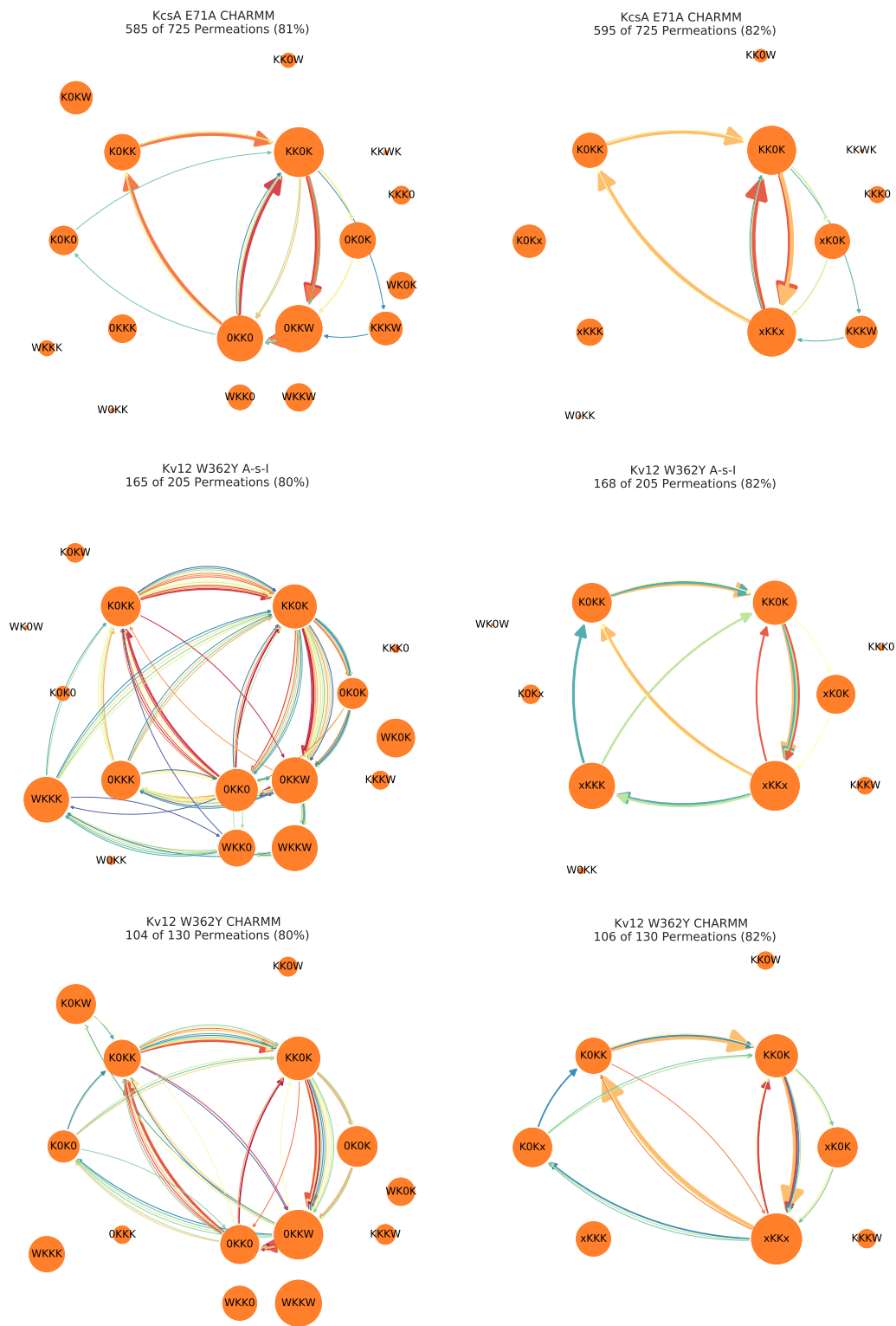
MthK AMBER
301 of 374 Permeations (80%)
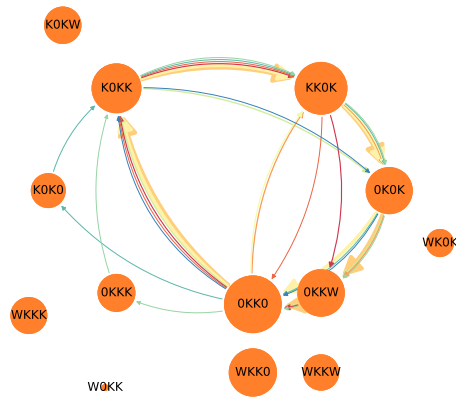
MthK AMBER
303 of 374 Permeations (81%)
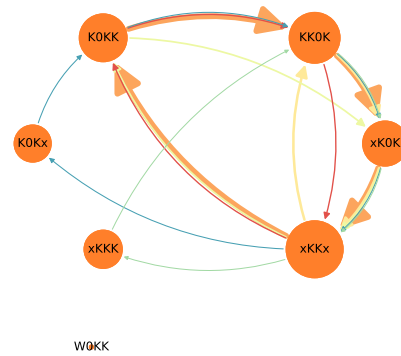
MthK CHARMM
93 of 116 Permeations (80%)
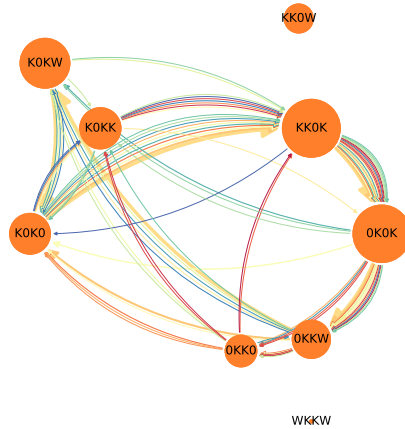
MthK CHARMM
93 of 116 Permeations (80%)
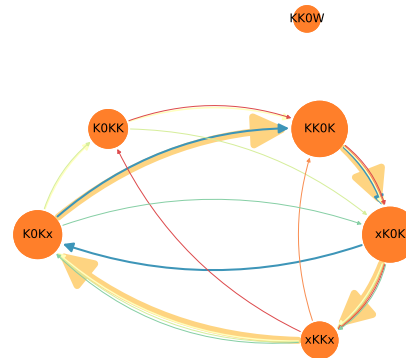
NaK2K F92A AMBER
500 of 625 Permeations (80%)
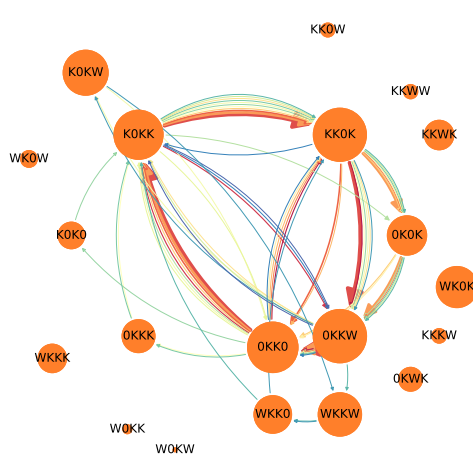
NaK2K F92A AMBER
505 of 625 Permeations (81%)
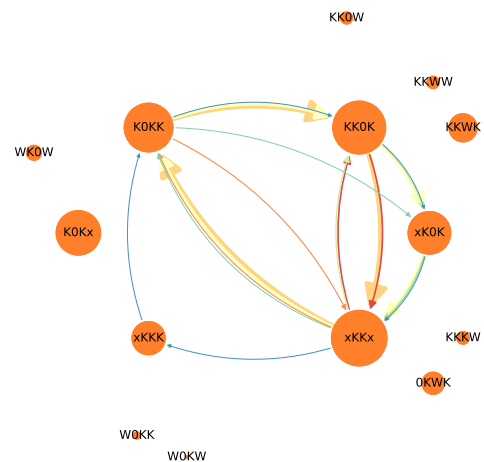
4. Results and Discussion
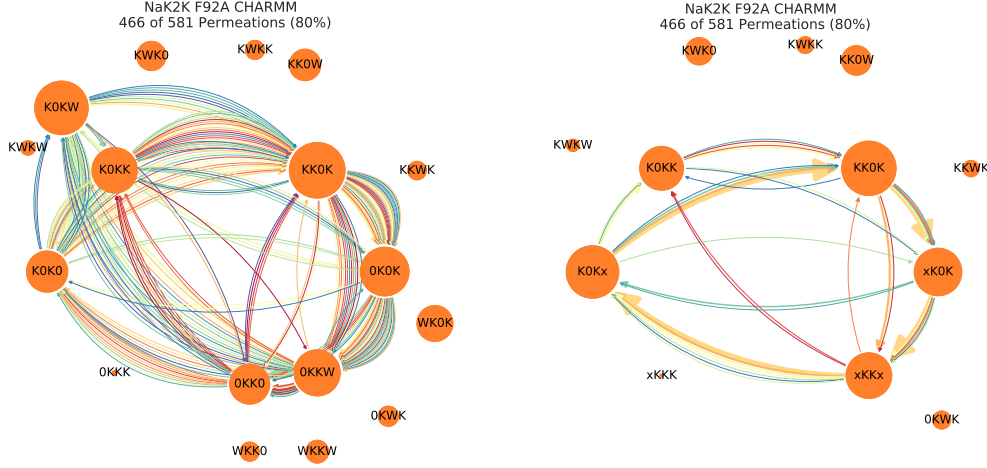
Figure 4.10.: The permeation cycles that occurred during the simulation. On the left hand side the full number of states is shown while on the right hand side states with the same K^+ occupation are clustered together.

The shortest possible cycles observed with a resolution of $\Delta t = 40\,\text{ps}$ are those moving 5 K^+ ions in one step, returning and starting at the same state. This happens extremely rarely (less than 1 ‰) and does not represent a typical cycle. If such cycles would happen more often, the output frequency of the simulation had to be increased.

The most frequent cycles one can find in figure 4.10 in the clustered versions on the r.h.s. are described in the following. The graphs in figure 4.10 are referenced via the channel – force field combination.

$$\text{KK0K} \;\; 2\uparrow \;\; \text{xKKx} \;\; 3\uparrow \;\; \text{KK0K},$$

which occurs mainly in KcsA both with and without mutation, to a smaller amount in Kv1.2 and not at all in MthK and NaK2K. The cycle

$$\text{KK0K} \;\; 2\uparrow \;\; \text{xKKx} \;\; 2\uparrow \;\; \text{K0KK} \;\; 1\uparrow \;\; \text{KK0K},$$

is observed in all KcsA variants, but more pronounced in the `CHARMM` versions. It also occurs frequently in Kv1.2 and NaK2K, for both in the `AMBER` version.

The cycle

$$\text{KK0K} \ 1\uparrow \ \text{xK0K} \ 1\uparrow \ \text{xKKx} \ 2\uparrow \ \text{K0KK} \ 1\uparrow \ \text{KK0K},$$

is just found in the `AMBER` versions of MthK and NaK2K, and

$$\text{KK0K} \ 1\uparrow \ \text{xK0K} \ 1\uparrow \ \text{xKKx} \ 1\uparrow \ \text{K0Kx} \ 2\uparrow \ \text{KK0K},$$

in MthK and NaK2K both in the `CHARMM` variant.

MthK and NaK2K prefer for both force field variants not to skip the first intermediate state xK0K, whereas this shortcut occurs in most of the other channels.
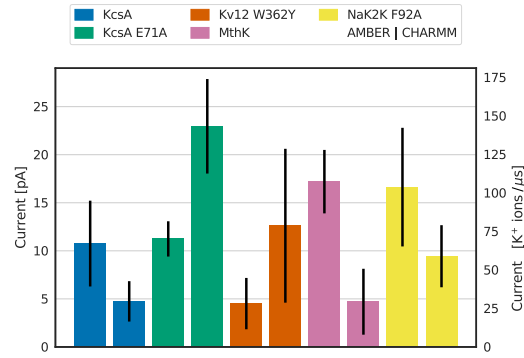
Figure 4.11.: The current per channel and force field.

It is interesting to see that there is no obvious correlation between the averaged current of the different channels, as figure 4.11 depicts, and their cycles' typical number of transitions. As already formulated in section 4.3, the differences in the current are caused by stalling periods in the trajectories where no ion motion is observed. These periods vary in length between both channels and force fields. This finding is confirmed by the distribution of dwell times which varies remarkably between the same states in different simulation systems, compare figure A.3.

4. Results and Discussion
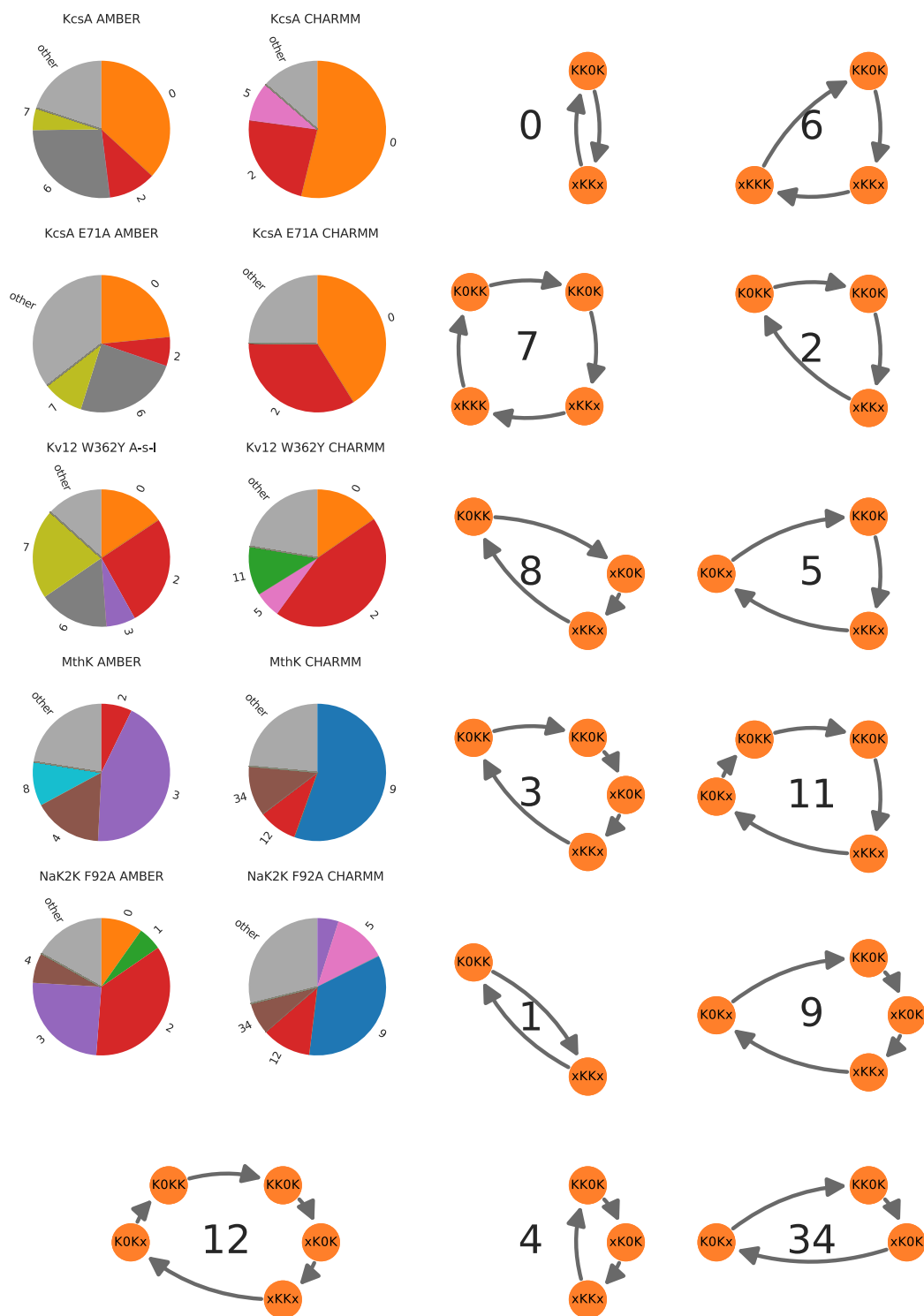
Figure 4.12.: A pie chart showing the occupancy of the cycles for different channels and force fields on the left hand side and next to the most frequent cycles. The wedge 'other' combines cycles with contributing less than 1 % to all permeations.

Figure 4.12 depicts the cycle occupations of the different channels as well as graphs of the most frequent cycles together with the respective id referenced in the pie charts' wedges. We see a large similarity in form of common cycles (0, 2) for KcsA, KcsA E71A and Kv1.2, which are the cycles combining the highest number of ion jumps in single transitions, namely double and triple jumps of K^+. The same holds for the CHARMM versions of MthK and NaK2K with cycles 9, 12, 34 and for the AMBER versions of MthK and NaK2K with cycles 2, 3, 4. The AMBER versions of KcsA, KcsA E71A and Kv1.2 do in addition have cycle 7 in common.

We can conclude that there exists a general theme behind the potassium permeation mechanism which is the cycle depicted in 4.9. It is a very robust mechanism because it is in its main theme conserved through all channels and force fields. Different channels take different shortcuts through the main permeation cycle and thereby combine multiple transitions and their K^+ jumps into one single transition. The difference that lies between the channels is whether the first shortcut, leaving out state xK0K, is taken or not. KcsA, KcsA E71A and Kv1.2 prefer not to visit this state while MthK and NaK2K do so. Whether state xKKK is skipped or visited depends on the force field, the AMBER versions of the former three channels do all include this state while most of the CHARMM versions skip it. This means that AMBER renders states with three adjacent binding sites occupied with K^+ more accessible than CHARMM. KK0K and K0KK are states that seem to be mandatory for all permeation cycles as all channel – force field combinations visit them.

4.4.1. Number of ions jumps per transition

Extracting quantitative and humanly comprehensible differences between the cycles each system visits is not a straightforward task due to the high diversity even after the clustering. Guided by the idea of similar shortcuts trough the main cycle, we found a characteristic property that we could extract from the permeation cycles of each of the systems. Focusing on the transitions of a system, we looked for the number of K^+ ions that are moved during each transition. Collecting the ratios of these transitions and multiplying them with

the amount of K^+ ions transported, gives a relation telling how important these different transition types are for the total K^+current. These values are not independent, as e. g. double jumps do always occur together with either single or triple jumps in order to reach five jumps in total. Figure 4.13 shows these relations.
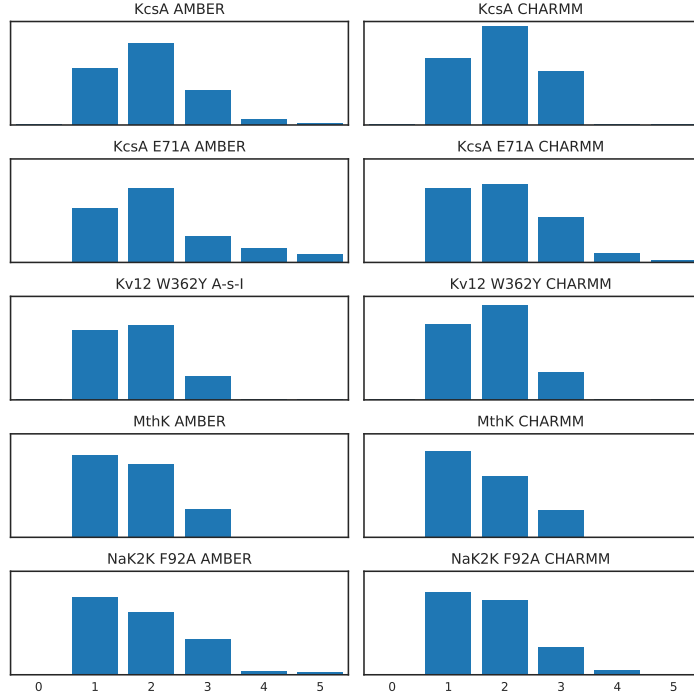
Figure 4.13.: Sorting transitions by the number of moved K^+ ions and scaling that ratio with the same number gives an idea of the importance of different transition types for the permeation in the corresponding channel. These relations are plotted here, the y-axis is the same for all variants and no units are shown as just the relation is important.

