New Strategies in Conformation Dynamics

Investigation of the μ-opioid receptor in healthy and inflamed tissue

Eine beim Fachbereich Mathematik und Informatik
der Freien Universität Berlin
eingereichte Dissertation zur Erlangung des
Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

vorgelegt von

Martina Klimm

aus Berlin

Berlin, 2018

1. Gutachter: PD Dr. Marcus Weber,
Mathematischer Molekülentwurf,
Mathematik für Lebens- und Materialwissenschaften,
Zuse-Institut Berlin

2. Gutachter: PD Dr. Konstantin Fackeldey,
Modellierung, Numerik, Differentialgleichungen,
Institut für Mathematik,
Technische Universität Berlin

Tag der Disputation: 08.06.2018

Abstract

The aim of conformation dynamics is the identification of metastable sets in the molecular conformation space, and the computation of their statistical weights. Because of the high dimensionality of the state space, coarse graining methods attempt to improve the computational performance by reducing the representation of a molecule's dynamics without losing relevant details. In this doctoral thesis, a coarse graining method in time, by clustering motions of particles according to a certain criterion, and a coarse graining strategy in space, by revealing a hierarchy in the descriptors of a protein, are developed. To find the global statistical weights of the metastable sets, reweighting strategies are necessary. We present two novel techniques that permit a direct calculation of statistical weights, both based on the estimation of free energy differences. The first method requires an inverse balance condition that leads to a direct calculation of the statistical weights as eigenvalues of a transition matrix. The second approach approximates entropy differences between metastable sets to derive according free energy differences. The performance of the novel strategies is tested on several small examples before applying them to the μ-opioid receptor.

The μ-opioid receptor is a transmembrane protein, especially important for the perception and the alleviation of pain. A medication with analgesics like morphine can lead to serious side effects. Therefore, it is desirable to find an opioid that activates the μ-opioid receptor in the inflamed tissue only. While Spahn et al. [94] already found a prototype that binds to the μ-opioid receptor in the inflamed tissue only, the intent of this doctoral thesis is to analyze the μ-opioid receptor in order to find possible reasons for the known better effect of analgesics in inflamed tissue. We model the receptor in a simulation box, embedded in a membrane and solvated with water, both in healthy and inflamed tissue, propose different hypotheses that may cause the better effect of the receptor in inflamed tissue, and investigate electrostatic potentials, Coulomb interactions, and conformational changes for the analysis of the receptor's behavior.

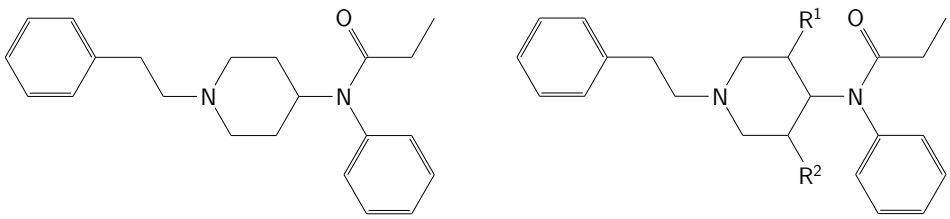
Contents

Chapter 1

Introduction

Many people suffer from acute or chronic pain. While some people have to struggle with their pain only temporarily, e.g. after surgery, during disease or after an accident, other people are afflicted with chronic pain their entire lives. Depending on the intensity, these pains can be medicated with different pharmaceuticals. For slight pains it is often sufficient to be treated with pain relieving drugs like paracetamol, whereas patients with acute pains, e.g. patients with end-stage cancer, need to be treated with a high dose of morphine to bear the pains at all. The annual sales of morphine are valued to about 40 billions USD [82]. Though, the use of analgesics is limited by serious, to some extent even life-threatening side effects. Especially opioids like morphine cross the blood-brain barrier and lead to tiredness, lack of concentration, memory defects, respiratory depression, nausea, addiction and tolerance in the central nervous system. Furthermore, they affect the endocrine system and restrict the activity of the intestine which leads to constipation and intestinal obstruction [118]. These side effects seem to be particularly dramatic considering that analgesics were used too often or wrongly during the last years worldwide. The consumption of opioids like morphine increased up to the tenfold in some regions [64]. Thus, the administration of opioids is heatedly discussed for several years. Thereby, not only abuse and overdosing (often with lethal consequence), but also illegal distribution, financial aspects of side effects and complications, as well as the strain of the health services are denounced. These debates have already lead to tighter legal restrictions in the United States of America [115]. Even in Germany, different expert associations worked out a guideline indicating the noxious side effects and the cautious use of analgesics [39].

The solution for the aforementioned problems would be the development of an opioid without noxious side effects but with comparable potency, an opioid that has an effect on the inflamed tissue only. A local injection of analgetics into the injured tissue involves the danger of an infection, furthermore, the active ingredient gets to the intestine anyhow, leading to the undesired side effects. Though,



(a) Fentanyl (b) NFEPP

Figure 1.1: Chemical structures of fentanyl and (±)-N-(3-fluoro-1-phenethylpiperidin-4-yl)-N-phenyl propionamide (NFEPP). Sites of fluorination are indicated as R^1 and R^2, compare [94].

an opioid shall be developed that can be administered systemically (e.g. orally or intravenously) but that acts only locally. From former research it is already known, that a decrease of the pH value in the affected tissue is one major characteristic of inflammation [49]. In the treatment with analgesics, the μ-opioid receptor (MOR) plays a fundamental role, since many analgesics, e.g. morphine and fentanyl, are binding to this receptor. For the activation of the MOR, there has to be a certain interaction between the receptor and the ligand (the analgesic) which requires a protonation of a specific N-group of the ligand. The pK_a value of this N-group is about 8 for most of the conventional analgesics, so that they activate the MOR both in the inflamed tissue (pH value of up to 5.5) and in the healthy tissue (pH value of 7.4). Though, a sufficient decrease of the pK_a value of the respective N-group could lead to an exclusive activation of the MOR in the inflamed tissue, even though the analgesic is administered systemically.

Within a cooperation of Charité Berlin and Zuse-Institute Berlin (ZIB) and with the help of computer simulations, a prototype (fluorinated fentanyl) was developed which activates the μ-opioid receptors in the inflamed tissue only. Thereby, the pK_a value of the specific N-group was reduced to 6 by replacing a proton near this N-group by a fluorine atom, see Figures 1.1(a) and 1.1(b). Theoretically, every opioid molecule can be fluorinated at several positions. The closeness to the N-atom and the number of substituted hydrogen atoms directly influence the decrease of the pK_a value. The optimal pK_a value, i.e. the value where the number of activated MOR is minimized in healthy tissue and maximized in inflamed tissue, depends on the efficiency of the particular opioid bound to the receptor, its solubility and pharmacokinetics. By means of computer simulations, several opioids with preferably different properties (i.a. solubility, efficiency, molecular geometry) were analyzed regarding their pK_a values, conformations, binding processes, pharmacokinetics and solubility. The aim was to develop an optimal opioid molecule with a pK_a value that is low enough to keep down the concentration of protonated molecules in healthy tissue, but high enough to ensure the activation of the

receptor in inflamed tissue. Thus, a novel agent was developed, (±)-N-(3-fluoro-1-phenethylpiperidin-4-yl)-N-phenyl propionamide (NFEPP), and already tested in vitro and in vivo, see [94]. Additionally, two patents were obtained [99, 98].

The challenge now is to analyze the μ-opioid receptor in order to gain a better understanding of the responsible mechanisms that lead to the activation of the receptor in the inflamed tissue. A better understanding of the behavior of the receptor provides the basis for the investigation of further analgesics with few or no side effects.

In this thesis, we will discuss possible reasons for the better behavior of the μ-opioid receptor in inflamed tissue. In chapter 2, we will introduce the receptor, illustrate the effect of pain and painkillers in the body, explain the computer-based molecular modeling and list the different hypotheses for the better effect of the receptor in inflamed tissue. The theory of molecular simulation is described in chapter 3, that is also used to introduce necessary variables and symbols. Since the μ-opioid receptor is a very complex molecule with lots of atoms, a fortiori embedded in a membrane and solvated with water, coarse graining methods for the dimension reduction of the state space both in time and in space are presented in chapter 4. These methods have already been published by the author [31]. Since single trajectories will not be able to sample the whole conformational space, we want to decompose the state space and start separate, parallel running trajectories. For the estimation of the global distribution, reweighting strategies are necessary. In chapter 5 direct strategies are presented which allow for the calculation of the weighting without producing excess data or approximating functions. The strategies have also been published before by the author [54]. After having explained the mathematical and chemical background, we will present and discuss the results of the analysis of the behavior of the μ-opioid receptor in inflamed and healthy tissue in chapter 6. Finally, in chapter 7, we will conclude this thesis with an outlook.

Chapter 2

The μ-opioid receptor, preliminaries

2.1 Introduction

The perception of pain is a complex and subjective sensory perception. However, the neural process of the painful stimulus, the so-called nociception, is an objective perception. The stimulation of sensory receptors, so-called nociceptors, produces a signal that is sent via the spinal cord to the brain. The nociceptors are free nerve endings that can be found both in external tissues like dermis and mucosa, and in internal organs like muscle or gut. If the nociceptors are stimulated, several ion channels are activated: calcium ion Ca^{2+} channels are opening and, because of the higher concentration of calcium ions outside of the cell, calcium ions are streaming inside, thus, activating the action potential. The action potential transfers the painful stimulus via the spinal cord to the brain, where the feeling of the pain arises in the cerebral cortex. In contrast, potassium ion K^+ channels are closing, hence, suppressing the opposed resting potential. The enzyme adenylyl cyclase converts adenosine triphosphate (ATP) to cyclic adenosine monophosphate (cAMP). This second messenger is important in many biological processes and used for intracellular signal transduction.

Analgesics like morphine help to alleviate the pain by binding to the transmembrane μ-opioid receptor (MOR). This binding leads to a conformational change of the receptor, so that the inhibitory G protein can bind to the intracellular C-terminus of the receptor. The G protein consists of three subunits G_α, G_β, G_γ, whereas the G_β and the G_γ can form a stable βγ-complex $G_{\beta\gamma}$. In the inactive state, guanosine diphosphate (GDP) is bound to G_α. By binding to the receptor, there is a conformational change in the G protein so that GDP dissociates and guanosine triphosphate (GTP) binds to the G protein in place of the GDP, which activates the G protein and leads to a dissociation into G_α and $G_{\beta\gamma}$. The subunit G_α binds to adenylyl cyclase, thus, inhibiting this enzyme and leading to a reduction of cyclic

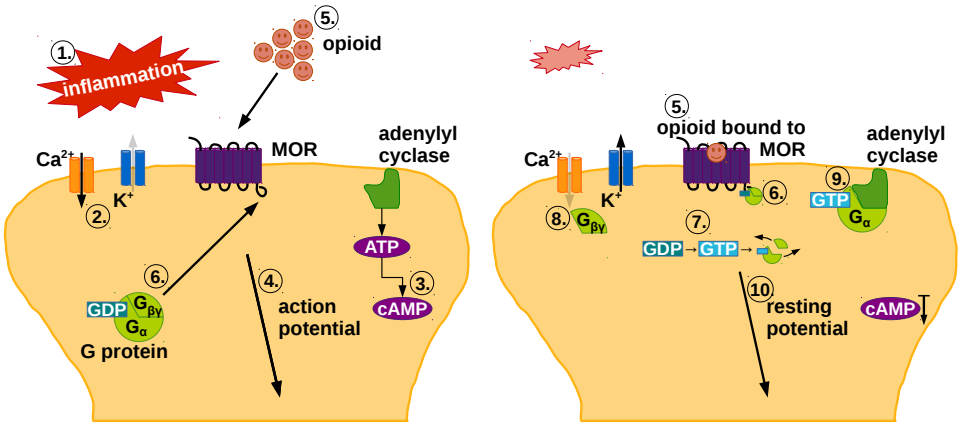


Figure 2.1: A very simple scheme to explain the transduction of pain (1-4) and the activation of the μ-opioid receptor by opioids (5-10): The stimulation of nociceptors by inflammation (1) leads to an opening of ion channels (2). The enzyme adenylyl cyclase converts ATP to cAMP (3) and acts as a second messenger, especially for signal transduction. The in this way initiated action potential (4) transfers the painful stimulus to the brain. If analgesics are taken, the binding to the μ-opioid receptor (5) results in a conformational change so that the inhibitory G protein can bind to the intracellular C-terminus (6). GDP dissociates to GTP (7) which leads to a dissociation of the G protein. The subunit $G_{\beta\gamma}$ induces the activation of ion channels (8), the subunit G_{α} inhibits adenylyl cyclase (9), causing a reduction of cAMP. In this way, the action potential is blocked and the resting potential (10) sends the pain relieving stimulus to the brain.

adenosine monophosphate (cAMP). The complex $G_{\beta\gamma}$ activates the calcium and potassium ion channels. The influx of calcium ions is constrained, which results in a decreased release of transfer agents, so that the action potential is blocked. On the other hand, the effusion of K^+ is increased, so that the excitability of neurons is reduced. In this way, the resting potential is initiated, which is the opposed part to the action potential, so that the pain relieving stimulus is sent to the brain. A very simple graphical representation is shown in Figure 2.1. More information about the μ-opioid receptor and its activation can be found here [2, 71, 95, 117]. Unfortunately, the salable analgesics like morphine take their effect not only in the inflamed tissue but in the healthy tissue as well, and, thus, lead to various side effects like nausea and addiction. It is already known that analgesics have a higher effect in inflamed tissue than in healthy tissue. This knowledge was used to develop a new type of analgesics that binds to the μ-opioid receptor in the inflamed tissue only [94]. In this doctoral thesis, the behavior of the μ-opioid receptor shall be investigated to understand the differing impact of analgesics to healthy and inflamed tissue, and to develop further analgesics with no or only few side effects.

2.2 Modeling

On the basis of the crystal structure of the μ-opioid receptor bound to a morphinan antagonist found in 2012 by Manglik et al. [65], we created a molecular model of the receptor. Thereby, the part of the morphinan antagonist, the β-funaltrexamine ($C_{25}H_{32}N_2O_6$), had to be deleted. This antagonist and five further small molecules, i.a. cholesterol, pentaethylene glycol and sulfate ion, were used to gain the crystal structure, i.e. to freeze the conformation of the μ-opioid receptor in order to study the crystal structure with X-ray diffraction, and are not necessary for our further investigations. Since there were missing atoms in the crystal structure, we tried to add them in the structure file. This was only possible for single atoms and not for whole amino acids, so that the model of the receptor is still incomplete. A molecular model of the μ-opioid receptor and its secondary structure is shown in Figure 2.3(a). This visualization was created with the molecular visualization program *Visual Molecular Dynamics* (VMD) [47].

In preparation for the simulation, we parametrized the molecule and put it into a simulation box with a lipid bilayer and water. The parameters that are used to calculate the potential energy are given by the force field which can be chosen according to the underlying molecular system and the aim of the molecular simulation. Because of the required membrane simulations, there are only a few force fields suitable, i.a. the *Coarse Grain Forcefield for Biomolecules* (MARTINI) [66], the *Assisted Model Building and Energy Refinement* (AMBER) [83] or the force field of the *GROningen MOlecular Simulation package* (GROMOS) [86, 74, 79, 81]. A good overview of force fields used in the last years for lipid membrane simulations is given by Lyubartsev and Rabinovich [63].

We decided to use the software package GROMACS [9] (*GROningen MAchine for Chemical Simulations*) with the force field GROMOS 53A6 [74]. Since normal force fields of GROMACS cannot handle both proteins and lipids, Berger et al. derived the Berger Lipids [10] which are the most widely-used parameters for the lipid part of membrane simulations under GROMACS. As lipid bilayer, we chose Dipalmitoylphosphatidylcholine DPPC [103], see Figure 2.2, This lipid bilayer belongs to the phospholipids. With the help of the PERL script InflateGRO [53], we packed the DPPC around the protein. Although water is a very small molecule with only three atoms, there are different models of water for molecular simu-