Similarities can be found between channels as well as on a force field level: Comparing all channels reveals that transitions with one or two transported K^+ ions are most prominent. Transitions with triple ion jumps do also occur in all variants but are less pronounced. Only in KcsA E71A transitions with more than three K^+ ions do contribute notably to the current. Both force field versions of KcsA, KcsA E71A and Kv1.2 have the biggest contribution from transitions with two K^+ ion jumps, MthK and NaK2K from single jump transitions. For higher numbers of transported K^+ ions, the contributions are

descending. The relations between the transition contributions show similarities between KcsA, KcsA E71A and Kv1.2 on the one hand and between MthK and NaK2K on the other hand. This fits to the impression one got from the force field – independent similarities between the channels MthK and NaK2K on the one hand and KcsA, KcsA E71A and Kv1.2 on the other hand.

Performing the same analysis using the Jensen-Shannon-Distance as in section 4.3 on the distribution of ion jumps per transition we can validate these similarities. Contrary to the high distances the stationary distributions had between each other for the same channels with different force fields, figure 4.14 shows smaller distances on the tri-diagonal elements in the rectangles. Especially the two force field versions of each Kv1.2, MthK and NaK2K do not differ severely from each other. Additionally we find that these three channels show small distances among each other, as is depicted by a dashed rectangle in figure 4.14. This was not identifiable from the graphs themselves.
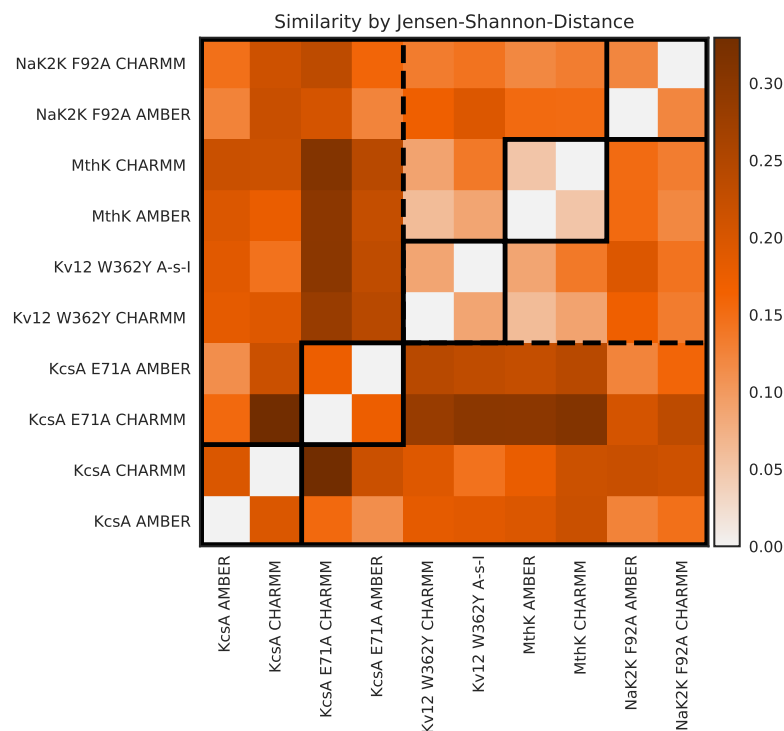
Figure 4.14.: The same analysis as for figure 4.8 was performed on the distribution of ion jumps per transition, thereby validating the similarities described in figure 4.13.

4.5. Voltage drop across selectivity filter

Calculating the electric potential in a simulation box is done by integrating twice over the spatial charge distribution. This is a difficult task as the high fluctuations that occur in the positions of the charged molecules lead to extreme variations in the voltage, which can just be overcome with extensive averaging over simulation time. The voltage drop across membranes with and without pores has been estimated for simulations with applied electric field in [GKASR12]. There it was shown that the voltage stays constant in the solution and drops over the membrane, as expected by physical intuition. Even in water filled cavities the voltage remains constant. This is in particular relevant to this work, since the channels examined here show filled cavities as well.

Using a reformulation of equation 2.19, we obtain a fundamentally different approach to estimate the voltage drop that arises over the length of the selectivity filter. The local detailed balance condition gives access to the free energy difference between two states, that equation can be combined for all the edges of the occurring cycles. Equation 4.1 allows the calculation of the free energy difference between both sides of the SF, by multiplying the quotient of forward and backward rates for each edge along a cycle,

$$\ln \prod_{\text{cycle}} \frac{w_{mn}}{w_{nm}} = (qU)/k_{\text{B}}T. \tag{4.1}$$

Assuming that the compartments on both sides of the SF do not differ in any other aspect but the voltage, the free energy difference is directly proportional to the voltage difference U with a factor of the charge of the transported ion, in this case $q = +1\,e^-$ for a K$^+$ ion. Figure 4.15 shows the voltage difference in units of $k_{\text{B}}T$ and mili Volt for all channels and force fields. All cycles that lead to permeations of K$^+$ ions are taken into account and the average is weighted by their individual counts.
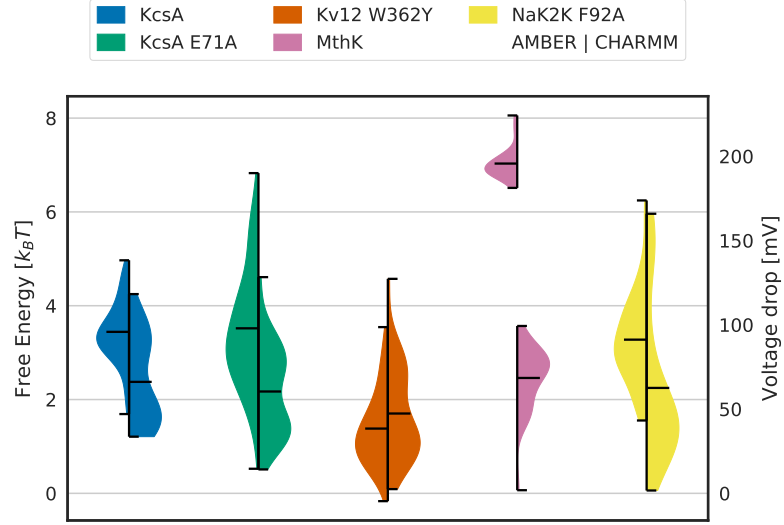
Figure 4.15.: Using the local detailed balance relation for all subsequent transitions of the occurring cycles we can estimate the free energy difference associated with the transport of one K$^+$ ion trough the selectivity filter. This violin plot compares the distributions of values for the free energy difference among all channels and force fields.

Most of the values for the free energy difference lie in the range of 1–5 $k_\mathrm{B}T$ (30–140 mV). The only exception is MthK `AMBER`, showing a free energy difference of 7 $k_\mathrm{B}T$ with a significantly narrower distribution. A possible explanation for this behaviour could be, that it does not suffer from the following shortcoming. One has to consider that the voltage drop is always underestimated due to the overestimation of transition counts. This overestimation is caused by recrossing events, attempted transitions that directly return to the state of origin and are therefore excess counts of both forward and backward transitions. In the equation, this leads to a ratio closer to one, the logarithm of which is then closer to zero.

With regard to the applied voltage of ($U \approx 220 - 280\,\mathrm{mV}$) over the full simulation box these results are in perfect agreement with the expected result, which is that most of the voltage should drop along the channel with a significant part along the selectivity filter which spans about one third of the length of the protein.

4. Results and Discussion

5. Conclusion

In this thesis we could show that the permeation of K^+ ions trough the selectivity filter of different K^+ channels can be described with a Markov State Model. We estimated transition matrices based on the discretized motion of K^+ ions in the SF under applied voltage. After confirming Markovianity for most transitions we applied Markov Chain Monte Carlo sampling to verify the reaching of a steady state for the Molecular Dynamics trajectories. Except for one exception, NaK2K `AMBER`, this analysis did not suffer from insufficient sampling.

The probability flux cycles that emerge in a non–equilibrium steady state system have been decomposed into cycles using the information in the discretized trajectories. A further clustering of the cycles revealed a common permeation mechanism for all of the channels, of which different variations are realized by different channels. The variations that occur differ both between the channels as well as the force fields.

It was not possible to track down these differences to the protein background stabilizing the selectivity filter because the differences between the force fields of the same channel are of the same magnitude as the inter–channel distances. However, the dualism between the similarity of the signaling sequence (section 1.1.1) and the selectivity filters occupation states (section 4.1) versus the differences in the proteins background (section 1.1.1) and the states occupation times (section 4.3) of the different cycles (section 4.4) give a consistent impression.

5. Conclusion

60

6. Outlook

The analysis performed can be extended in several ways.

For a time resolution of $\Delta t = 40\,\text{ps}$, some channels shortcut transitions during permeation. A first question arising from this finding is, whether this behaviour is maintained at higher time resolutions or if it is an artifact of the chosen simulation output frequency. It would be interesting to spawn new simulations with a higher output frequency to address this question.

So far the co–permeation of water was excluded from the analysis because it seems to follow a significantly different pathway trough the channel. Additional states that allow for multiple occupations per binding site could cope with the states visited during water co–permeation. Such states could be introduced 'manually', as an extension of the existing 81 states or with a completely different approach of clustering states in the 3N dimensional configuration space with N being the number of all K^+ ions and water molecules that permeate the channel.

A more straightforward extension would be to analyze the permeation patterns of different ions, e. g. Na^+ ions. Thereby we could extend the findings which explain the selectivity in these channels [KKV$^+$18]. Simulations of these are available. Sodium has been shown to block the permeation in potassium channels in both experiments [FW77] and simulations [KKV$^+$18]. However, enough permeations have been observed to encourage further analysis.

In this work, only four Channels were examined. However, there exists a huge variety of channels similar to the analyzed ones with respect to the geometry of the selectivity filter. Some of these exhibit only two or three binding sites. The developed formalism could be easily extended to incorporate these channels, provided that a discrete description of their permeation processes seems rationale.

6. Outlook

As noticed in the analysis of dwell times in different states, several processes do overlap in some cases. These processes are out of reach in the current subspace of the phase space but a first idea to search for these is in the motion of the selectivity filter. A state of the art method for modeling full protein trajectories as Markov models is `pyEMMA` [STSP$^+$15]. Unpublished data from Paul and Noe allows the hypothesis that the selectivity filter itself undergoes various transitions.

A possibility to combine the `pyEMMA` approach with the analysis conducted in this thesis is to include the states of the selectivity filter in form of SF state dependent transition rates in our model. The assumption therein is that the phase space of permeation stays the same.

The definition of an additional macro–state for a collapsed SF would hereby provide a way to avoid the splitting of trajectories.

To overcome the problem of overestimated transition counts due to the re-crossing problem, the introduction of core sets is common in the MSM community [LK16]. This could also be implemented as an extension.

Bibliography

[AMS⁺15] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd
Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs:
High performance molecular simulations through multi-level par-
allelism from laptops to supercomputers. *SoftwareX*, 1:19–25,
2015.

[Bat01] S. S. Batsanov. Van der waals radii of elements. *Inorganic
Materials*, 37(9):871–885, Sep 2001.

[BBO⁺83] Bernard R Brooks, Robert E Bruccoleri, Barry D Olafson,
David J States, S a Swaminathan, and Martin Karplus.
Charmm: a program for macromolecular energy, minimization,
and dynamics calculations. *Journal of computational chemistry*,
4(2):187–217, 1983.

[BH09] Robert B Best and Gerhard Hummer. Optimized molecular dy-
namics force fields applied to the helix- coil transition of polypep-
tides. *The journal of physical chemistry B*, 113(26):9004–9015,
2009.

[BM12] Manasi P Bhate and Ann E McDermott. Protonation state of
E71 in KcsA and its role for channel collapse and inactivation.
Proceedings of the National Academy of Sciences, 2012.

[Bol64] Ludwig Boltzmann. Vorlesungen über Gastheorie, 1923, 2 vols.
leipzig: Barth, 1896, 1898. This book has been translated into
English by S.G. Brush, Lectures on Gas Theory, 1964.

Bibliography

[BR01]   Simon Berneche and Benoit Roux. Energetics of ion conduction through the K+ channel. *Nature*, 414(6859):73, 2001.

[BR03]   Simon Bernèche and Benoît Roux. A microscopic view of ion conduction through the K+ channel. *Proceedings of the National Academy of Sciences*, 100(15):8644–8648, 2003.

[BT06]   Julius Bernstein and A Tschermak. Untersuchungen zur Thermodynamik der bioelektrischen Ströme. *Archiv für die gesamte Physiologie des Menschen und der Tiere*, 112(9-10):439–521, 1906.

[BWF+06]   Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank, 1999–. In *International Tables for Crystallography Volume F: Crystallography of biological macromolecules*, pages 675–684. Springer, 2006.

[BZS+12]   Robert B Best, Xiao Zhu, Jihyun Shim, Pedro EM Lopes, Jeetain Mittal, Michael Feig, and Alexander D MacKerell Jr. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi1$ and $\chi2$ dihedral angles. *Journal of chemical theory and computation*, 8(9):3257–3273, 2012.

[CJC+10]   Luis G Cuello, Vishwanath Jogini, D Marien Cortes, Albert C Pan, Dominique G Gagnon, Olivier Dalmas, Julio F Cordero-Morales, Sudha Chakrapani, Benoît Roux, and Eduardo Perozo. Structural basis for the coupling between activation and inactivation gates in k+ channels. *Nature*, 466(7303):272, 2010.

[CJCP10]   Luis G Cuello, Vishwanath Jogini, D Marien Cortes, and Eduardo Perozo. Structural mechanism of c-type inactivation in k+ channels. *Nature*, 466(7303):203, 2010.

[CMT+11]   Wayland WL Cheng, Jason G McCoy, Ameer N Thompson, Colin G Nichols, and Crina M Nimigean. Mechanism for

selectivity-inactivation coupling in kcsa potassium channels. *Proceedings of the National Academy of Sciences*, 2011.

[DCP⁺98] Declan A Doyle, Joao Morais Cabral, Richard A Pfuetzner, Anling Kuo, Jacqueline M Gulbis, Steven L Cohen, Brian T Chait, and Roderick MacKinnon. The structure of the potassium channel: molecular basis of k+ conduction and selectivity. *science*, 280(5360):69–77, 1998.

[DSZ⁺11] Mehabaw G Derebe, David B Sauer, Weizhong Zeng, Amer Alam, Ning Shi, and Youxing Jiang. Tuning the ion selectivity of tetrameric cation channels by changing the number of ion binding sites. *Proceedings of the National Academy of Sciences*, 108(2):598–602, 2011.

[DTGG11] Serena Donnini, Florian Tegeler, Gerrit Groenhof, and Helmut Grubmüller. In Silico Titration of Biomolecules: Explicit Solvent Constant pH Molecular Dynamics. *Biophysical Journal*, 100(3):613a, 2011.

[EML⁺05] Elan Z Eisenmesser, Oscar Millet, Wladimir Labeikovsky, Dmitry M Korzhnev, Magnus Wolf-Watz, Daryl A Bosco, Jack J Skalicky, Lewis E Kay, and Dorothee Kern. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438(7064):117, 2005.

[FHB99] K Anton Feenstra, Berk Hess, and Herman JC Berendsen. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry*, 20(8):786–798, 1999.

[FS02] D. Frenkel and B. Smit. Understanding molecular simulation: from algorithms to applications, 2002.

[FT04] Bent Fuglede and Flemming Topsoe. Jensen-shannon divergence and hilbert space embedding. In *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, page 31. IEEE, 2004.

Bibliography

[FW77] Robert J French and Jay B Wells. Sodium ions as blocking agents and charge carriers in the potassium channel of the squid giant axon. *The Journal of general physiology*, 70(6):707–724, 1977.

[GKASR12] James Gumbart, Fatemeh Khalili-Araghi, Marcos Sotomayor, and Benoît Roux. Constant electric field simulations of the membrane potential illustrated with simple systems. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1818(2):294 – 302, 2012. Membrane protein structure and function.

[Gro13] Gerrit Groenhof. Introduction to QM/MM simulations. *Methods in Molecular Biology*, 924:43–66, 2013.

[HAO+06] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.

[HH39] Alan L Hodgkin and Andrew F Huxley. Action potentials recorded from inside a nerve fibre. *Nature*, 144(3651):710, 1939.

[Hil05] Terrell L Hill. *Free energy transduction and biochemical cycle kinetics*. Courier Corporation, 2005.

[HK55] Al L Hodgkin and RD Keynes. The potassium permeability of a giant nerve fibre. *The Journal of Physiology*, 128(1):61–88, 1955.

[HKM08] Ramesh Hariharan, Telikepalli Kavitha, and Kurt Mehlhorn. Faster algorithms for minimum cycle basis in directed graphs. *SIAM Journal on Computing*, 38(4):1430–1447, 2008.

[HKVDSL08] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.

[HLAM94] Lise Heginbotham, Zhe Lu, Tatiana Abramson, and Roderick MacKinnon. Mutations in the k+ channel signature sequence. *Biophysical journal*, 66(4):1061–1067, 1994.

[HMN$^+$81] Owen P Hamill, A Marty, Erwin Neher, Bert Sakmann, and FJ Sigworth. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Archiv*, 391(2):85–100, 1981.

[HRN$^+$16] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L de Groot, Helmut Grubmüller, and Alexander D MacKerell Jr. Charmm36m: an improved force field for folded and intrinsically disordered proteins. *nature methods*, 14(1):71, 2016.

[IOTS10] Shunsuke Imai, Masanori Osawa, Koh Takeuchi, and Ichio Shimada. Structural basis underlying the dual gate properties of kcsa. *Proceedings of the National Academy of Sciences*, 107(14):6216–6221, 2010.

[JCM$^+$83] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[JJES13] Morten Ø Jensen, Vishwanath Jogini, Michael P Eastwood, and David E Shaw. Atomic-level simulation of current–voltage relationships in single-file ion channels. *The Journal of general physiology*, 141(5):619–632, 2013.

[Joy11] James M Joyce. Kullback-leibler divergence. In *International encyclopedia of statistical science*, pages 720–722. Springer, 2011.

[JQQ03] Da-Quan Jiang, Min Qian, and Ming-Ping Qian. *Mathematical theory of nonequilibrium steady states: on the frontier of probability and dynamical systems*. Springer, 2003.

Bibliography

[Kal07] Sophia L Kalpazidou. *Cycle representations of Markov processes*, volume 28. Springer Science & Business Media, 2007.

[KGDZ11] Carsten Kutzner, Helmut Grubmüller, Bert L. De Groot, and Ulrich Zachariae. Computational electrophysiology: The molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophysical Journal*, 101(4):809–817, 2011.

[Kir47] Gustav Kirchhoff. Ueber die auflösung der gleichungen, auf welche man bei der untersuchung der linearen vertheilung galvanischer ströme geführt wird. *Annalen der Physik*, 148(12):497–508, 1847.

[KKV+18] Wojciech Kopec, David A Köpfer, Owen N Vickery, Anna S Bondarenko, Thomas LC Jansen, Bert L de Groot, and Ulrich Zachariae. Direct knock-on of desolvated ions governs strict ion selectivity in k+ channels. *Nature Chemistry*, 10(8):813, 2018.

[KLS83] Sheldon Katz, Joel L Lebowitz, and H Spohn. Phase transitions in stationary nonequilibrium states of model lattice systems. *Physical Review B*, 28(3):1655, 1983.

[KM02] M Karplus and J McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9):646–652, 2002.

[KSG+14] David A. Köpfer, Chen Song, Tim Gruene, George M. Sheldrick, Ulrich Zachariae, and Bert L. de Groot. Ion permeation in k+ channels occurs by direct coulomb knock-on. *Science*, 346(6207):352–355, 2014.

[Kub66] Rep Kubo. The fluctuation-dissipation theorem. *Reports on progress in physics*, 29(1):255, 1966.

[KVF+10] Jeffery B Klauda, Richard M Venable, J Alfredo Freites, Joseph W O'Connor, Douglas J Tobias, Carlos Mondragon-Ramirez, Igor Vorobyov, Alexander D MacKerell Jr, and Richard W Pastor. Update of the charmm all-atom additive

force field for lipids: validation on six lipid types. *The journal of physical chemistry B*, 114(23):7830–7843, 2010.