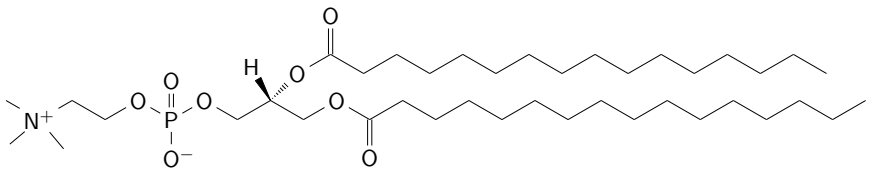
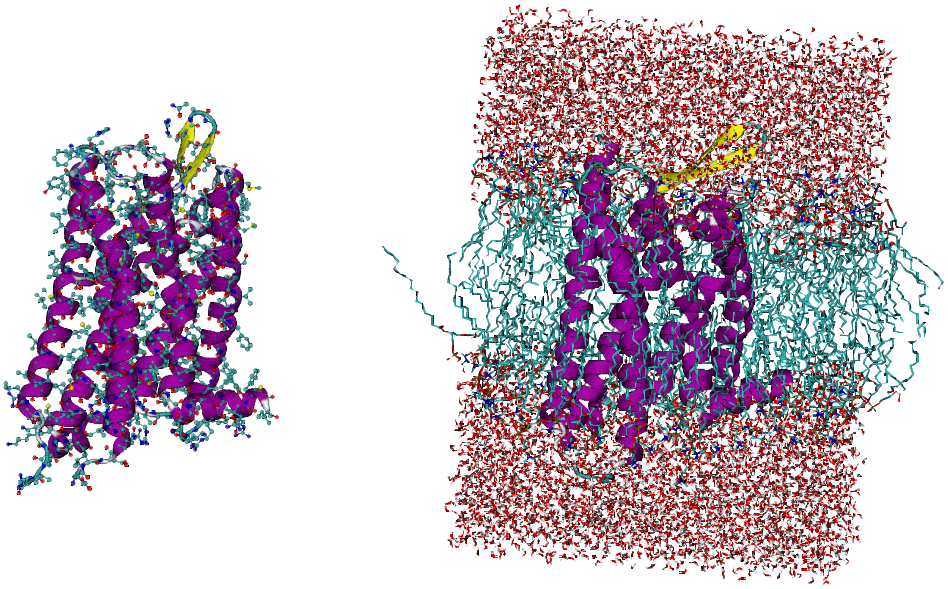


Figure 2.2: DPPC as lipid bilayer for the simulation box.



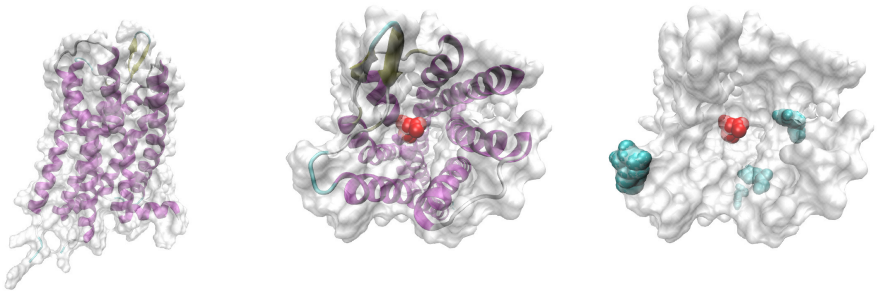
(a) MOR in vacuum.

(b) MOR in simulation box.

Figure 2.3: VMD snapshots of μ-opioid receptor (MOR). Secondary structure in purple. In (b) the tails of the lipid bilayer are represented by turquoise lines, water molecules are in red/white.

lations with diverse complexity. We decided for the spc water model [8]. The following solvation with water was a bit tricky since gaps in the lipid acyl chains might be filled with random water molecules. By increasing the van der Waals radius of carbon, the addition of water within the lipids gets less likely, so that the few stray molecules can be deleted manually. Needless to say, the radius has to be reset to the default value afterwards. After adding ions to neutralize the system, and minimizing the energy to exclude inappropriate geometry, we had to run an equilibration to stabilize the temperature and the pressure of the system. This equilibration phase was also necessary for the lipids to orient themselves around the protein, and for the water molecules to reorient around the lipid headgroups and any exposed parts of the protein. Finally, the simulation box consisting of 31640 atoms, thereof 2950 atoms belonging to the receptor, can be used for molecular dynamics simulations with GROMACS, see the VMD snapshot in Figure 2.3(b). The total number of atoms depends on a variety of factors, and will differ in individual modeling phases.

Another abstract representation of the receptor is achieved by calculating its molecular surface, see Figure 2.4(a), which is important for computing molecular interactions and for displaying molecular properties on the surface, i.e. electrostatic potential, atomic charge and hydrophobicity. There are different surface models, the simplest one is the van der Waals (vdW) surface, the first extension is the sol-



(a) Front view. (b) View into binding pocket. (c) View into binding pocket.

Figure 2.4: The molecular surface (white/gray) of the MOR visualized with VMD. For orientation, the secondary structure is drawn in purple. The aspartate is highlighted in red as the center of the binding pocket; the histidines, that are protonated in the model of the inflamed tissue, are highlighted in turquoise.

vent accessible surface (SAS), and the first smooth molecular surface is the solvent excluded surface (SES). A good overview is given by Kozlíková et al. [55]. The molecular surface in VMD is calculated on the basis of SAS. With this graphical representation, the binding pocket can be visualized, see the top view in Figure 2.4(b). Here, the aspartate Asp147 is highlighted in red as the center of the binding pocket [58, 90].

2.3 Comparing healthy and inflamed tissue

To compare the different behavior of morphine in healthy and inflamed tissue, the μ-opioid receptor has to be modeled in inflamed tissue as well. Therefore, the pH value is important, which is a numeric measure to characterize the acidity or basicity of an aqueous solution. It is defined as the negative decimal logarithm of the hydrogen ion activity in a solution

$$\mathrm{pH} = -\log_{10}(a_{\mathrm{H}^+}).$$

Since a value of pH = 7 (e.g. pure water) is classified as neutral, blood in healthy tissue is slightly alkaline with a pH value of pH = 7.4, and blood in inflamed tissue is acidic with a value of approximately up to pH ≈ 5.5. To model the μ-opioid receptor in inflamed tissue, the lower pH value has to be transferred to the molecular model. This can be achieved by adding additional hydrogens to amino acids that have a $\mathrm{p}K_\mathrm{a}$ value between 5.5 and 7.4. The $\mathrm{p}K_\mathrm{a}$ value is the negative decimal logarithm of the acid dissociation constant

$$\mathrm{p}K_\mathrm{a} = -\log_{10}(K_\mathrm{a}),$$

i.e. the smaller the pK_a value, the more acidic is the solution. With the definition of the acid dissociation constant as

$$K_a = \frac{[A^-][H^+]}{[HA]},$$

where HA is a generic acid that dissociates into A^-, it is obvious, that an increase of hydrogens H^+ leads to a higher value of K_a, and, thus, a lower pK_a value. Since the amino acids in the receptor are bound to each other, only the pK_a values of the side chains are relevant. With a pK_a value of the side chain of approximately p$K_a \approx 6$, see e.g. [102], the amino acid histidine is the only one that is affected by a decrease of the pH value from 7.4 to 5.5. Regarding the μ-opioid receptor, the amino acid histidine can be found four times, three of them are lying in the center of the binding pocket or nearby, see Figure 2.4(c). In practice, the parametrization of the μ-opioid receptor in inflamed tissue with GROMACS is expanded to the choice of the histidine type. Selecting *HISH* for each of the four histidines leads to the desired result. With the changed files for the structure and the topology, the whole modeling of the last section has to be rerun in order to generate a similar simulation box with a lipid bilayer and water. Note that, because of the protonation of the four histidines, the number of ions to neutralize the system has to be greater by four than for the molecular model of the receptor in healthy tissue.

2.4 Hypotheses

The μ-opioid receptor plays an important role both in the perception and in the alleviation of pain. We want to study the behavior of this receptor in healthy and inflamed tissue to understand the better effect of analgesics in inflamed tissue, and to find possible reasons therefor. In this chapter, the molecular modeling of the μ-opioid receptor was shown, and the protonation of the receptor resulting from the characteristics of the inflamed tissue was discussed. Concerning the simulation in the inflamed tissue, we assumed that the receptor is the only part in the simulation box that is pH-dependent since the appearance of H_3^+ ions in such a little box is rather improbable, the headgroups of the lipids are not pH-dependent, and the rest has no charge at all. Now, if we investigate the better effect of analgesics in inflamed tissue, we can assume that this results either from differing properties of the μ-opioid receptor or from varying characteristics of the analgesic or from different interactions between ligand and receptor. Summarizing, there are four hypotheses that might explain the better drug-receptor-interactions in inflamed tissue:

(1) The differing Coulomb interactions between ligand and receptor enable the analgesic to find its way into the binding pocket of the μ-opioid receptor better in the inflamed tissue.

(2) The conformation of the μ-opioid receptor in the inflamed tissue is more suitable for the analgesic than in healthy tissue.

(3) Changes in intracellular pH value lead to changes in intracellular signaling.
(4) The better effect depends on the ligand, i.e. the analgesic only.

Isom et al. showed 2013 [48] that the G_α subunit detects changes in the intracellular pH value and transduces signals of the G protein coupled receptor. Their results indicate that changes in intracellular pH value can provoke a disruption of G protein-mediated signaling, i.e. one reason for the better effect of analgesics in inflamed tissue can probably be found in changes of intracellular pH value (3). The development of a new analgesics by Spahn et al. [94] that binds to the receptor in the inflamed tissue only, shows that also the chemical properties of the ligand have an influence on the effect of the analgesics (4). In this thesis, we will investigate the μ-opioid receptor concerning its effect in inflamed tissue to analyze assumptions (1) and (2). Before we start with this analysis in chapter 6, an introduction to the theory of molecular simulation shall be given. We want to explain the most significant terms and definitions, discuss different techniques and present new strategies.

Chapter 3

Theory of molecular simulation

3.1 Introduction

Considering a canonical ensemble, also abbreviated as (N,V,T), that is a system where the number of atoms N, the volume V and the temperature T is constant. The latter is achieved by a heat bath at a fixed temperature. The state $x \in (\Omega,\Gamma) \subset \mathbb{R}^{6N}$ of a molecule is described by its position coordinates $q \in \Omega$ and momenta $p = Mv \in \Gamma$ with the mass matrix $M \in \mathbb{R}^{3N\times 3N}$ and the velocity vector $v \in \mathbb{R}^{3N}$. The probability for a molecule to be in a state x is defined by the Boltzmann distribution

$$\pi(x) = \pi(q,p) = \frac{\exp\left(-\beta H(q,p)\right)}{\int\limits_{\Omega}\int\limits_{\Gamma} \exp\left(-\beta H(\tilde{q},\tilde{p})\right) \mathrm{d}\tilde{x}} \tag{3.1}$$

with the thermodynamic beta $\beta = 1/(k_B T)$, the Boltzmann constant k_B, the differential $\mathrm{d}\tilde{x} = \mathrm{d}\tilde{q}\,\mathrm{d}\tilde{p}$ and the total energy $H(q,p) = E_k(p) + E_p(q)$ as the sum of the kinetic energy $E_k(p)$ and the potential energy $E_p(q)$. Since H is obviously separable, it yields

$$\begin{aligned} \pi(q,p) &= \underbrace{\frac{\exp\left(-\beta E_k(p)\right)}{\int\limits_{\Gamma} \exp\left(-\beta E_k(\tilde{p})\right) \mathrm{d}\tilde{p}}}_{\pi_p(p)} \cdot \underbrace{\frac{\exp\left(-\beta E_p(q)\right)}{\int\limits_{\Omega} \exp\left(-\beta E_p(\tilde{q})\right) \mathrm{d}\tilde{q}}}_{\pi_q(q)} \\ &= \pi_p(p) \cdot \pi_q(q) \end{aligned} \tag{3.2}$$

While the first term is a rather simple quadratic function, the second term is too complex to be computed analytically, particularly because of the partition function $Z_q = \int_{\Omega} \exp\left(-\beta E_p(\tilde{q})\right) \mathrm{d}\tilde{q}$. The spatiotemporal development of a molecule is described by the Hamilton's equations

$$\begin{aligned} \dot{q}(t) &= \frac{\partial H}{\partial p} = M^{-1} p(t) \\ \dot{p}(t) &= -\frac{\partial H}{\partial q} = -\nabla_q E_p(q(t)) \end{aligned} \tag{3.3}$$

which is a system of ordinary differential equations of first order with three important properties:

(1) reversibility: if $(q,p) \xrightarrow{\pi} (\tilde{q},\tilde{p})$ then $(\tilde{q},-\tilde{p}) \xrightarrow{\pi} (q,-p)$
(2) symplecticity: $d\omega = 0$ with $\omega = \mathrm{d}p \wedge \mathrm{d}q$
(3) energy conservation: $H(q(t),p(t)) = H(q(0),p(0)) \quad \forall t \in (-\infty,\infty)$

These equations of motion require infinitesimal time steps for an exact solution which is not feasible for complex molecular systems. In the following section, numerical approximations are presented.

3.2 Molecular dynamics and stochastic thermodynamics

Numerical solutions based on a discretization of time $t_{i+1} = t_i + \Delta t$, $i = 1,\dots,n-1$, with time step Δt are called molecular dynamics (MD). A very simple approach to solve equations (3.3) is the Euler method

$$\begin{aligned}
\frac{q(t+\Delta t)-q(t)}{\Delta t} &\approx \dot{q}(t) = \quad M^{-1}p(t) \qquad &\Rightarrow\ q(t+\Delta t) \approx q(t) + \Delta t M^{-1}p(t) \\
\frac{p(t+\Delta t)-p(t)}{\Delta t} &\approx \dot{p}(t) = -\nabla_q E_p(q(t)) \quad &\Rightarrow\ p(t+\Delta t) \approx p(t) - \Delta t \nabla_q E_p(q(t))
\end{aligned}$$

While this method destroys the physical properties (no reversibility, no symplecticity, no energy conservation), other integrators like the velocity Verlet [100]

$$\begin{aligned}
q(t+\Delta t) &\approx q(t) + \Delta t M^{-1}p(t) + \frac{\Delta t^2}{2} M^{-1}\nabla_q E_p(q(t)) \\
p(t+\Delta t) &\approx p(t) + \frac{\Delta t}{2}\left(\nabla_q E_p(q(t)) + \nabla_q E_p(q(t+\Delta t))\right)
\end{aligned}$$

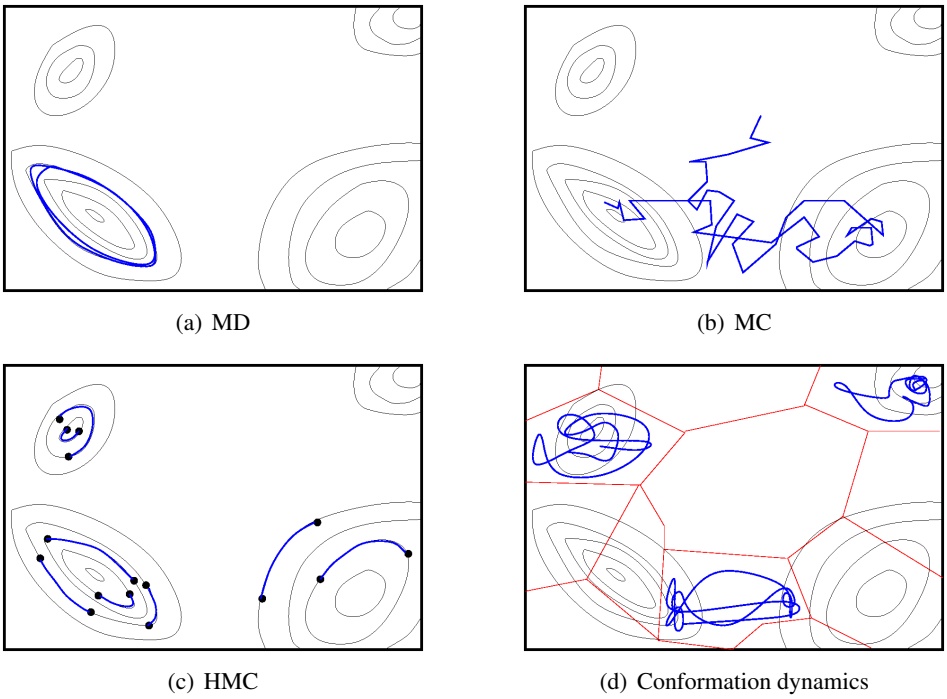
conserve the important characteristics. Anyway, the in this way generated long trajectories are mathematically ill-conditioned because of cumulative errors. Additionally, these trajectories are trapped in local minima because of the energy conserving property, see Figure 3.1(a).