[LHM01] Meredith LeMasurier, Lise Heginbotham, and Christopher Miller. KcsA: it's a potassium channel. *The Journal of general physiology*, 118(3):303–314, 2001.

[LK16] Oliver Lemke and Bettina G Keller. Density-based cluster algorithms for the identification of core sets. *The Journal of chemical physics*, 145(16):164104, 2016.

[LLPP+10] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

[MCZM01] João H Morais-Cabral, Yufeng Zhou, and Roderick MacKinnon. Energetic optimization of ion conduction rate by the k+ selectivity filter. *Nature*, 414(6859):37, 2001.

[MN12] Jason G McCoy and Crina M Nimigean. Structural correlates of selectivity and inactivation in potassium channels. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1818(2):272–285, 2012.

[MO70] P Mazur and I Oppenheim. Molecular theory of brownian motion. *Physica*, 50(2):241–258, 1970.

[Mor65] Hazime Mori. Transport, collective motion, and brownian motion. *Progress of theoretical physics*, 33(3):423–455, 1965.

[MRR+53] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

Bibliography

[MSO82] U Mohanty, KE Shuler, and I Oppenheim. On the exact and phenomenological langevin equations for a harmonic oscillator in a fluid. *Physica A: Statistical Mechanics and its Applications*, 115(1-2):1–20, 1982.

[NGB⁺07] Heike Neubauer, Natalia Gaiko, Sylvia Berger, Jörg Schaffer, Christian Eggeling, Jennifer Tuma, Laurent Verdier, Claus AM Seidel, Christian Griesinger, and Andreas Volkmer. Orientational and dynamical heterogeneity of rhodamine 6g terminally attached to a dna helix revealed by nmr and single-molecule fluorescence spectroscopy. *Journal of the American Chemical Society*, 129(42):12746–12755, 2007.

[Phi13] Rob Phillips. *Physical biology of the cell.* Garland Science, London : New York, NY, second edition edition, 2013.

[PPS⁺13] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R Shirts, Jeremy C Smith, Peter M Kasson, David Van Der Spoel, et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 2013.

[Pur77] Edward M Purcell. Life at low reynolds number. *American journal of physics*, 45(1):3–11, 1977.

[PWS⁺11] Jan-Hendrik Prinz, Hao Wu, Marco Sarich, Bettina Keller, Martin Senne, Martin Held, John D Chodera, Christof Schütte, and Frank Noé. Markov models of molecular kinetics: Generation and validation. *The Journal of chemical physics*, 134(17):174105, 2011.

[RABI04] Benoit Roux, Toby Allen, Simon Berneche, and Wonpil Im. Theoretical and computational models of biological ion channels. *Quarterly reviews of biophysics*, 37(1):15–103, 2004.

[RK91]   Benoit Roux and Martin Karplus. Ion transport in a gramicidin-like channel: dynamics and mobility. *The Journal of Physical Chemistry*, 95(12):4856–4868, 1991.

[Röp87]   Gerd Röpke. *Statistische Mechanik für das Nichtgleichgewicht.* Deutscher Verlag der Wissenschaften, 1987.

[Rou08]   Benoît Roux. The membrane potential and its representation by a constant electric field in computer simulations. *Biophysical Journal*, 95(9):4205–4216, 2008.

[Sch26]   E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, 28(6):1049–1070, 1926.

[Sch76]   Jürgen Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Reviews of Modern physics*, 48(4):571, 1976.

[Sei12]   Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75(12):126001, 2012.

[SFHD99]   Ch Schütte, Alexander Fischer, Wilhelm Huisinga, and Peter Deuflhard. A direct approach to conformational dynamics based on hybrid monte carlo. *Journal of Computational Physics*, 151(1):146–168, 1999.

[SS97]   MC Sanguinetti and PS Spector. Potassium channelopathies. *Neuropharmacology*, 36(6):755–762, 1997.

[STSP+15] Martin K Scherer, Benjamin Trendelkamp-Schroer, Fabian Paul, Guillermo Perez-Hernandez, Moritz Hoffmann, Nuria Plattner, Christoph Wehmeyer, Jan-Hendrik Prinz, and Frank Noe. Pyemma 2: A software package for estimation, validation, and analysis of markov models. *Journal of chemical theory and computation*, 11(11):5525–5542, 2015.

Bibliography

[SYA+06] Ning Shi, Sheng Ye, Amer Alam, Liping Chen, and Youxing Jiang. Atomic structure of a na+-and k+-conducting channel. *Nature*, 440(7083):570, 2006.

[SZC+13] David B Sauer, Weizhong Zeng, John Canty, Yeeling Lam, and Youxing Jiang. Sodium and potassium competition in potassium-selective and non-selective channels. *Nature communications*, 4:2721, 2013.

[SZRJ11] David B Sauer, Weizhong Zeng, Srinivasan Raghunathan, and Youxing Jiang. Protein interactions central to stabilizing the k+ channel selectivity filter in a four-sited configuration for selective k+ permeation. *Proceedings of the National Academy of Sciences*, 2011.

[VDSLH+05] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. Gromacs: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–1718, 2005.

[VLH+05] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E. Mark, and Herman J C Berendsen. GROMACS: Fast, flexible, and free, 2005.

[VNU51] John Von Neumann and Stanislav Ulam. Monte carlo method. *National Bureau of Standards Applied Mathematics Series*, 12:36, 1951.

[VR+07] Guido Van Rossum et al. Python programming language. In *USENIX Annual Technical Conference*, volume 41, page 36, 2007.

[WBB+10] Beth G Wensley, Sarah Batey, Fleur AC Bone, Zheng Ming Chan, Nuala R Tumelty, Annette Steward, Lee Gyan Kwa, Alessandro Borgia, and Jane Clarke. Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature*, 463(7281):685, 2010.

72

[WCV11] Stéfan van der Walt, S Chris Colbert, and Gael Varoquaux. The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30, 2011.

[WWC+04] Junmei Wang, Romain M Wolf, James W Caldwell, Peter A Kollman, and David A Case. Development and testing of a general amber force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.

[Yel98] Gary Yellen. The moving parts of voltage-gated ion channels. *Quarterly reviews of biophysics*, 31(3):239–295, 1998.

[YLJ10] Sheng Ye, Yang Li, and Youxing Jiang. Novel insights into k+ selectivity from high-resolution structures of an open k+ channel pore. *Nature structural & molecular biology*, 17(8):1019, 2010.

[ZMCKM01] Yufeng Zhou, Joao H Morais-Cabral, Amelia Kaufman, and Roderick MacKinnon. Chemistry of ion coordination and hydration revealed by a k+ channel–fab complex at 2.0 r a resolution. *Nature*, 414(6859):43, 2001.

[Zwa61] Robert Zwanzig. Lectures in theoretical physics, 1961.

Bibliography

74

A. Appendix

77
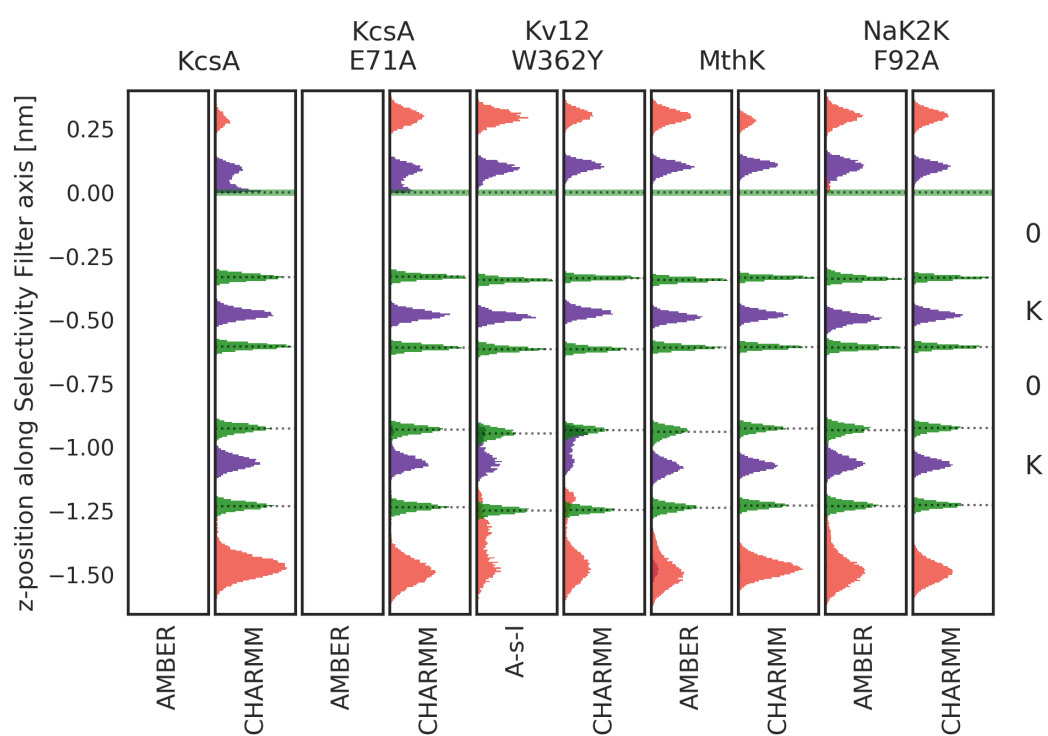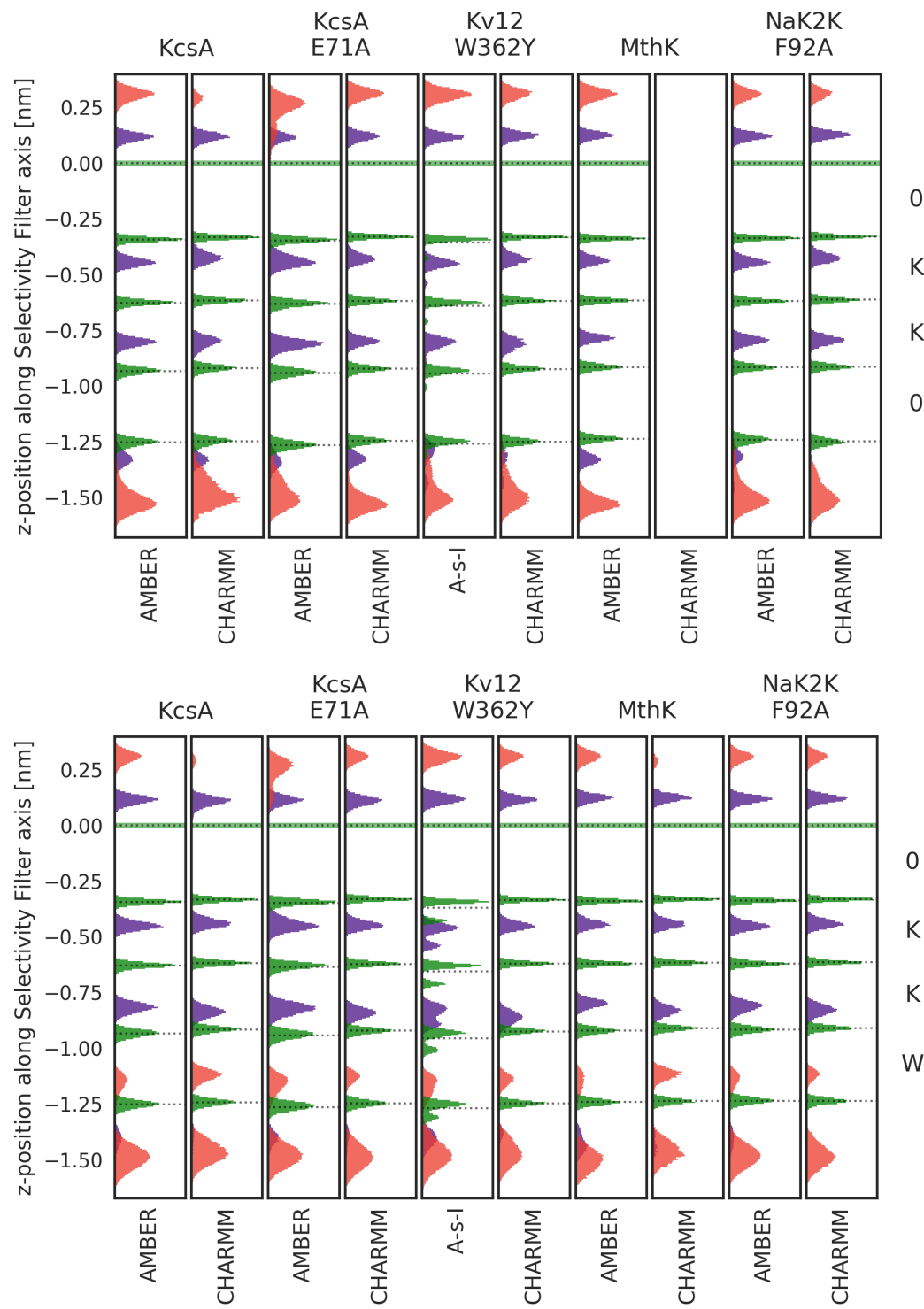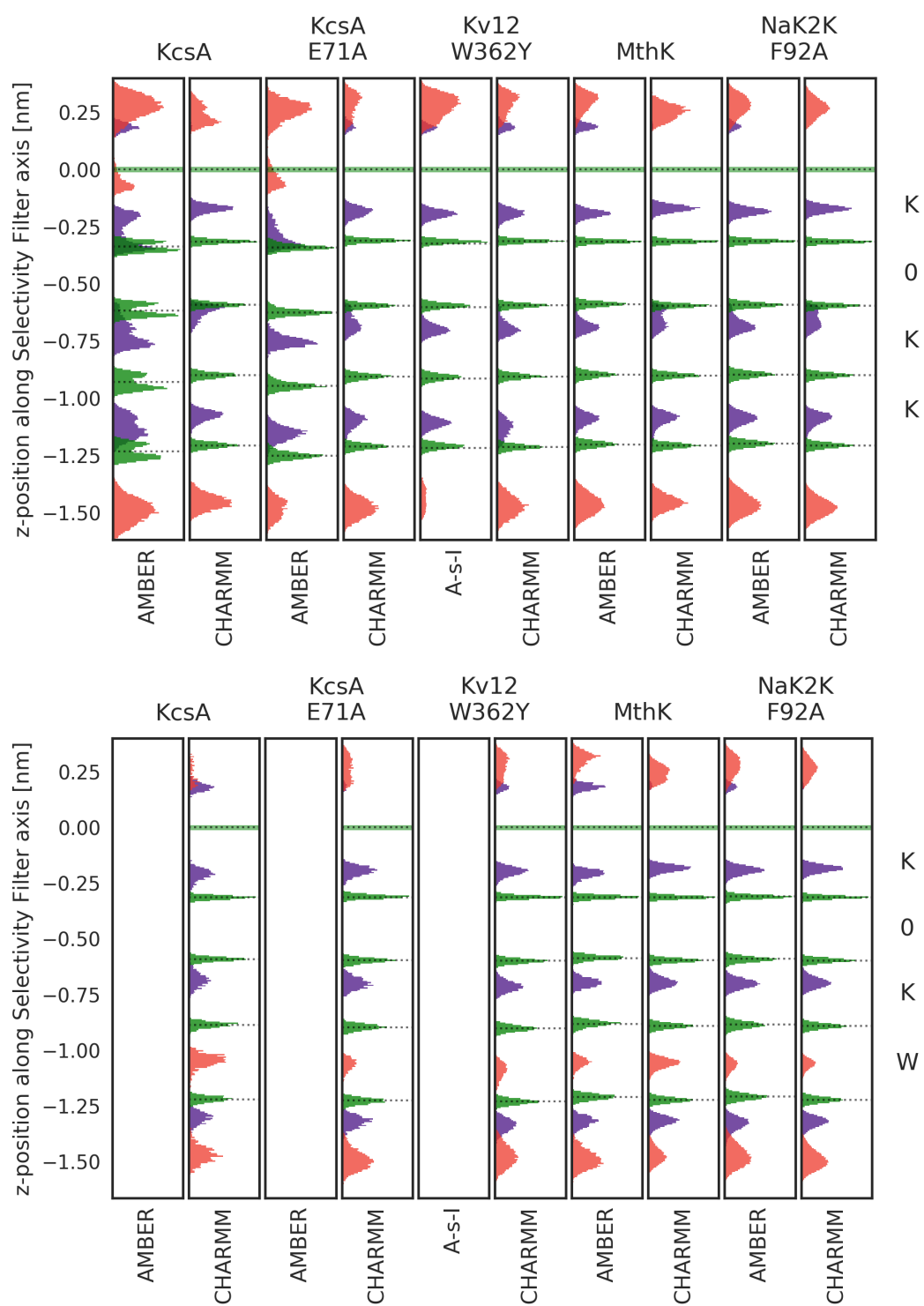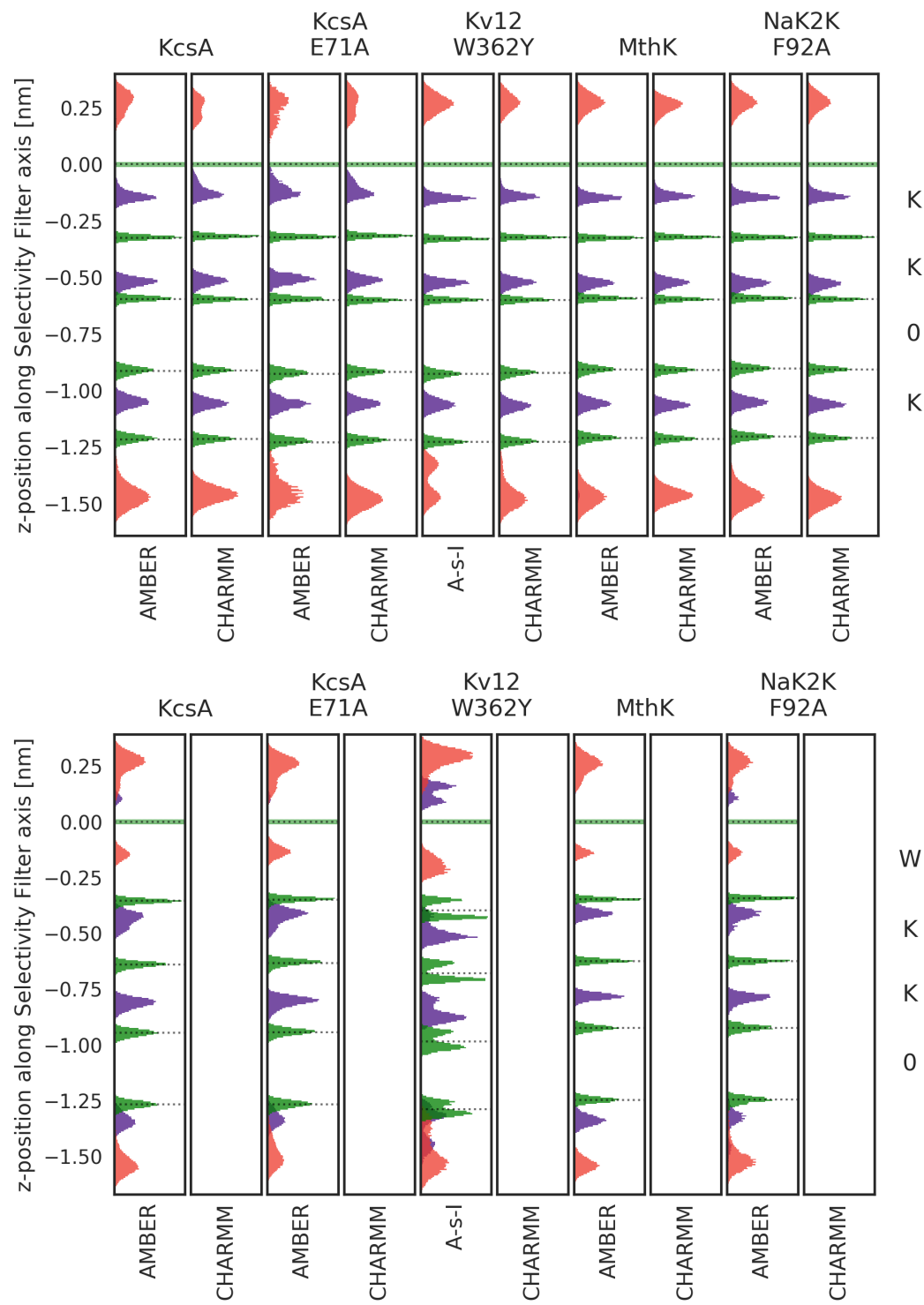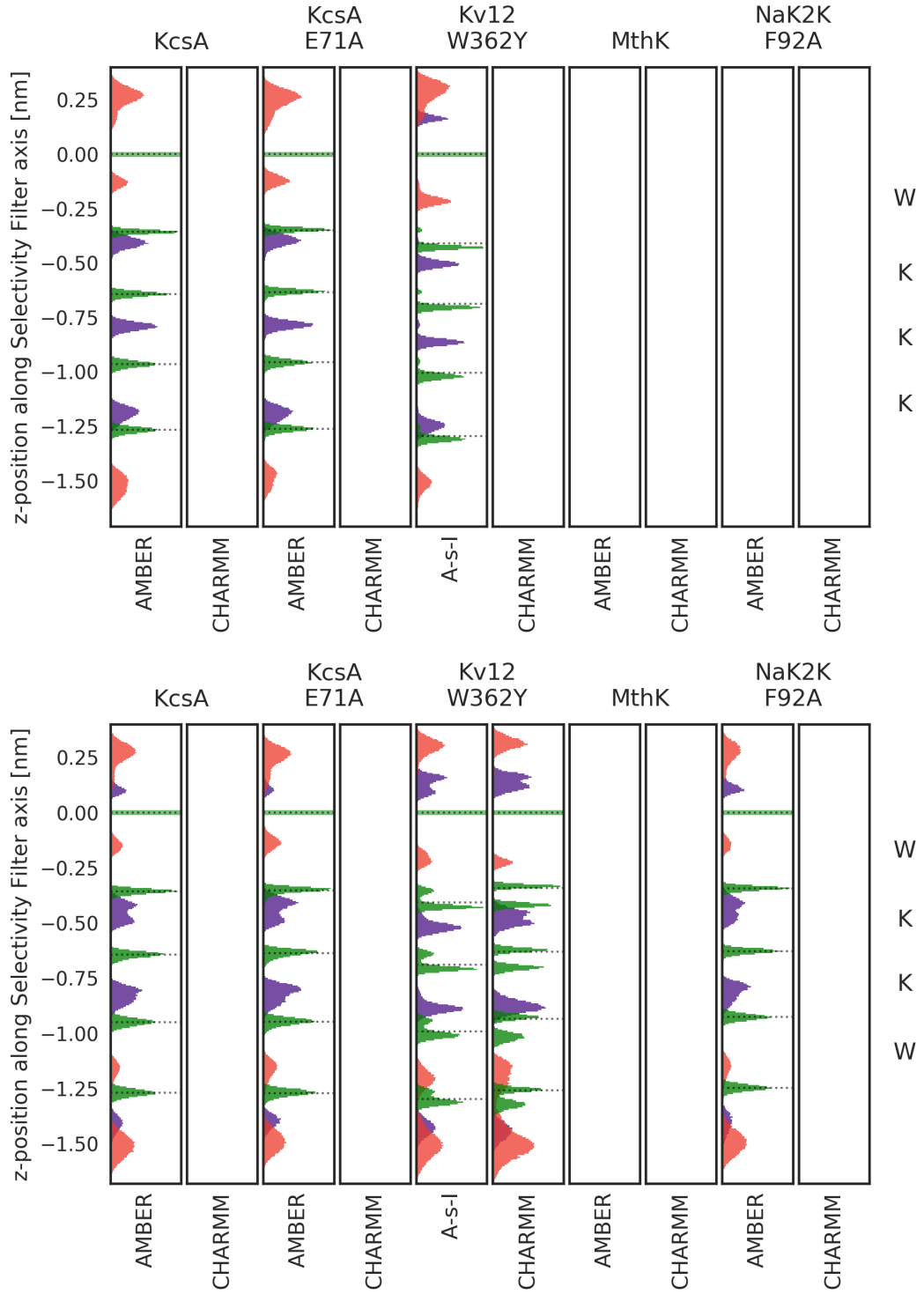
Figure A.1.: Histograms of the z-positions of oxygen atoms of the SF (green peaks), the K^+ ions (violet peaks) and the water molecules (red peaks). This is the breakdown of figure 4.1 into the individual states, confirming that the states themselves do not differ between the channels and force fields.
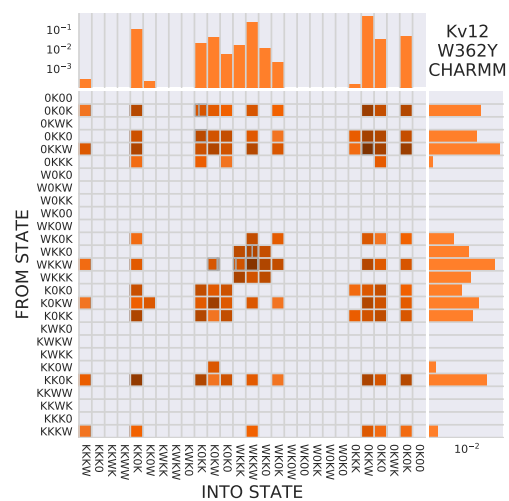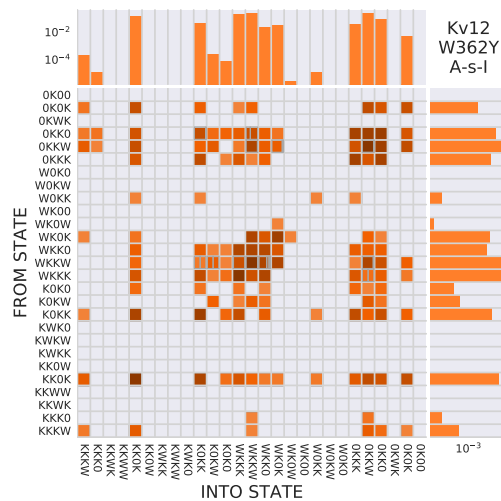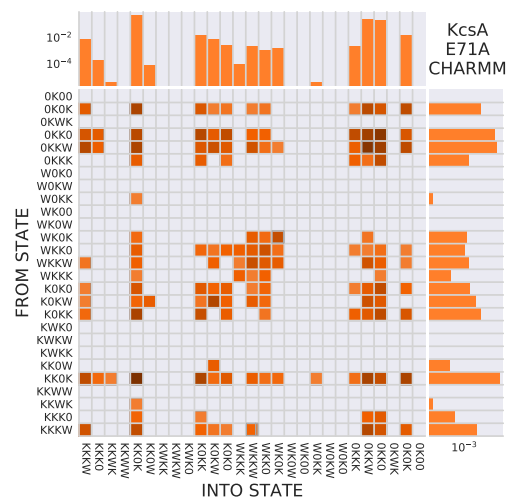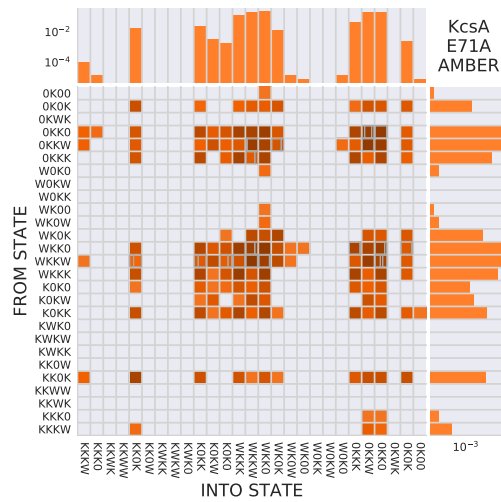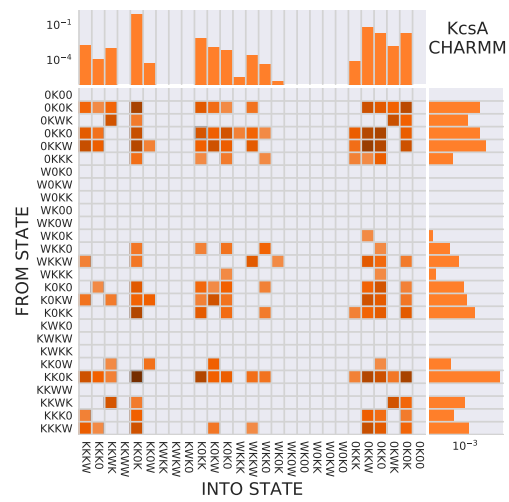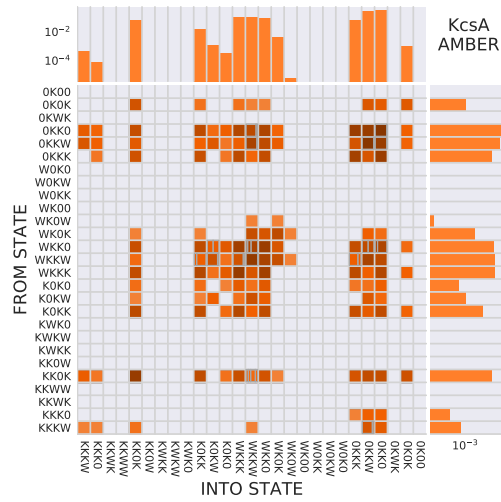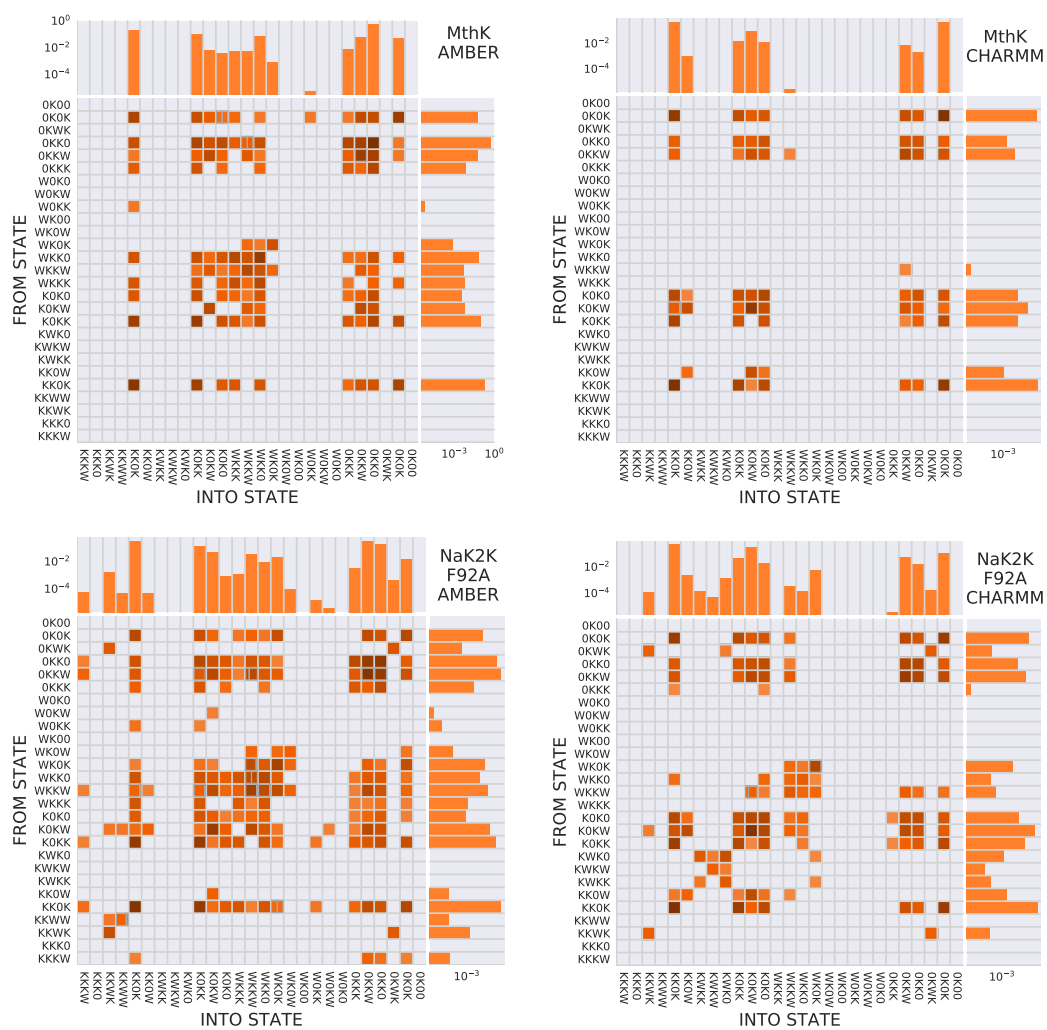
Figure A.2.: All transition matrices, scale bar and ordering as in figure 4.3 to simplify comparison.
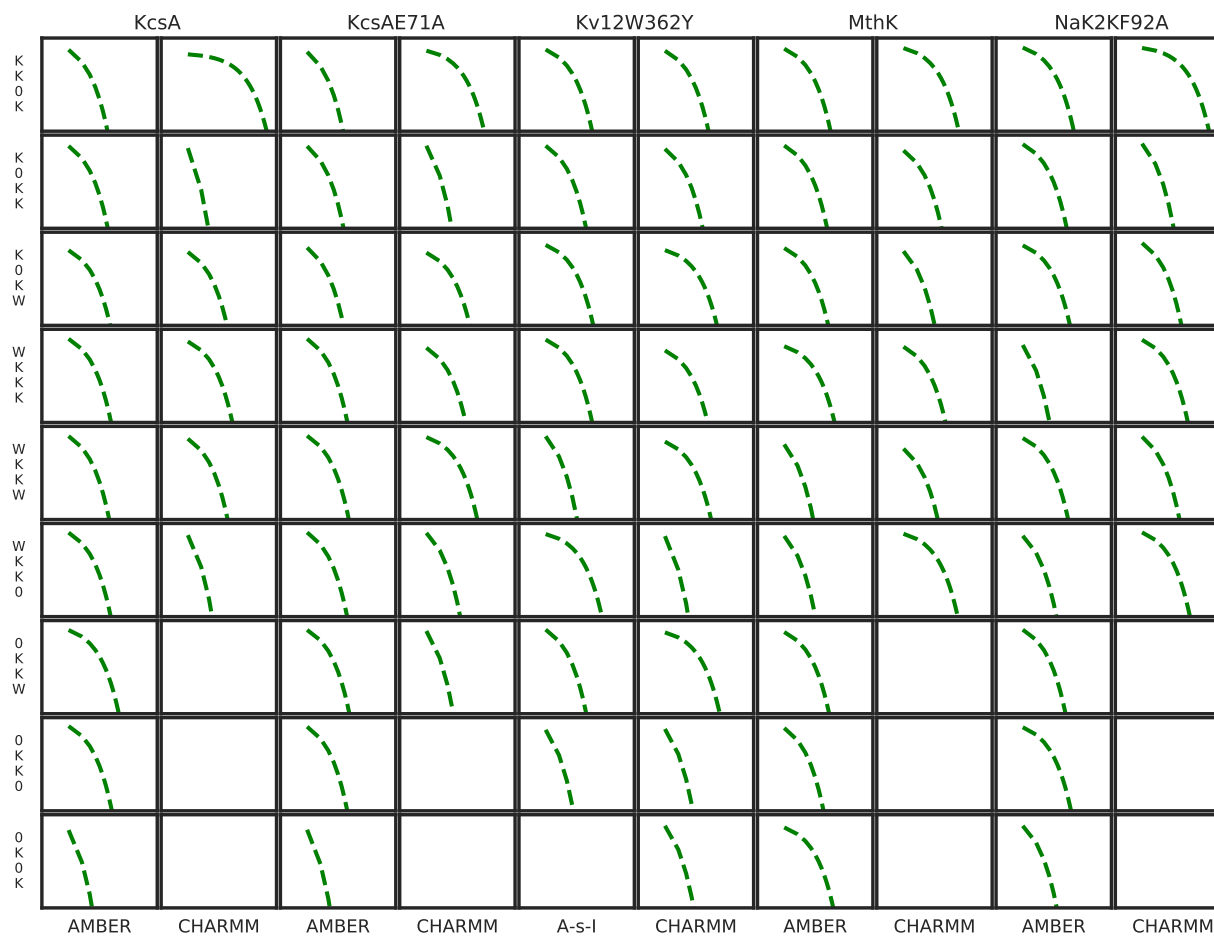
Figure A.3.: Dwell time distributions for the most prominent common states for all the channels.