Besides deterministic approaches, there are stochastic methods as well. Instead of a trajectory, a Markov Chain [20] is generated

$$q_1 \xrightarrow{P(q_1 \to q_2)} q_2 \xrightarrow{P(q_2 \to q_3)} \dots \xrightarrow{P(q_{n-1} \to q_n)} q_n$$

with a transition probability $P(q_i \to q_{i+1})$. These Markov Chains are random and memoryless, i.e. state q_{i+1} depends on state q_i only. Methods that are based on Markov Chains are leading to the thermodynamically correct distribution if the transition probability is ergodic and fulfills the thermodynamic equilibrium, also called detailed balance condition

$$\pi(q)\,P(q \to \tilde{q}) = \pi(\tilde{q})\,P(\tilde{q} \to q) \tag{3.4}$$



(a) MD (b) MC

(c) HMC (d) Conformation dynamics

Figure 3.1: A sketch of MD vs. MC vs. HMC vs. conformation dynamics in phase space.

which means that the reciprocal process is equiprobable. Using equation (3.2), the detailed balance condition (3.4) can be stated as

$$\frac{P(q \to \tilde{q})}{P(\tilde{q} \to q)} = \exp\Big(-\beta\Big(E_p(\tilde{q}) - E_p(q)\Big)\Big) = \exp\Big(-\beta\,\Delta E_p\Big). \tag{3.5}$$

One prominent method based on these Markov Chains is the Metropolis Monte Carlo [69]. The basic idea is that the transition probability is divided into a proposal and an acceptance step, i.e. $P(q \to \tilde{q}) = P^{(p)} \cdot P^{(a)}$. Choosing a symmetric proposal step $P^{(p)}(q \to \tilde{q}) = P^{(p)}(\tilde{q} \to q)$, the restated detailed balance condition (3.5) leads to the Metropolis acceptance probability $P^{(a)}(q \to \tilde{q}) = \min\{1, \exp(-\beta\Delta E_p)\}$. Although this method converges to the correct distribution, the dynamics are not physically correct mapped, and the algorithm is very slow, see Figure 3.1(b).

Combining both deterministic and stochastic methods, the Hybrid Monte Carlo (HMC) [30] is more effective and robust. The proposal step is not symmetric but generated by a short MD simulation with an initial impulse $p \propto \pi_p$ starting from q. The integrator for this short MD has to be reversible and symplectic. Using the reversibility of the proposal step in equations (3.5) and (3.2), the acceptance probability yields $P^{(a)}(q \to \tilde{q}) = \min\{1, \exp(-\beta\Delta H)\}$. For this method, a convergence criterion was developed by Gelman and Rubin [33]. Starting with at least two parallel running trajectories, convergence is reached if the variance within each

chain is approximately the same as the variance across the chains. Although this approach leads to the correct distribution, and the well-conditioned short-time trajectories represent the correct physics, see Figure 3.1(c), it is still difficult, using this method, to sample the whole phase space because of high energy barriers, high dimensions and limited computing time.

Since the trajectory-based, global sampling strategies are not suitable to analyze the conformational space of complex molecular structures, we want to present sampling strategies on the basis of a decomposition of the state space in the following section.

3.3 Conformation dynamics

The term conformation dynamics was formed by Deuflhard, Schütte et al. [27, 25, 28, 87, 26]. Their set-based approach was motivated by papers of Dellnitz et al. [23, 24], and is nowadays mostly called Markov State Models (MSM).

For the introduction of meshless methods in conformation dynamics [111], the term metastability has to be defined. Metastabilities are subsets of the phase space where the configurations of the molecule remain constant for a certain period of time, basically. In conformation dynamics, the dynamics of a molecule is described by transition probabilities between metastabilities. Based on a smart decomposition of the state space Ω into non-intersecting subsets B_i, $i = 1,\dots,b$, with $\Omega = \bigcup_{i=1}^{b} B_i$, local samplings are run in each subset B_i, possibly in parallel, see Figure 3.1(d). An overlapping of subsets may lead to a poor sampling. The restriction of the local samplings to the corresponding subsets can be achieved by modifying the (local) potential functions or by simply rejecting proposed states that are not lying inside the respective subset. As the transition probabilities between the subsets do not match with the statistical weights $w_i = \int_{B_i} \pi_q(q)\,\mathrm{d}q, i = 1,\dots,b$, we developed direct reweighting strategies for the estimation of the global distribution that are presented in chapter 5.

Since the state space is not known in advance, it is difficult to decompose it into non-intersecting subsets previous to the local samplings without additional analytical techniques. Often used is a fast presampling to examine the potential energy landscape E_p. This can be done with a high-temperature HMC trajectory to overcome as much energy barriers as possible. Although such a high-temperature sampling does not lead to the correct distribution of states, the sampling points will be, for the most part, physically reasonable. The selection of the nodes out of the HMC sampling points can be implemented with a clustering algorithm like k-means [61] based on internal coordinates, e.g. dihedral angles or bond lengths. These nodes mark the centers of the subsets B_i and are used to set up radial basis functions that can reach values between 0 and 1. In this way, these basis functions are indicating the belonging of a configuration q to the subset. The potential functions can be modified so that the following local samplings stay inside the sup-

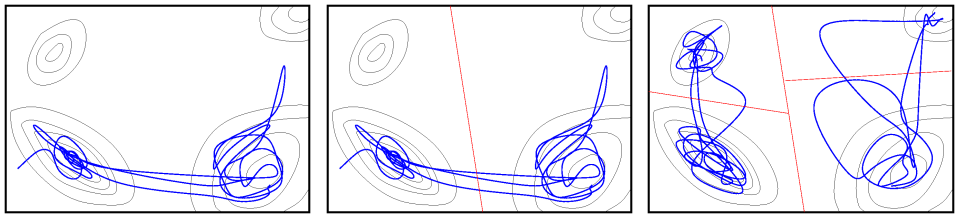


Figure 3.2: An adaptive strategy: starting with a global sampling, the state space is divided according to detected local minima.

port of the basis functions. If a local sampling does not converge in terms of the Gelman-Rubin criterion, the node can be refined or expanded later on, and the local sampling can be restarted.

However, in case of proteins, high-temperature samplings are not feasible since proteins are denaturing at high temperature so that the molecular system drifts apart. Thus, for the molecular simulation of the μ-opioid receptor we have to find another way to generate starting points for the parallel running local samplings. This will be explained in section 6.3.

More efficient would seem an adaptive strategy where the decomposition of the conformational space happens during the sampling. Starting a global sampling, if local minima are found, the state space is divided so that each local minimum has its own subset, and local samplings are continued in each subset, compare Figure 3.2. This is difficult to implement because transitions are rare events and, thus, a long simulation time would be necessary to find all minima. In the next chapter, a coarse graining method for the dimension reduction in time is presented.

Chapter 4

A coarse graining method for the dimension reduction of the state space

4.1 Introduction

From the trajectories of a molecular dynamics (MD) simulation the position and momenta of a molecule on the atomistic level are given. Since the first article about MD simulations [3], much effort has been put into the development of fast and reliable algorithms [5, 9, 109, 105, 106] and their parallelization [35, 40, 67, 78]. However, in trajectory based MD, the maximum size of the integration step is bound to femtoseconds whereas protein folding, for instance, ranges in microseconds which implies the simulation of long trajectories [62, 92] or multiple trajectories [6, 70, 116]. Moreover, having the data of a long trajectory, it is still unclear how to interpret these data in the high dimensional state space.

Focussing on the time scale, we want to establish a coarse graining strategy in time by considering conformational changes only instead of using confining small time steps in a trajectory. Based on Markov State Models (MSM) [12, 13, 17, 18, 19, 32, 88], the dynamics of a molecule can be described in terms of transition probabilities between metastable sets, taking advantage of the fact, that for suitable chosen time steps the transitions appear stochastic and memoryless. Since metastable states can be identified as subsets in the high dimensional conformation space, we will decompose the state space into disjoint subsets. Following the approach of [87] and applying a transfer operator $\mathcal{T}$, the metastable decomposition of the state space can be represented by linear combinations of characteristic or meshfree basis functions employing a Galerkin discretization of $\mathcal{T}$ [28, 88]. We will discuss this approach in section 4.2.