KcsA AMBER
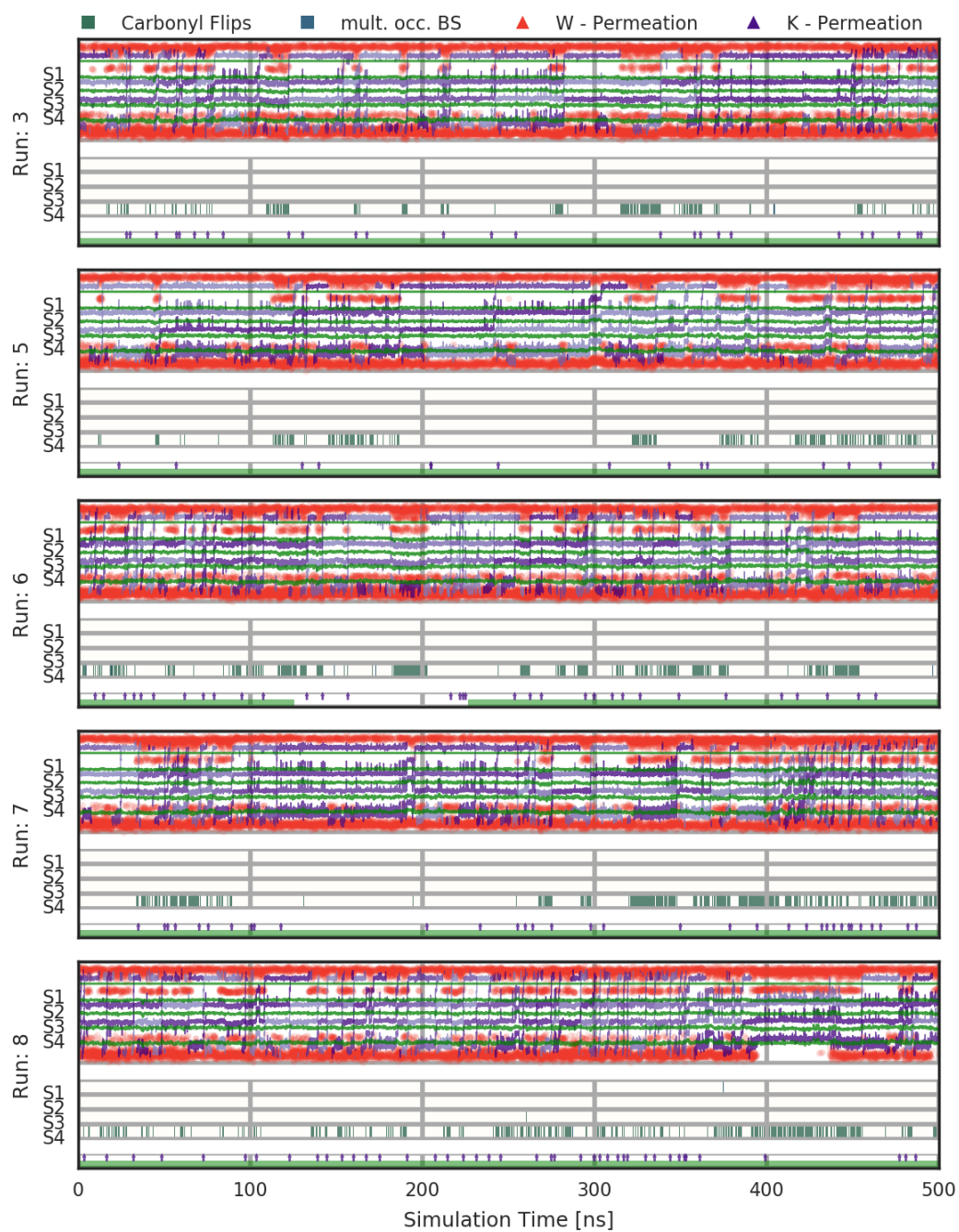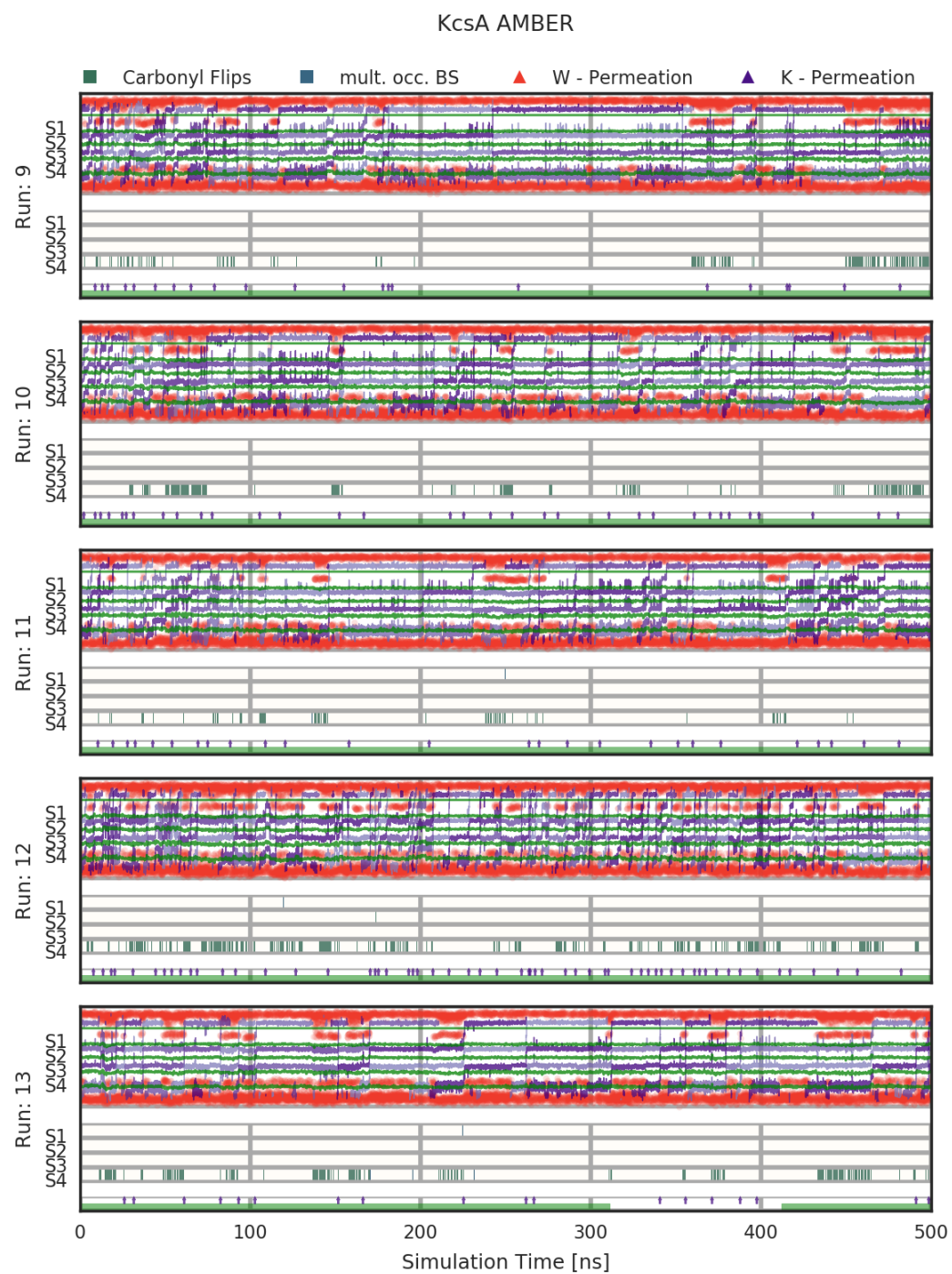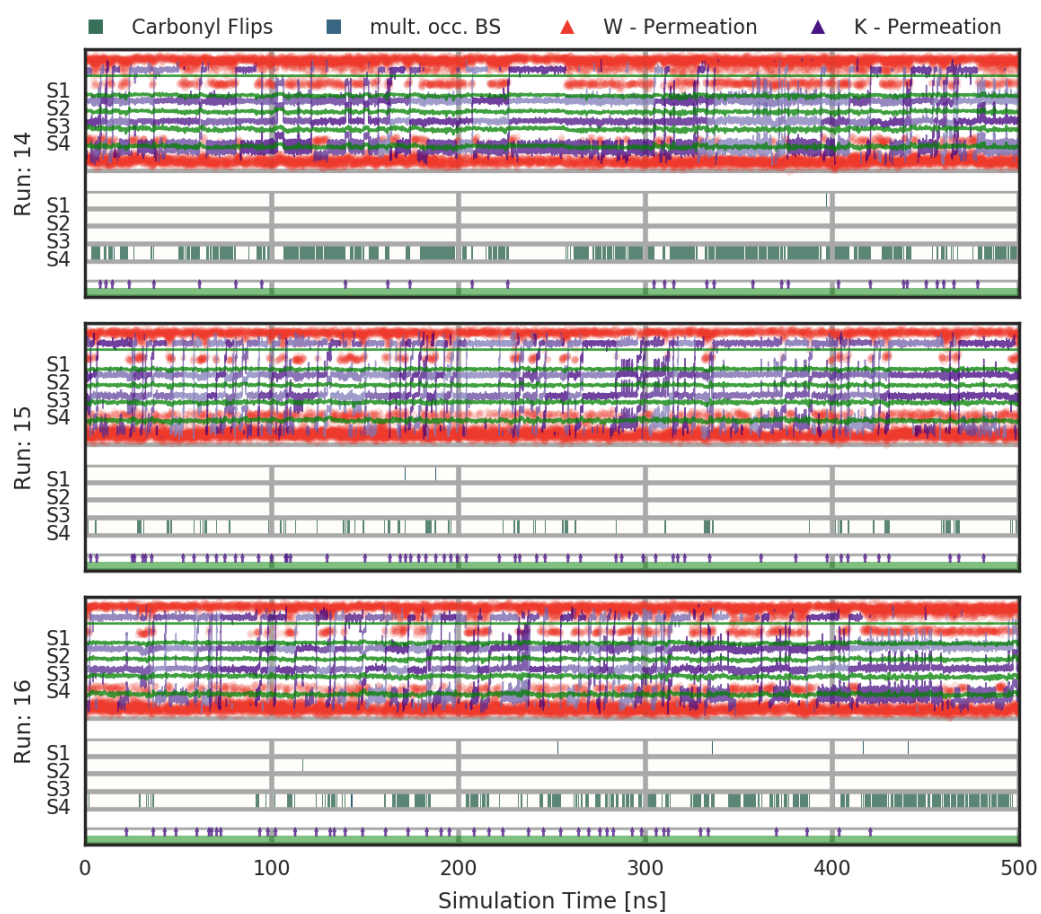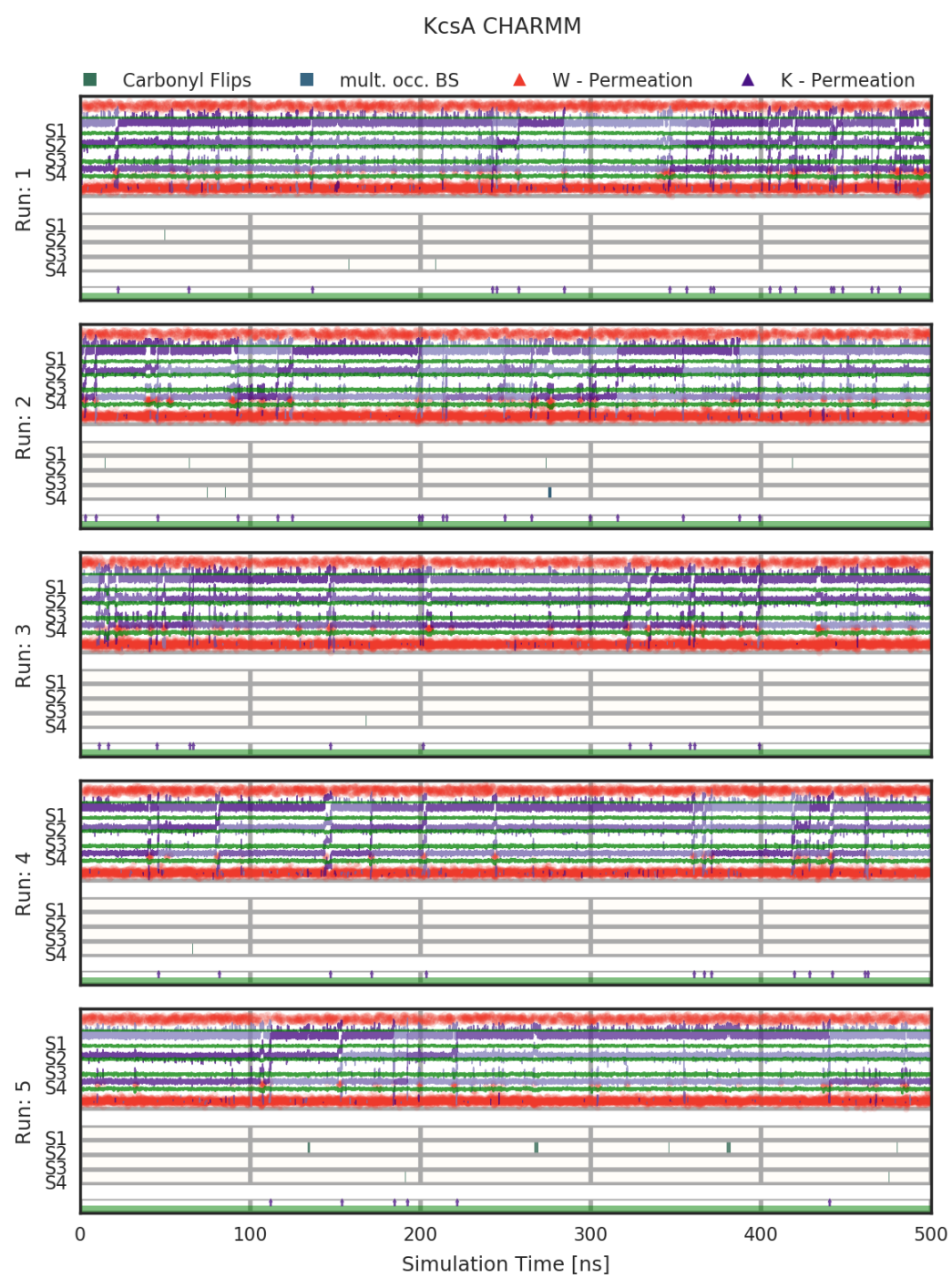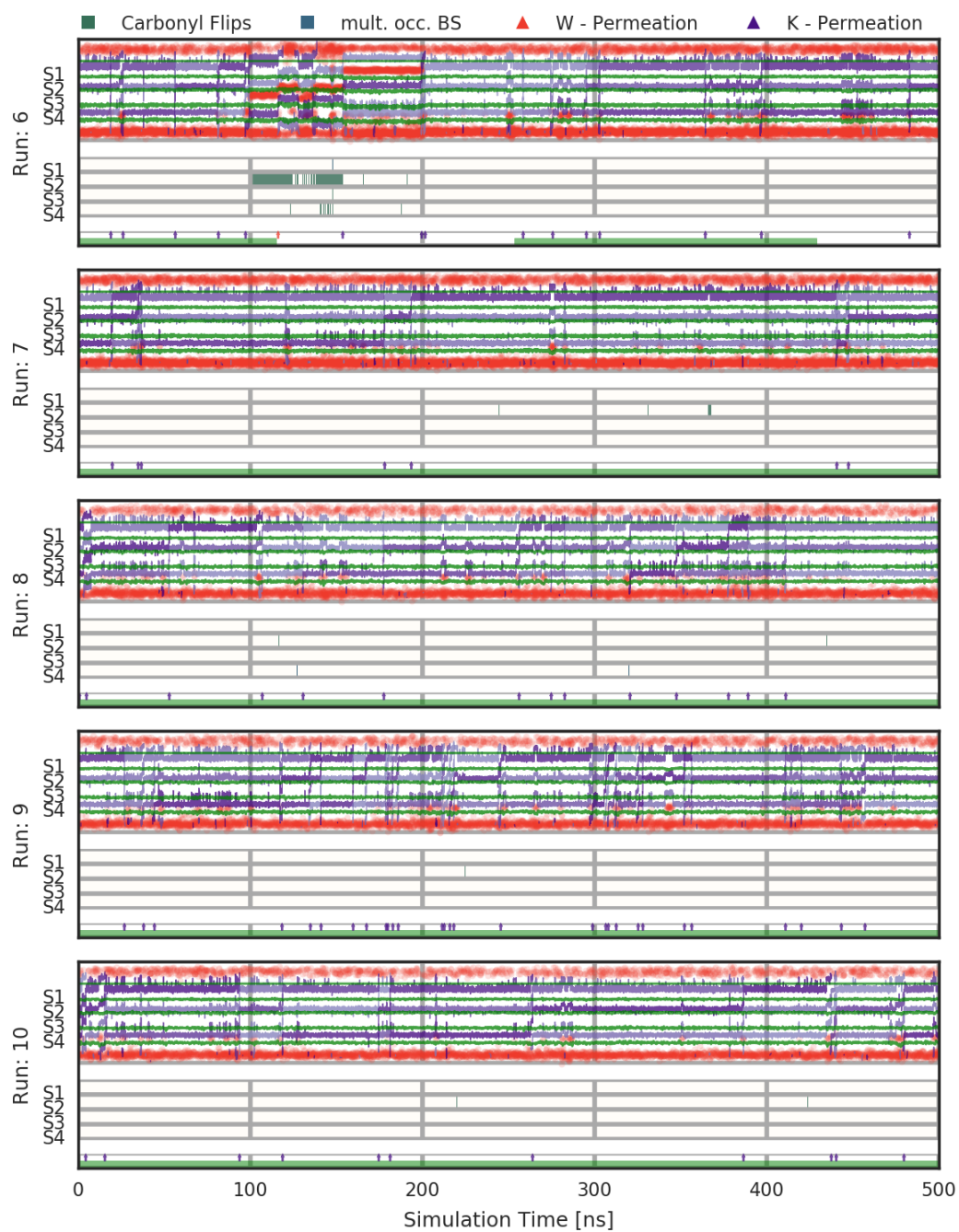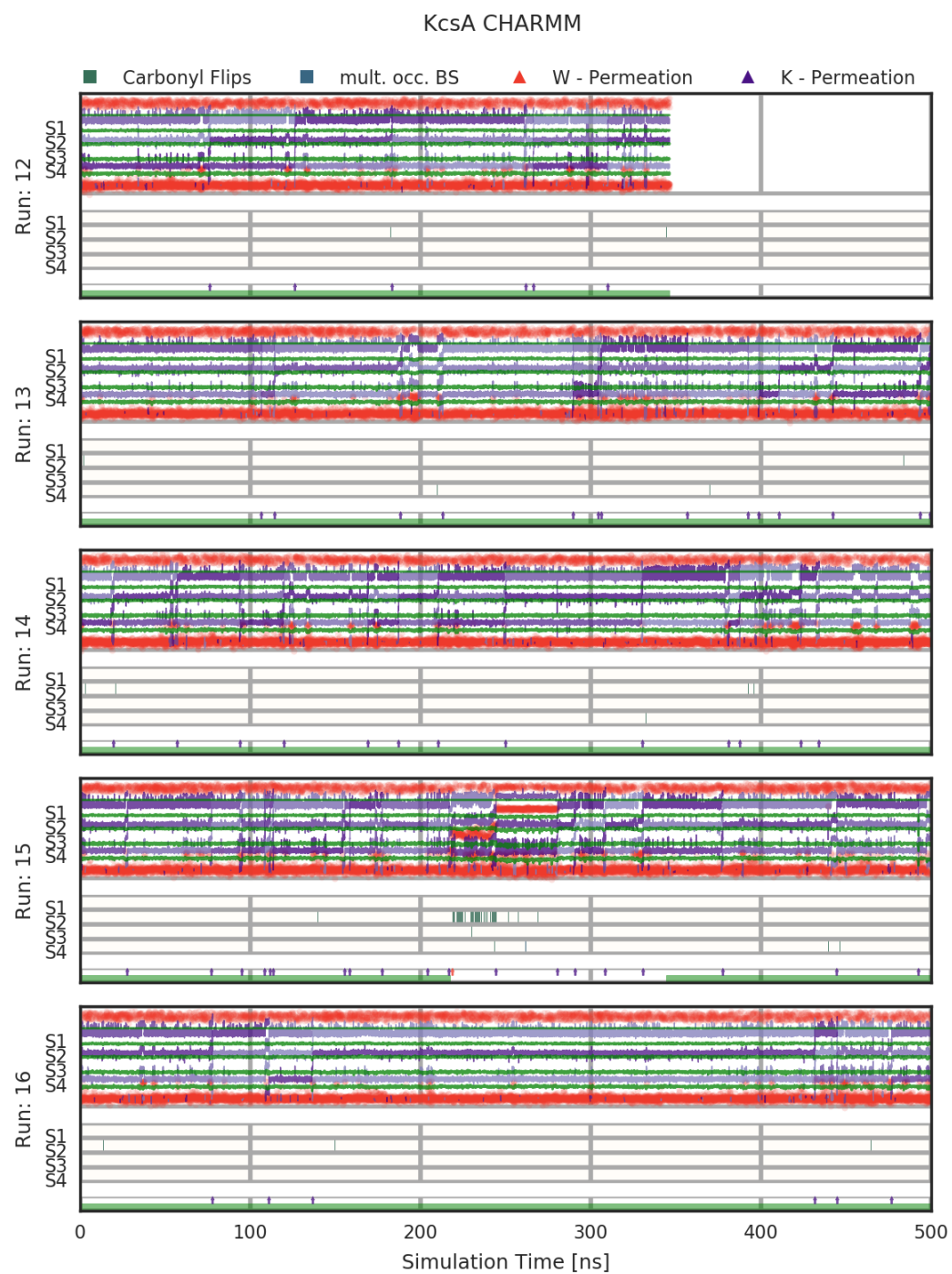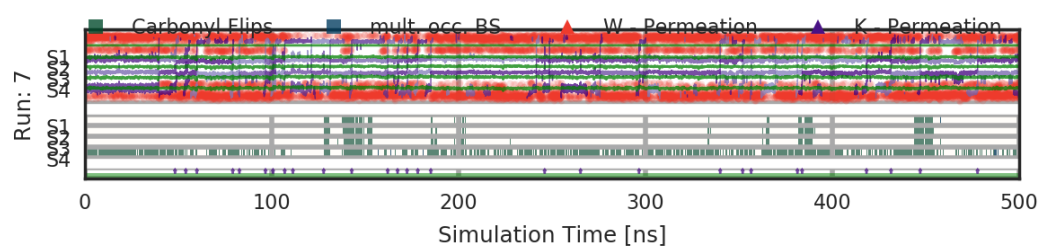
KcsA AMBER