To reduce the high dimensionality of the state space by still keeping relevant information, we will describe conformational changes in a lower dimensional space using dynamical informations of a trajectory only. In general, descriptors describe the structure-activity relationship in molecules, they can be defined as a collection

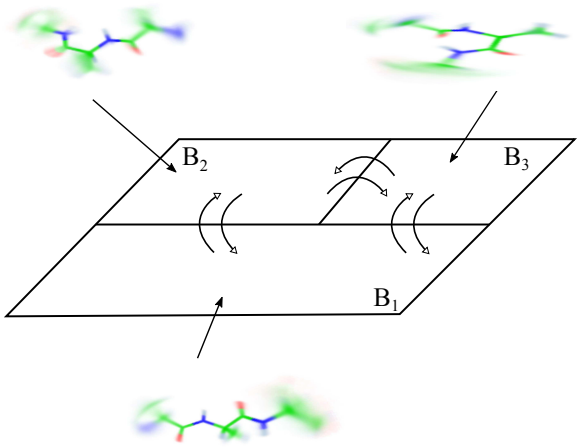


Figure 4.1: Sketch of a position space Ω and its metastable subsets B_1, B_2 and B_3.

of data characterizing the molecule [104]. In contrast to QSAR methods [11, 22], we will not use descriptors to characterize or compare structures of molecules but to describe the dynamics of a molecule using less degrees of freedom. For instance, the relevant dynamics (conformational changes) of the well-known alanine dipeptide can be described by using the two dihedral angles ϕ and ψ only. In section 4.3, we present the coarse graining strategy in space, the Hierarchical Relevant Descriptor Detector (HRDD).

4.2 Metastability and coarse graining in time

A molecule passes through intermediate states on its way from the initial to the final state. Thereby, a protein in an intermediate state is not stable but almost stable (metastable), which means that the molecule stays for a certain time span in this configuration before it switches to another state. In the following, we give the above mentioned terms metastability and conformational change a mathematical foundation. Thereby, employing MSM, metastable states can be identified as subsets in the phase space, and the dynamics as transition probabilities of a transition matrix. More precisely, the entries of the transition matrix give the probability that a system being in metastable state B_1 switches to a metastable state B_2, compare Figure 4.1. Note that the different configurations belonging to a certain metastable state are only kinetically closely related, but not necessarily "geometrically" or "homotopically" similar. However, Hammond's postulate [38] states that energetically closely related configurations are also geometrically similar.

Mathematically, we understand a conformation as a part of the conformation space, which comprises the collection of structurally related configurations. Now, considering a molecular system of N atoms with the spatial coordinates $q_i \in \mathbb{R}^3$,

$i = 1,\dots,N$, and their N generalized momenta $p_i \in \mathbb{R}^3$, then $(q,p) \in (\Omega,\Gamma) \subset \mathbb{R}^{6N}$ is the phase space and Ω the position space. The states in the phase space are distributed according to the Boltzmann distribution $\pi(q,p)$ in equation (3.1). The canonical density can be split into a distribution of momenta $\pi_p(p)$ and positions $\pi_q(q)$, compare equation (3.2). A direct computation of the distribution is often hardly feasible, since for its evaluation a high dimensional integral has to be solved. For the dynamics of the system we can employ a corresponding flow Φ^τ for a time span $\tau > 0$. This flow is given by

$$(q_i,p_i) = \Phi^{i\tau}(q_0,p_0), \qquad i = 1,\dots,N.$$

Let Π_q be the projection of the state (q,p) onto the position q and let further p be chosen randomly according to the distribution $\pi_p(p)$, then

$$q_{i+1} = \Pi_q \Phi^\tau(q_i,p_i)$$

describes a Markov process. Note, that the ith state depends on the preceding state only. It can be shown that this assumption of Markovianity implies the time independence of the corresponding transfer operator, e.g. [87]. Hence, we will define the above described metastability in a mathematical framework. Therefor, the Boltzmann weighted scalar product $\langle f,g\rangle := \int f(q)g(q)\pi_q(q)\,\mathrm{d}q$ and the characteristic function $\chi_B(q)$, being 1 if $q \in B$ and 0 otherwise, have to be introduced. Then, the conditional probability of the system to move during time span τ from subset B_1 to subset B_2 can be defined by

$$p(B_1,B_2,\tau) = \frac{\langle \chi_{B_1}, \mathcal{T}^\tau \chi_{B_2}\rangle}{\langle \chi_{B_1}, \chi_{B_2}\rangle}$$

with the transfer operator

$$\mathcal{T}^\tau f(q) = \int_\Gamma f(\Pi_q \Phi^{-\tau}(q,p))\pi_p(p)\,\mathrm{d}p \tag{4.1}$$

which has been introduced by Schütte [87]. The operator in (4.1) can be explained as follows: Applying the Hamiltonian dynamics backwards to a lag time τ, $\Phi^{-\tau}(q,p)$ is obtained and projected by Π_q onto the state space. The integral then, averages over all possible initial momentum variables with given Boltzmann distribution π_p. These tools enable to characterize a subset $B \subset \Omega$ metastable if it is almost invariant under the transfer operator $\mathcal{T}^\tau$, i.e.

$$p(B,B,\tau) = \frac{\langle \chi_B, \mathcal{T}^\tau \chi_B\rangle}{\langle \chi_B, \chi_B\rangle} \approx 1 \tag{4.2}$$

which can be restated as

$$p(B,B,\tau) \approx 1 \iff \langle \chi_B, \mathcal{T}^\tau \chi_B\rangle \approx \langle \chi_B, \chi_B\rangle.$$

Here we can attain to two conclusions: For the first, metastable sets can be identified by computing the eigenfunctions of the propagator $\mathcal{T}^\tau$. For the second, the more the fraction in (4.2) is close to one, the more metastable is the set in B. Assuming, that the space can be partitioned into b metastable sets, i.e. $B_1, ..., B_b$, which approximate metastable eigenfunctions of the operator $\mathcal{T}^\tau$. Their characteristic basis functions are given by $\chi_{B_1}, ..., \chi_{B_b}$. Thus, the dynamics of the system can be approximated by the transition probabilities between the metastabilities. The resulting linear operator $P : \text{span}(\chi_{B_1}, ..., \chi_{B_b}) \to \text{span}(\chi_{B_1}, ..., \chi_{B_b})$ is given by the stochastic matrix

$$(P_{ij})_{i,j=1,...,b} \text{ with } P_{ij} = \frac{\langle \mathcal{T}^\tau \chi_{B_i}, \chi_{B_j} \rangle}{\langle \chi_{B_i}, \chi_{B_i} \rangle}. \tag{4.3}$$

According to the celebrated Perron-Frobenius theorem, e.g. [51], the row stochastic matrix P has the eigenvalues $\lambda_1, ..., \lambda_b$, which can be arranged such that $\lambda_1 > \lambda_2 \geq ... \geq \lambda_b$. Moreover, let v_i be the b dimensional right eigenvector of eigenvalue λ_i, and u_i the left eigenvector, then $\lambda_1 = 1, v_1 = e = (1, \ldots, 1)^T$, $u_1 = \pi_b$ and $\lambda_1 > \lambda_i \quad i = 2, \ldots, b$. Here, π_b is a b-dimensional vector, whose elements are the stationary probabilities of the b characteristic functions. Hence, the right eigenvectors $v_1, \ldots, v_b$ form an eigenbasis which can be used to express any vector x as

$$x = \sum_{i=1}^{b} \alpha_i v_i = \sum_{i=1}^{b} \langle x, v_i \rangle v_i \,.$$

Since $Pv_j = \lambda_j v_j$, it yields

$$Px = P \sum_{j=1}^{b} \langle x, v_j \rangle v_j = \sum_{j=1}^{b} \langle x, v_j \rangle P v_j = \sum_{j=1}^{b} \lambda_j \langle x, v_j \rangle v_j.$$

Thus, we can explain the dynamics of a molecule in terms of transition probabilities (given by matrix P) between subsets in the conformation space.

4.3 Coarse graining in space: HRDD

The introduction of the conformation concept allowed us for a coarse graining in time, which we established in the foregoing section. Here, we exploit, that in most molecules a conformational change can be detected by very few descriptors. One prominent example of this fact is the Ramachandran plot, where even larger molecules can be described admissible by using only two descriptors per amino acid. In the following, we introduce a novel method, which reveals a hierarchy in the descriptors of a molecule and, thus, allows to map the relevant motions (e.g. conformational change) by using only a few degrees of freedom (descriptors). Having such a hierarchy, we can start with a one dimensional model, by taking only the first (and most relevant) descriptor. Using the second most relevant descriptor we

can approximate the motion of the integral by a two dimensional model and so on. This method does not use any deeper chemical details, in fact, it only relies on the long-term trajectory. The spatial structure of each molecule can be characterized by its descriptors $D^i, i = 1, \ldots, n_d$, which are a collection of all internal degrees of freedom of the molecule such as distances between two arbitrary atoms, bond or dihedral angles between three or four bonded atoms. However the full number of descriptors of a molecule n_d to specify it, would lead to an over-determined system, which means that we have a redundancy in the descriptors. For instance, pentane consists of 17 atoms, thus, it has $17 \times 3 - 6 = 45$ degrees of freedom. Even if we only consider all possible distances between two atoms, we obtain $17 \times 16/2 = 136$ possible descriptors. This also leads to the assumption, that the descriptors of a molecule adhere a hierarchy, according to their influence on the spatial structure of the molecule. As a consequence, only a fraction of all possible descriptors, the *relevant descriptors* suffice to characterize the spatial structure and, thus, we do not need the high dimensional full phase space, but a low dimensional space, the so called *conformational space*.