KcsA AMBER

KcsA CHARMM

KcsA CHARMM

KcsA CHARMM

KcsA E71A AMBER

KcsA E71A AMBER

KcsA E71A CHARMM

■ Carbonyl Flips   ■ mult. occ. BS   ▲ W - Permeation   ▲ K - Permeation

KcsA E71A CHARMM

KcsA E71A CHARMM

Kv12 W362Y A-s-I

Kv12 W362Y A-s-I

Carbonyl Flips ■ mult. occ. BS ▲ W - Permeation ▲ K - Permeation

Kv12 W362Y A-s-I

Kv12 W362Y CHARMM

Kv12 W362Y CHARMM

Kv12 W362Y CHARMM

MthK AMBER

MthK AMBER

MthK CHARMM

MthK CHARMM

■ Carbonyl Flips ■ mult. occ. BS ▲ W - Permeation ▲ K - Permeation

NaK2K F92A AMBER

NaK2K F92A AMBER

■ Carbonyl Flips   ■ mult. occ. BS   ▲ W - Permeation   ▲ K - Permeation

NaK2K F92A AMBER
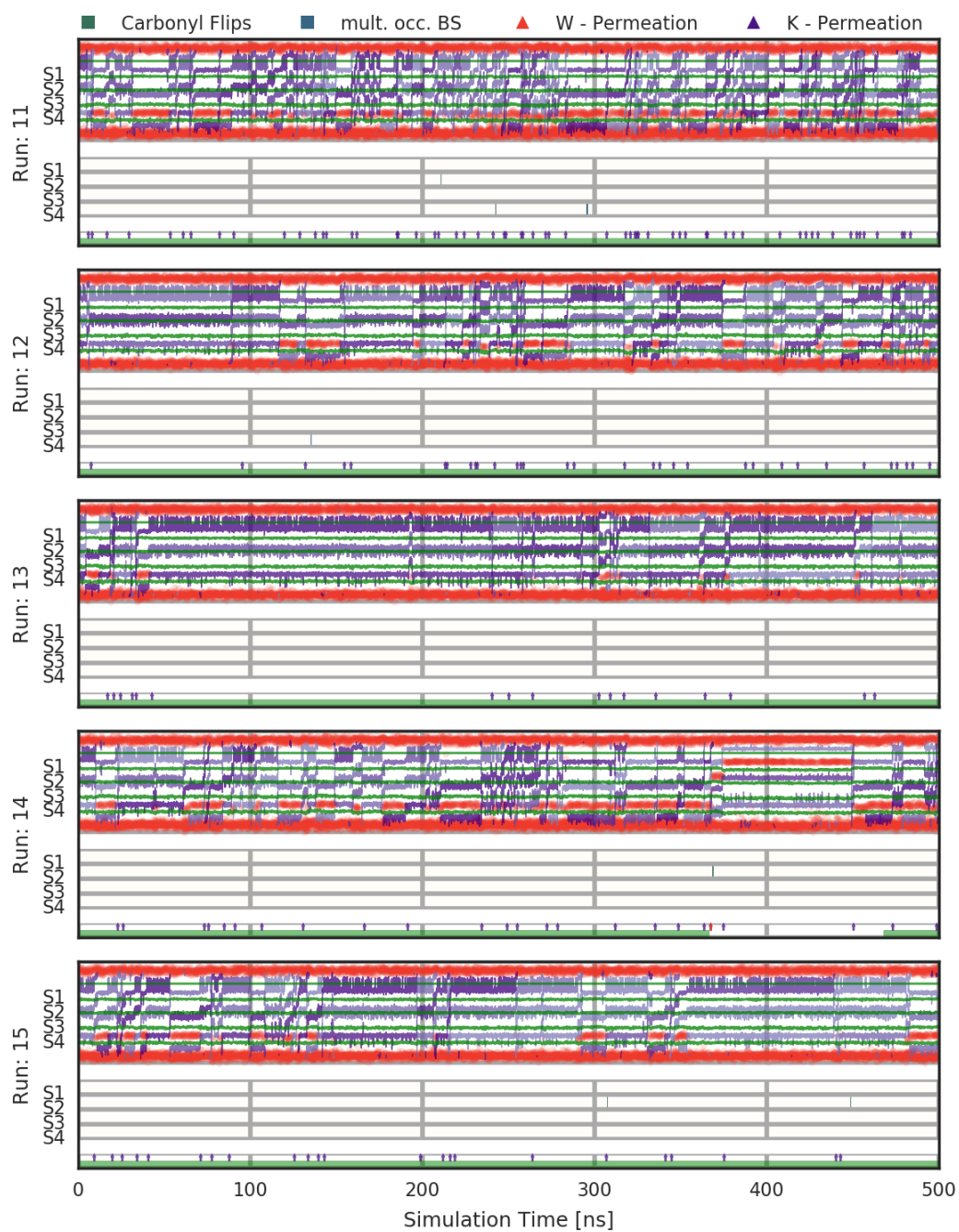

Simulation Time [ns]
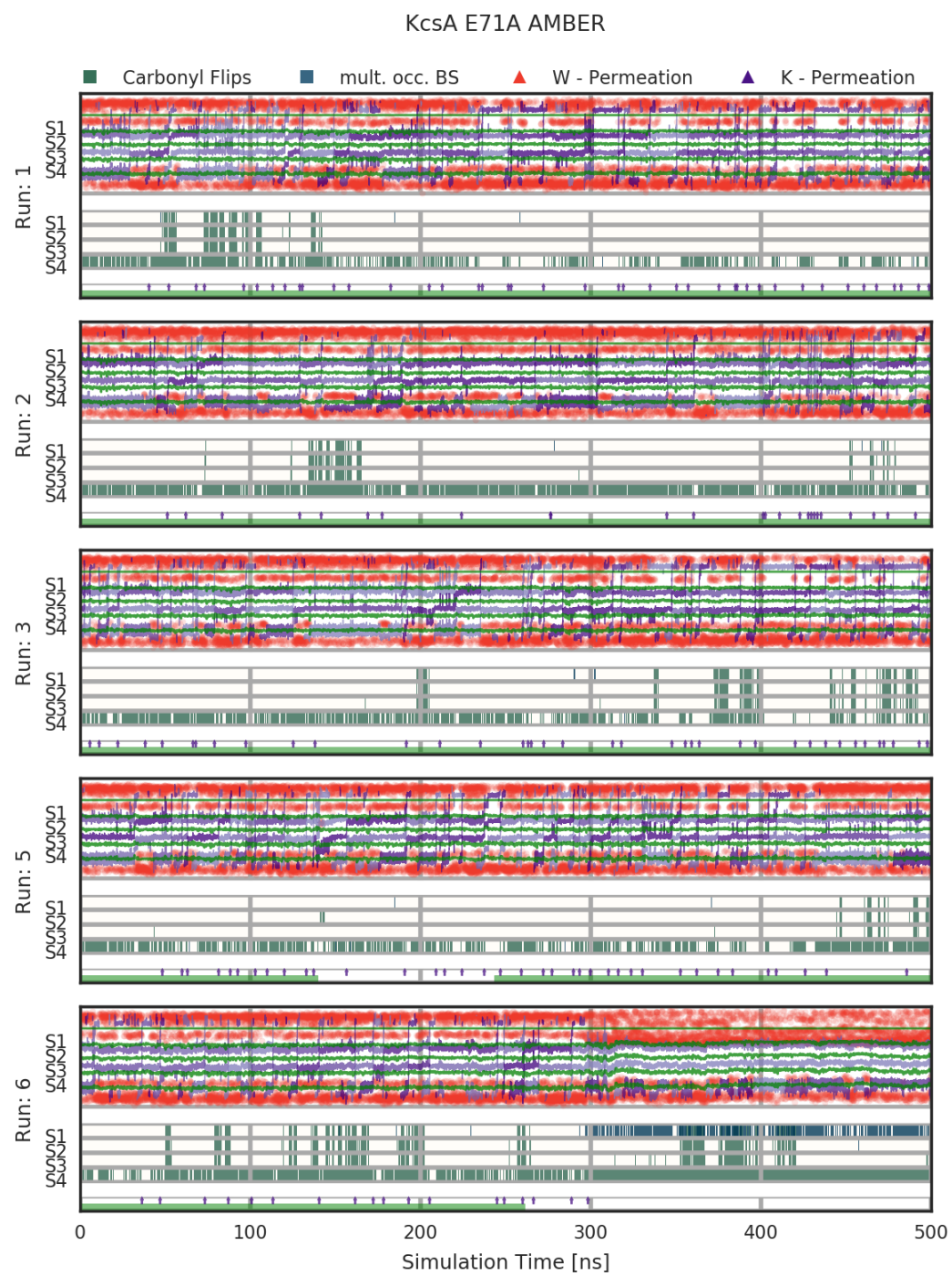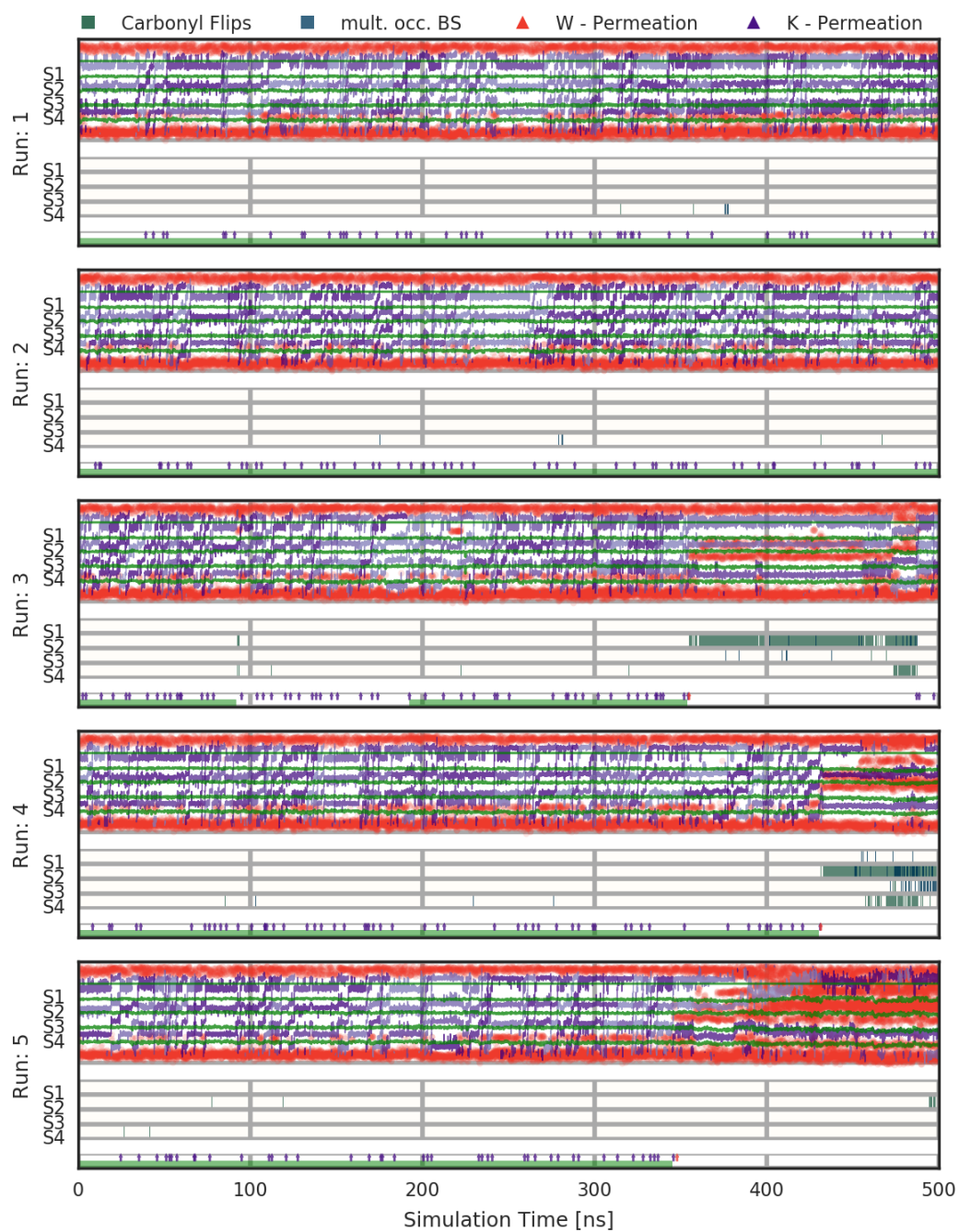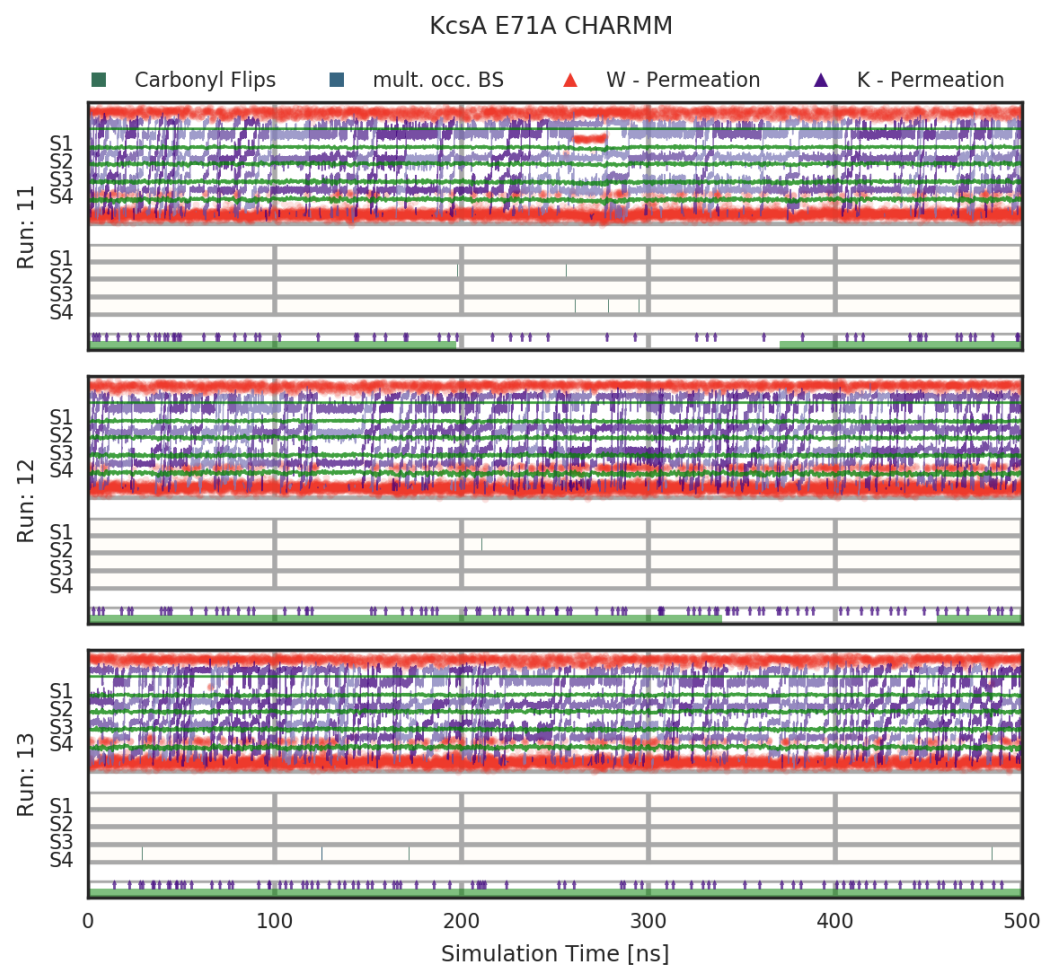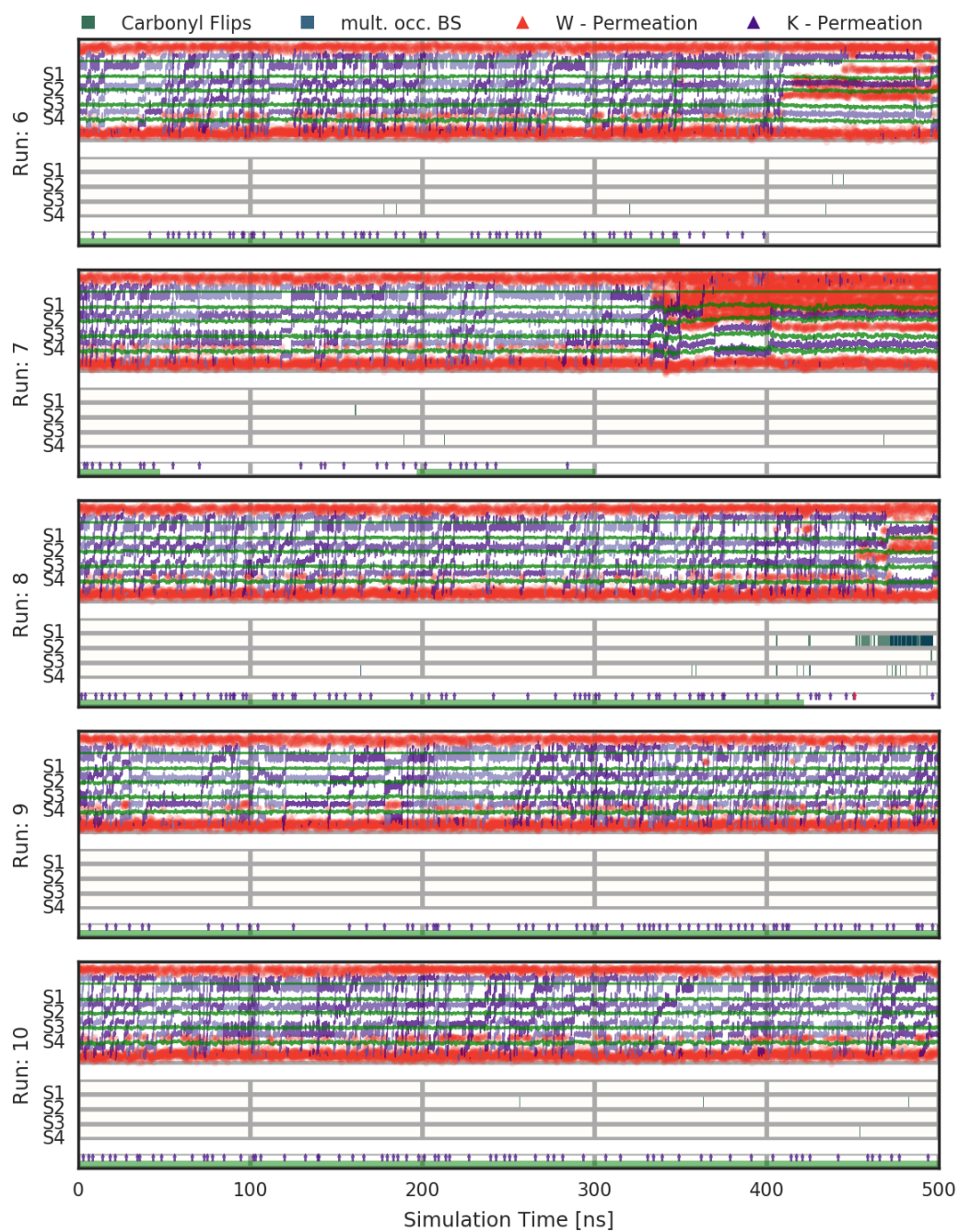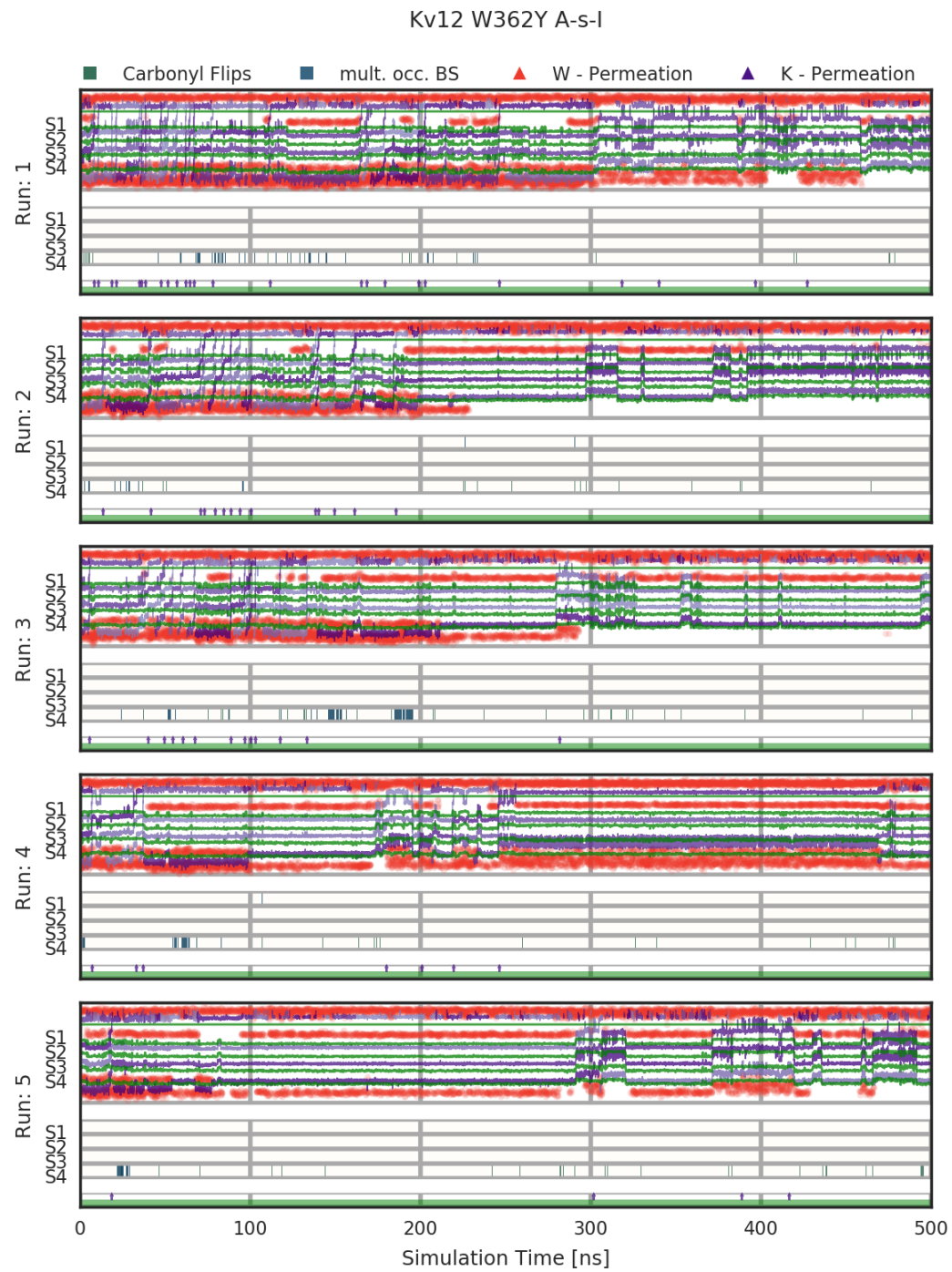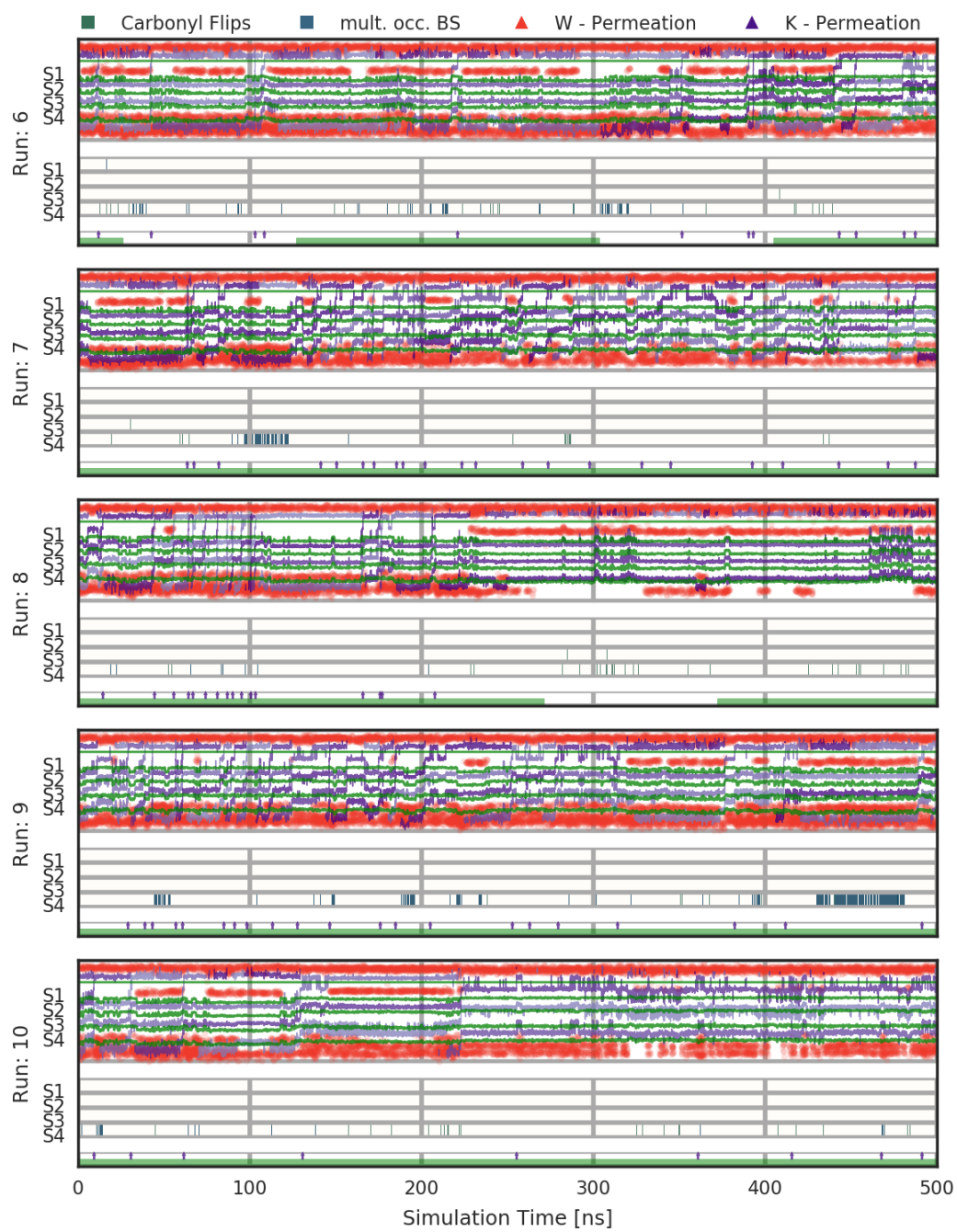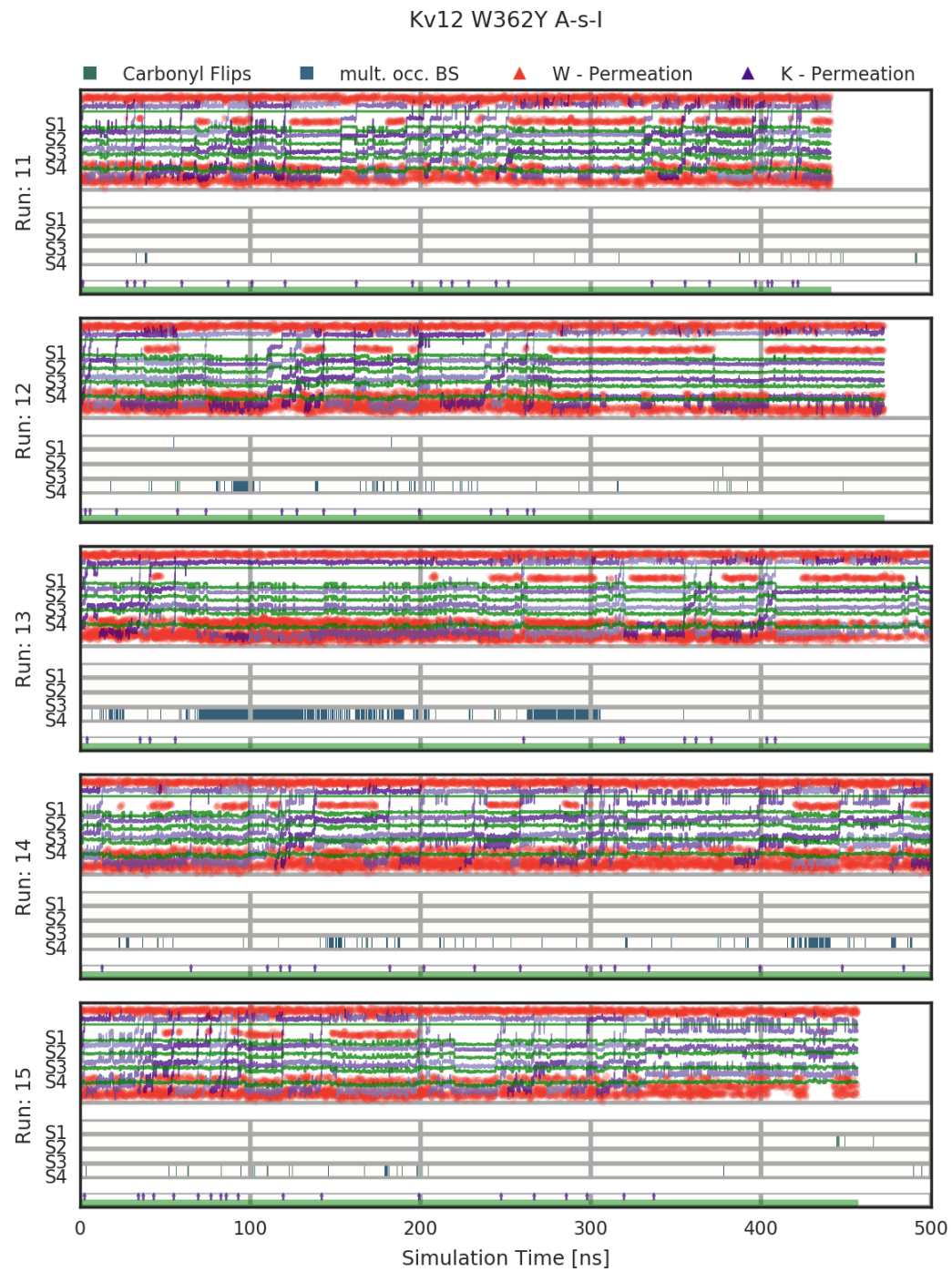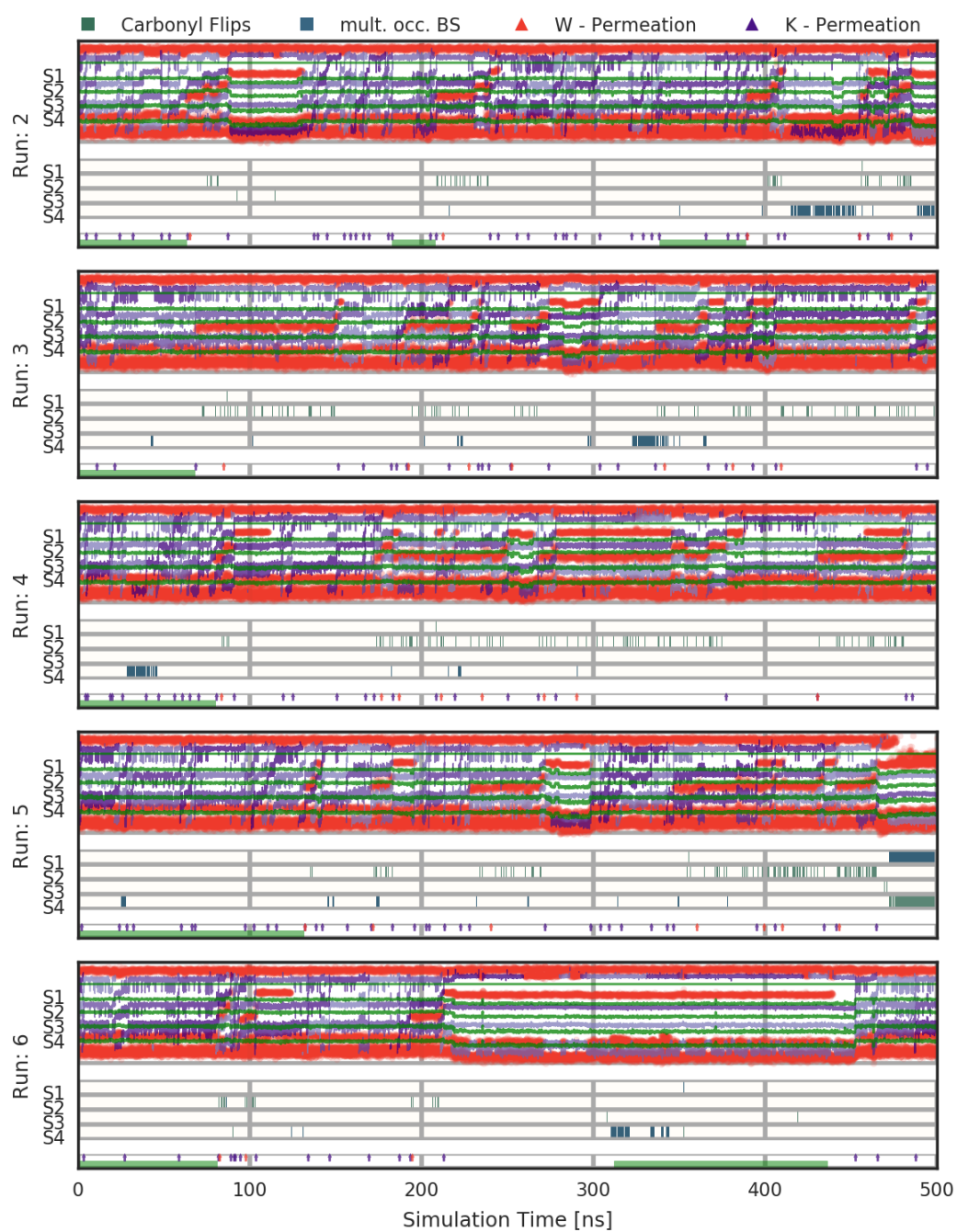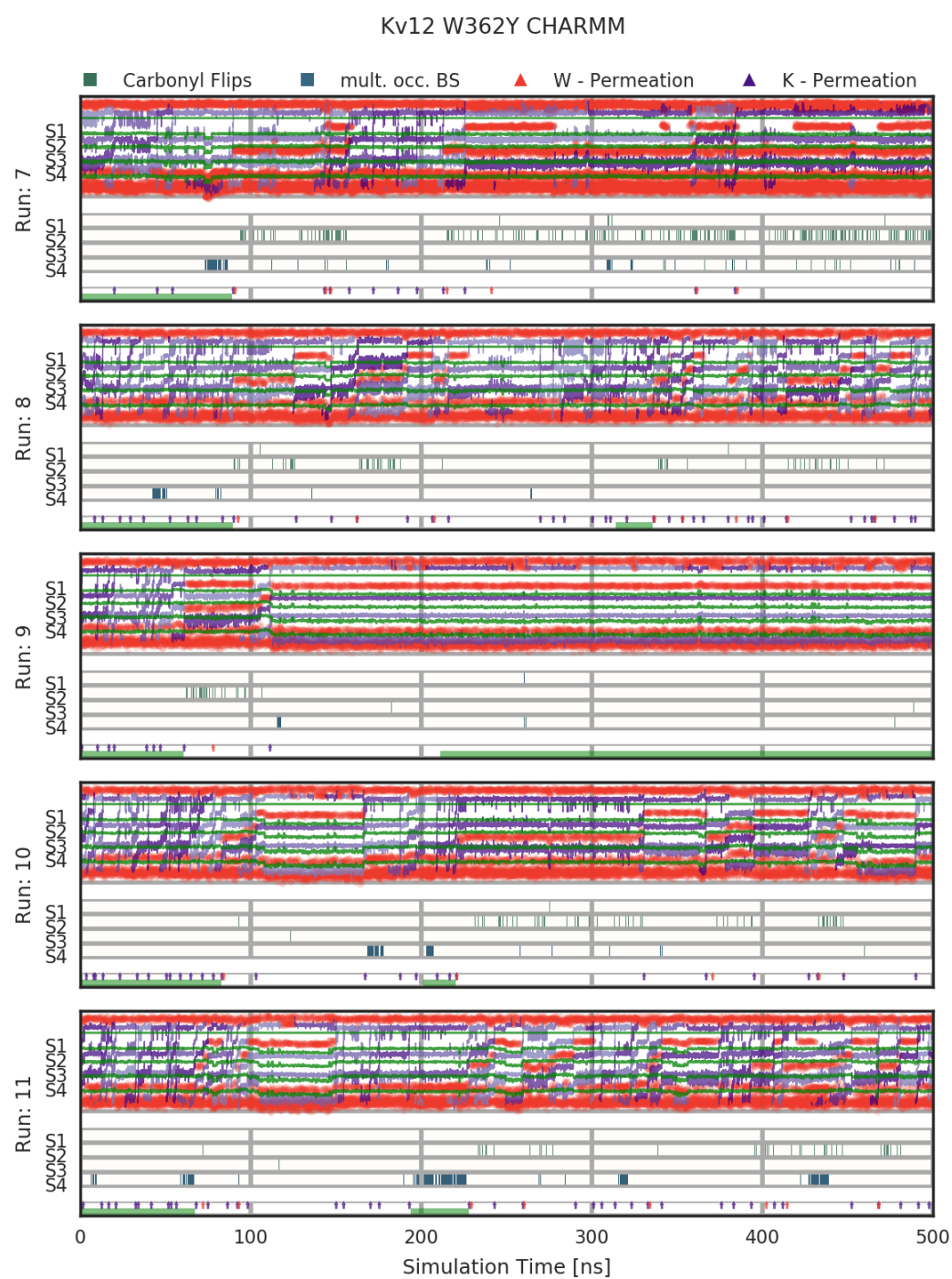
NaK2K F92A CHARMM
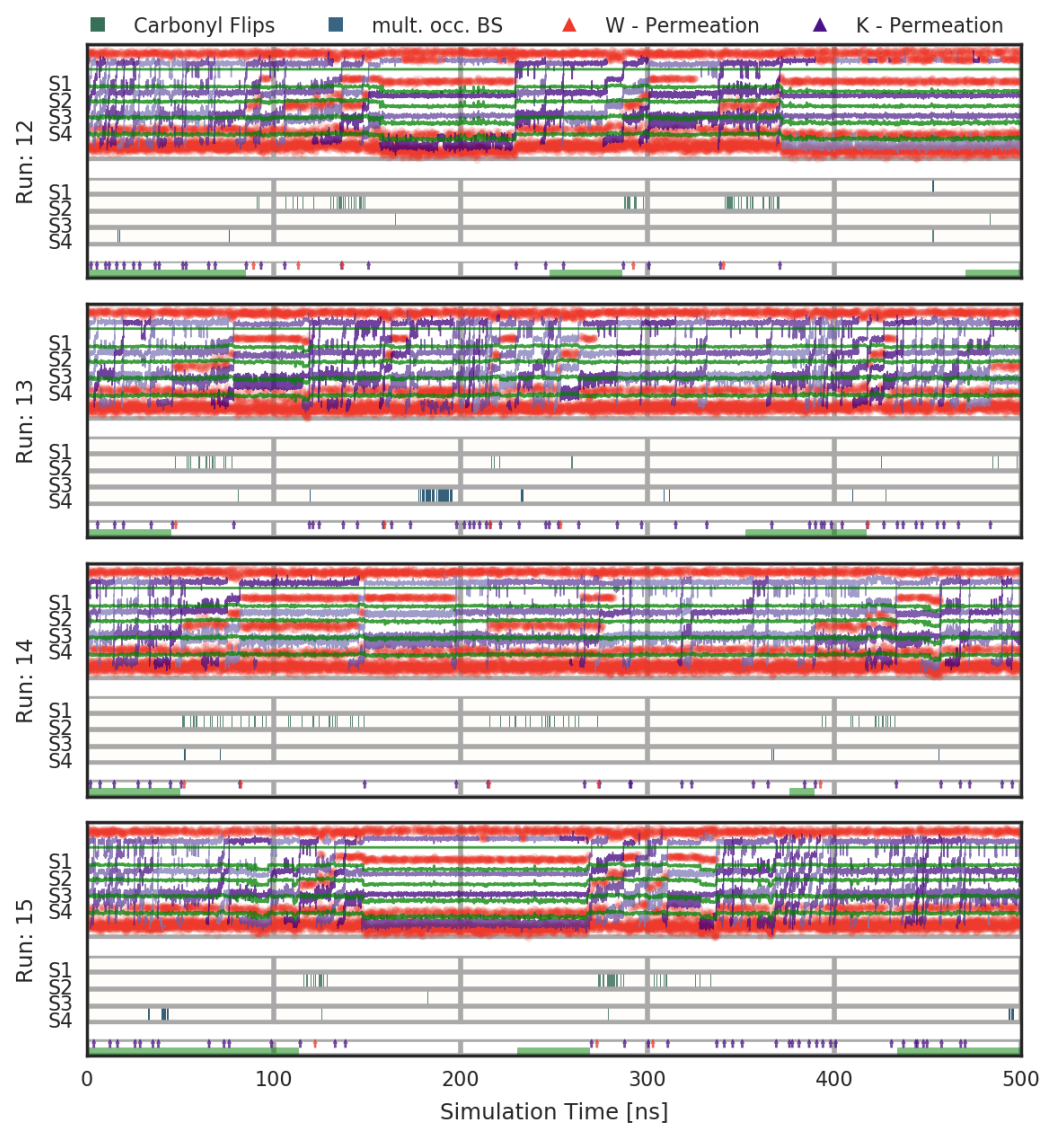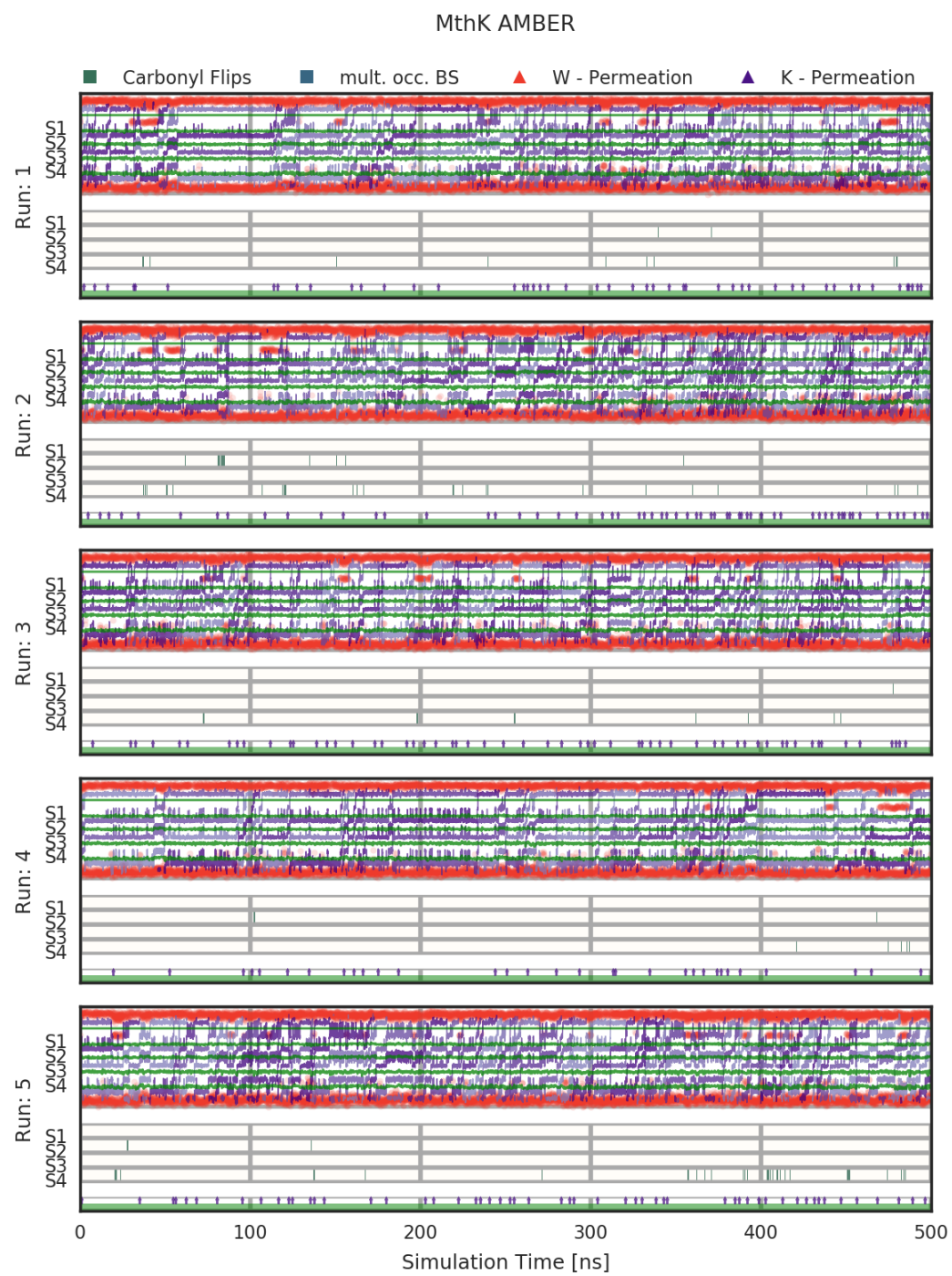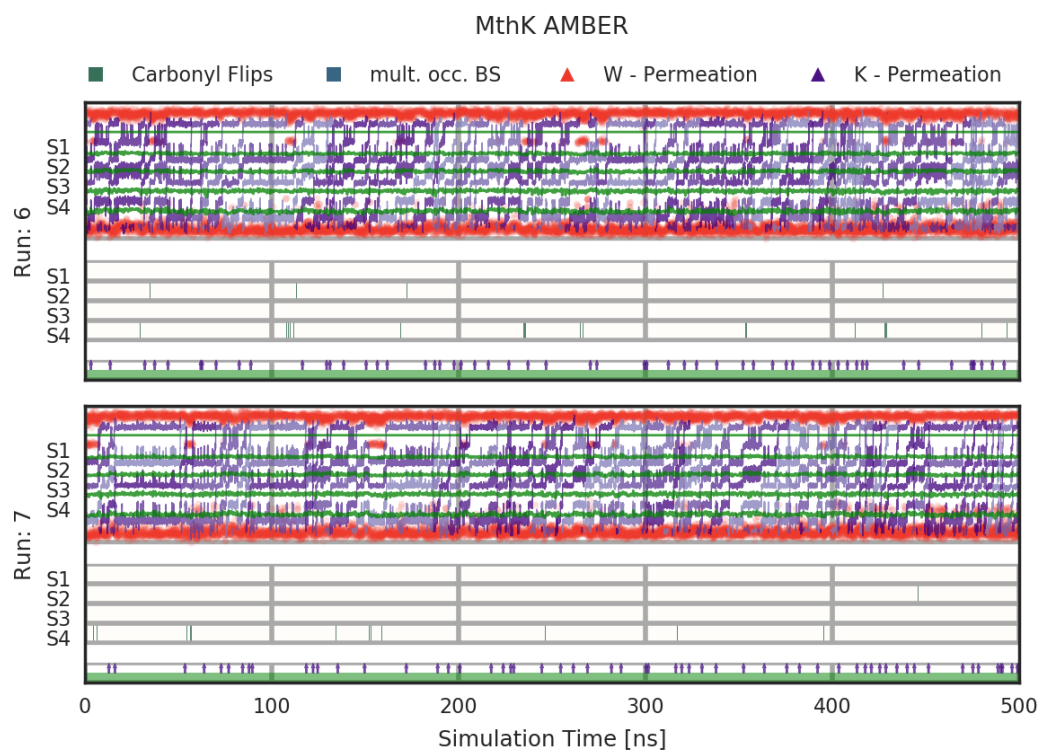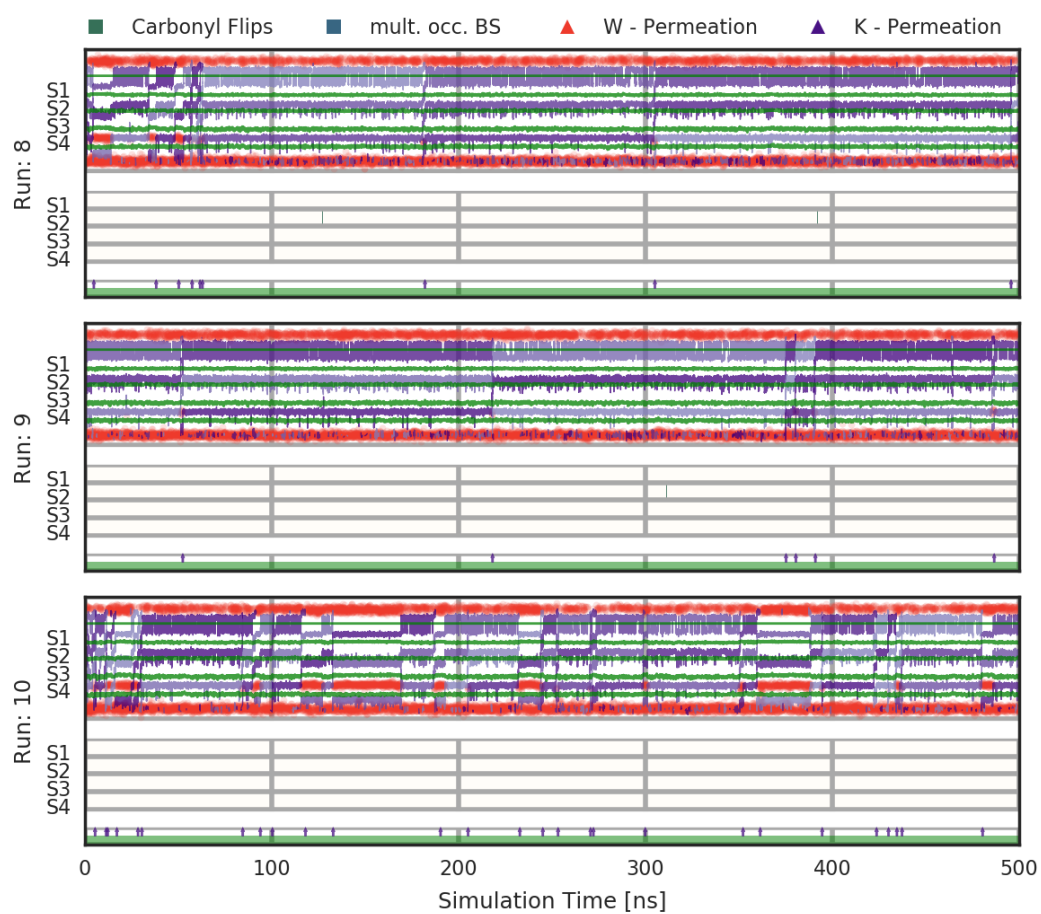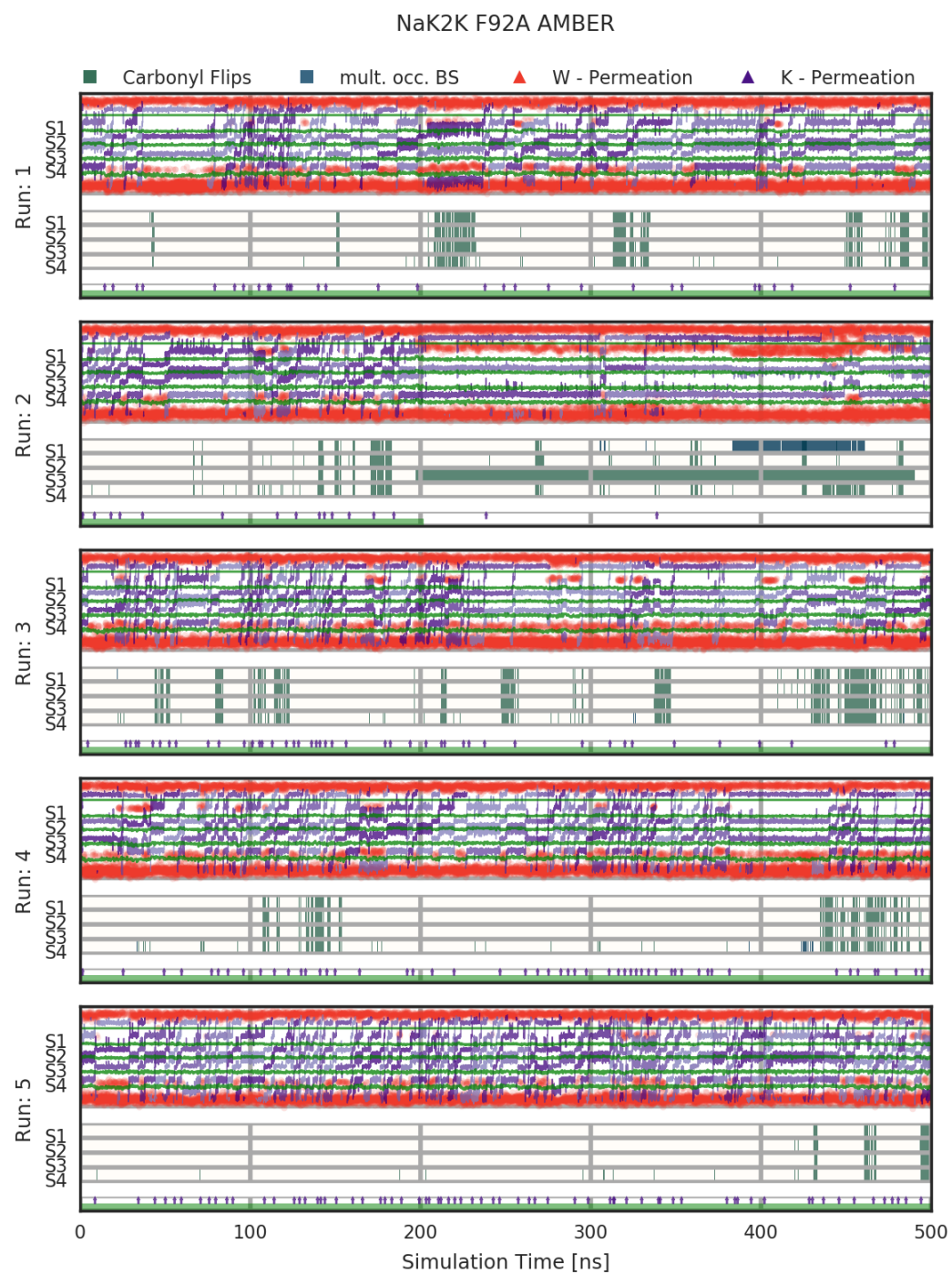
NaK2K F92A CHARMM