The detection of the relevant descriptors can either be done by employing detailed chemical expert knowledge or by analyzing the time series of the molecule's dynamics. In the following, we concentrate ourselves to the latter by introducing the Hierarchical Relevant Descriptor Detector (HRDD). Let us, therefore, assume that we have a sampling covering the whole conformational space of a molecule with at least two conformations.

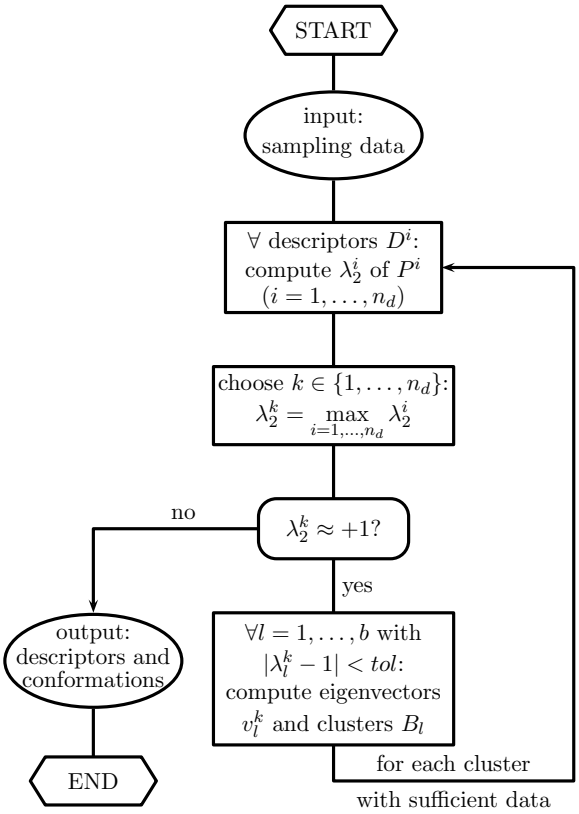
For all possible and given descriptors D^i, $i = 1, \ldots, n_d$, we employ a fine uniform discretization of (the subset of) the conformational space represented by the sampled data. For the ease of notation we assume that the number of discretization intervals (bins) b is equal for all descriptors, though, we remark that our method allows for different dimensions in each descriptor.

For a given trajectory, we calculate the transition matrix $P^i \in \mathbb{R}^{b \times b}$, for each descriptor D^i, i.e. increase the entries of P^i_{jk} and P^i_{kj} by one if the trajectory switches within two consecutive time-steps from bin j to bin k. In order to gain a row-stochastic matrix, that is $\sum_{k=1}^{b} P^i_{jk} = 1 \ \forall j = 1, \ldots, b$, we normalize each row by dividing each entry by its row-sum. Following the discussion of the foregoing section we compute for each matrix P^i the corresponding second largest eigenvalue λ_2^i and select the respective descriptor D^k with

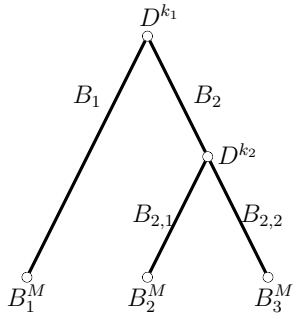
$$\lambda_2^k = \max\{\lambda_2^{(i)} \,|\, i = 1, \ldots, n_d\}, \tag{4.4}$$

i.e. largest of the second largest eigenvalues.

If this λ_2^k is close to 1, then the corresponding descriptor D^k characterizes at least two metastable conformations within the sampling, and we can continue to calculate the actual number of metastabilities in this descriptor D^k. Otherwise, the algorithm stops in this level (but might continue in another branch, compare



(a) Flow chart.



(b) Rooted tree.

Figure 4.2: The Hierarchical Relevant Descriptor Detector (HRDD): a recursive algorithm requiring sufficient sampling data of a molecule with at least two conformations supposed.

Figure 4.2(b)). The number of clusters n_c is gained by estimating all eigenvalues λ_l^k, $l = 1, \ldots, b$, which fulfill

$$|\lambda_l^k - 1| < tol,\ l = 1, \ldots, n_c. \tag{4.5}$$

The tolerance value must satisfy equation (4.5) for $l = 1, 2$, since we already required that λ_2^k is close to 1 and, thus, at least two metastabilities exist. We remark, that this tolerance has to be chosen with care: If the tolerance is too large, the number of clusters is large and, thus, we might have a poor reduction of the dimension. However, if the tolerance is too small, relevant dynamical information of the molecule might get lost.

For the decomposition of the space into clusters B_l, $l = 1, \ldots, n_c$, we compute the corresponding eigenvectors v_l^k, $l = 1, \ldots, n_c$, building a basis according to the respective eigenvalues λ_l^k. Finally the algorithm is recalled for each cluster found in the last iteration that covers sufficient data. For illustration purpose, Figure 4.2(a)

depicts a flow chart of the recursive algorithm.

Each non decomposable cluster is related to a metastable conformation. Starting from the top, the number of branches is given by all second largest eigenvalues satisfying *tol*. Going down one arm corresponds to "freezing" the first metastable state described by the most relevant descriptor. In the next branch the metastabilities except for the frozen parts are computed. For the illustration a rooted tree shall be drawn, see Figure 4.2(b). The number of branches on each level ("width") is given by the number of eigenvalues according to criterion (4.5). The height of the tree is determined by (4.4).

Each leaf is a not decomposable cluster and, thus, a metastable conformation B_l^M. Each parent corresponds to a descriptor. The higher the tree-order (height of a tree from root to leaf) the more descriptors are necessary to characterize a metastable conformation. In Figure 4.2(b) the algorithm finds the descriptor D^{k_1} in the first instance, which decomposes the whole conformational space into two clusters B_1 and B_2. The first one is not decomposable and, thus, a metastable conformation B_1^M characterized by the single descriptor D^{k_1}. The second cluster is most metastable in descriptor D^{k_2} and comprises two clusters $B_{2,1}$ and $B_{2,2}$, which prove to be metastable conformations B_2^M and B_3^M characterized by both D^{k_1} and D^{k_2}.

We remark, that such a hierarchical scheme has also been applied in the context of temperature and decomposition of the conformation space [32].

4.4 Application

For testing the applicability of the method, we analyzed the conformations and descriptors of two small molecules, pentane and alanine dipeptide. Both molecules have been investigated years before, see [87] and [93] respectively.

To obtain the data for our algorithm, we ran different trajectories with Hybrid Monte Carlo (HMC) [30] and MD in vacuum at a temperature of $T = 300\,\mathrm{K}$. In advance, we parameterized the molecules according to different force fields. In case of HMC simulation we used the Merck molecular force field (MMFF) [37] for both. For the MD simulation with GROMACS we chose the Amber Force Field ffAmber99sb [43] for pentane and the Optimized Potential for Liquid Simulations All-Atoms (OPLS-AA) [52] for alanine dipeptide.

For the HMC samplings, five chains were started in parallel, for each HMC step a 60 fs MD trajectory was calculated to generate a trial state. Altogether, the samplings covered 6 ns for pentane and 12 ns for alanine dipeptide. We reached an average acceptance probability of more than 95 %. The convergence was monitored by the Gelman-Rubin acceptance criterion with a threshold value of 1.2.

In order to generate MD trajectories we used GROMACS [9, 59, 108, 40] with a modified Berendsen thermostat [14], where the correct distribution of the energy

Table 4.1: The eigenvalue tolerances of the pentane simulation.