NaK2K F92A CHARMM

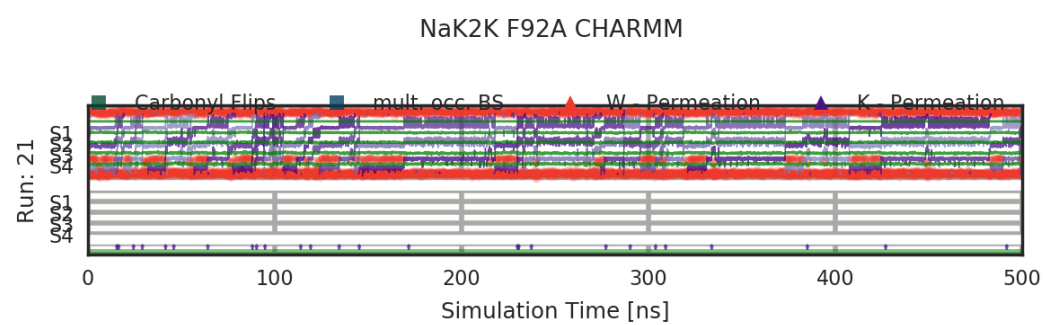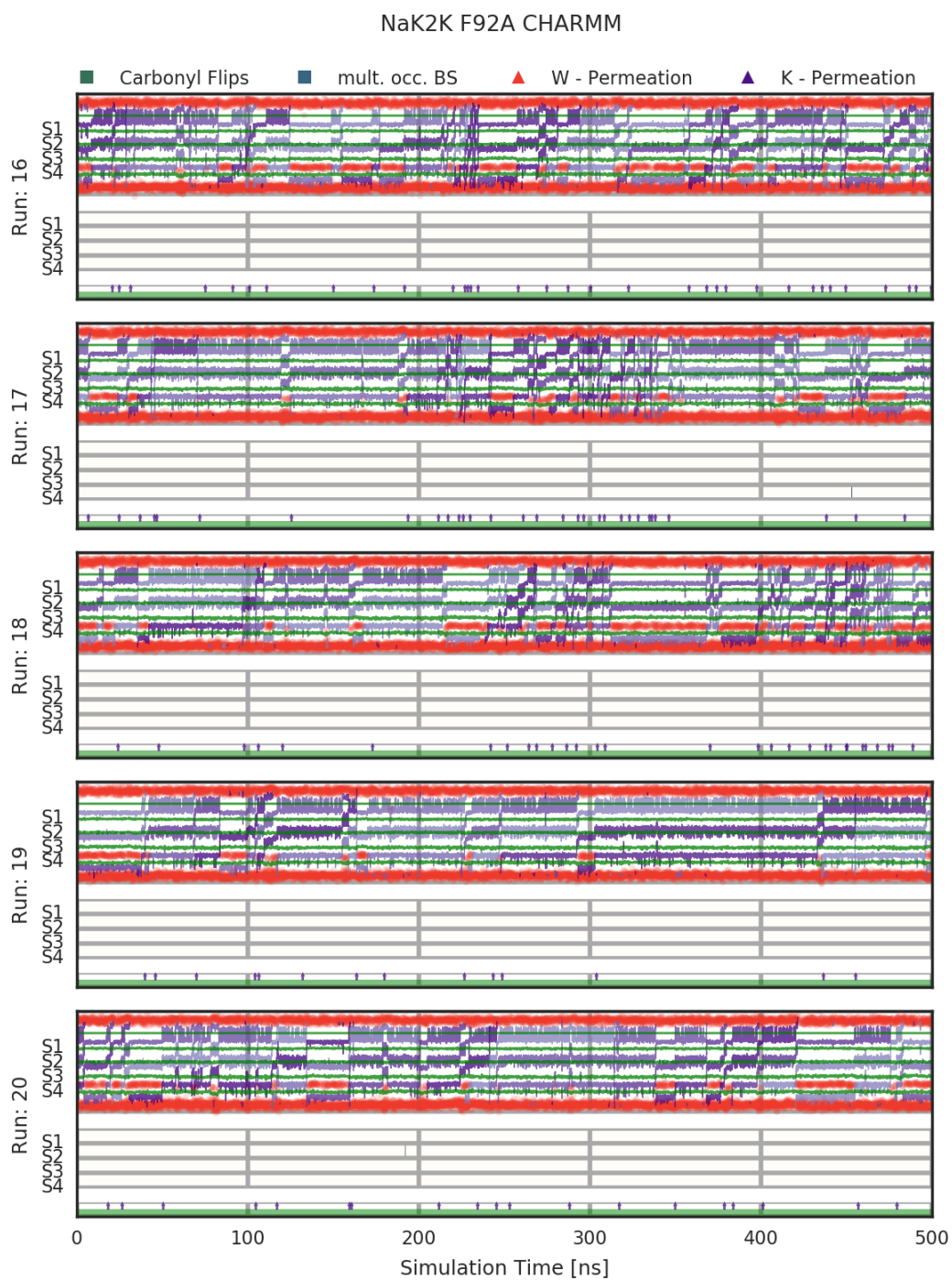## A. Appendix

NaK2K F92A CHARMM

Figure A.4.: These plots give an overview over all trajectories used in this thesis. Depicted are the z – movements of the K$^+$ ions and the water molecules in the selectivity filter in the upper panels of the plots. In the lower panels indications for double occupations and flips of the SF – carbonyls are given. Below each plot a green bar marks which part of the trajectory was used.

A. Appendix

Erklärung Ich versichere hiermit, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die von mir angegebenen Quellen und Hilfsmittel verwendet habe. Wörtlich oder sinngemäß aus anderen Werken entnommene Stellen habe ich unter Angabe der Quellen kenntlich gemacht. Die Richtlinien zur Sicherung der guten wissenschaftlichen Praxis an der Universität Göttingen wurden von mir beachtet. Eine gegebenenfalls eingereichte digitale Version stimmt mit der schriftlichen Fassung überein. Mir ist bewusst, dass bei Verstößen gegen diese Grundsätze die Prüfung mit nicht bestanden bewertet wird.

Göttingen, den August 7, 2018

(Maximilian Alexander Vossel)