	$\lambda_2^k \approx +1$	$\lvert\lambda_l^k - 1\rvert < tol$
HMC	0.05	0.1
MD	0.01	0.05

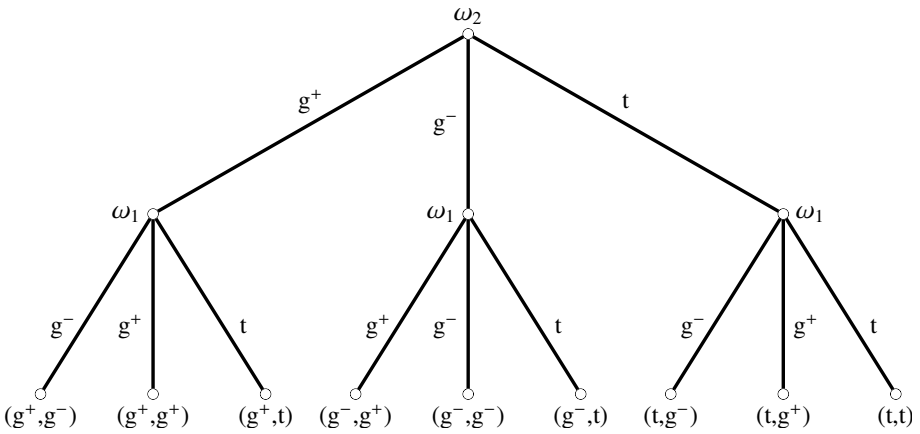
is enforced by a constructed force. Furthermore, we took an integration step of $\Delta t = 1\,\text{fs}$ and a reference temperature of $T = 300\,\text{K}$, and simulated over the same time as with HMC, i.e. 6 ns for pentane and 12 ns for alanine dipeptide. To have a comparable amount of data we wrote out the coordinates every 60 fs.

4.4.1 Pentane

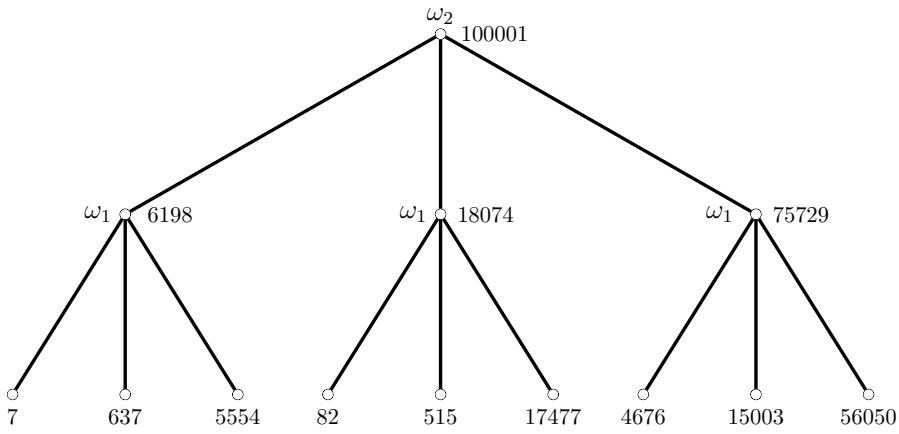
Pentane is able to adopt nine metastable conformations which can be described by the two dihedral angles ω_1 between the carbon atoms C_1, C_2, C_3, C_4 and ω_2 between C_2, C_3, C_4, C_5 (e.g. [56]). Out of a number of 138 descriptors of the pentane molecule (both the distances between any two atoms and the dihedral angles between the bonded carbon atoms), HRDD has selected the two relevant ones, ω_1 and ω_2. The dihedral angles can take the values ±180° (trans), 60° (gauche$^-$) and −60° (gauche$^+$) which we want to abbreviate to t, g$^-$ and g$^+$, respectively. As depicted in Figure 4.3(a), the HRDD algorithm found all nine conformations and the two descriptors ω_1 and ω_2 both for the sampling data generated by HMC and by MD. The chosen tolerances, compare equations (4.4) and (4.5), are listed in Table 4.1. In Figure 4.3(b) the absolute number of states for the respective conformations are given.

4.4.2 Alanine dipeptide

The alanine dipeptide which is a, by acetyl and methyl, terminally-blocked alanine amino acid, has a more complex structure and the metastable conformations are not that well-defined which reflects in the inconsistent results of different articles, see Table 4 in [93] for an overview of some of them. Additionally, the conformational space in vacuum is much smaller than in solution. Thus, we considered 175 possible descriptors of alanine dipeptide, thereof 171 distances and four dihedral angles. For the methylene group at C_β, we chose the united atom presentation since rotation in this side group leads to structurally identical molecules. The algorithm should automatically identify the descriptors which include the most relevant information about conformational transitions. In vacuum, the method found the two dihedral angles ϕ (C-N-Cα-C) and ψ (N-Cα-C-N) describing three metastable conformations, namely C_7^{eq} for $\phi \approx$-80° and $\psi \approx$70°, C_5 for $\phi \approx$-150° and $\psi \approx$155°, and C_7^{ax} for $\phi \approx$80° and $\psi \approx$-50°, see Figure 4.4(a) (for both HMC and MD). In Figure 4.4(b) the corresponding 2-dimensional plot clearly shows the three conformations of alanine dipeptide.



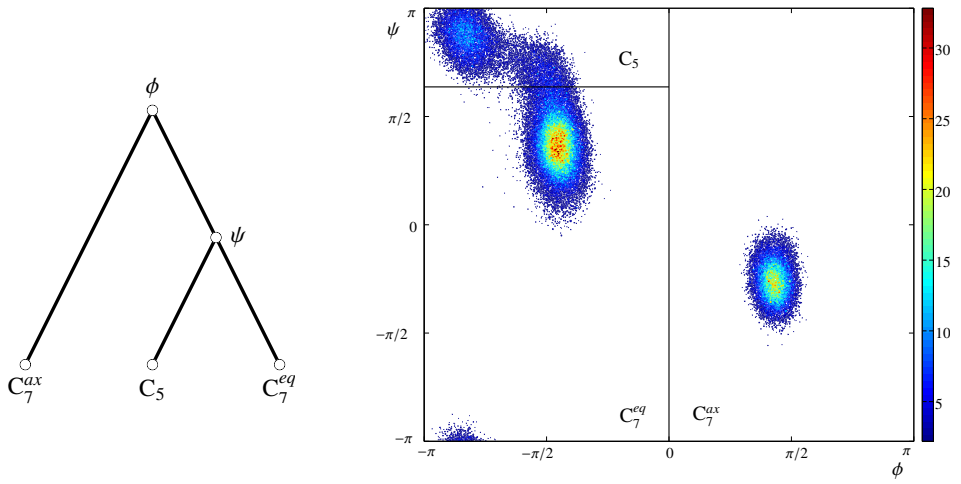
(a) On the leaves the types of conformations are shown.



(b) The numbers on the leaves show the numbers of states according to the respective conformation.

Figure 4.3: The resulting HRDD tree for pentane with nine conformations described by the dihedral angles (ω_2, ω_1).

This Figure 4.4(b) shall be used as well to illustrate the division of the conformational space for the two dihedral angles in case of the MD simulation. The color typifies the absolute number of states visited by the MD simulation with the coordinates rounded to two decimal places. In the first instance the conformational space is divided along the ϕ-axis into two clusters. The first cluster cannot be divided furthermore, this subset matches the conformation C_7^{ax}, while the second cluster can be split along the ψ-axis into two clusters, which are finally the two other conformations C_7^{eq} and C_5.



(a) Rooted tree.

(b) The ϕ-ψ-frequency-plot (the color describes the absolute number of states visited by the MD sampling).

Figure 4.4: The results of the HRDD algorithm for alanine dipeptide: three conformations described by the dihedral angles ϕ and ψ.

4.5 Conclusion

This chapter is geared on two aspects: The first one is the conformational change and the second one is the observation of hierarchies in the descriptors. These two aspects lead us to coarse graining strategies, making the computation of a molecule's dynamics more tractable by requiring less computational power.

We briefly reviewed the detection of metastable sets by MSM, and revealed that the dynamics of a molecule can be explained in terms of transition probabilities between subsets in the conformation space.

The main focus of this chapter is the introduction of the Hierarchical Relevant Descriptor Detector (HRDD) as a method for the identification of the relevant internal degrees of freedom of molecular structures. Whereas principal component analysis (PCA) extracts the coordinate directions with maximal variance of the data, HRDD selects the degrees of freedom which include the "rarest" transitions, therefore, leading to a good separation of conformations. The performance of this method has been tested on two small molecules, namely pentane and alanine dipeptide. In Figure 4.3(a), the relevant dihedral angles of pentane have been identified. For alanine dipeptide, HRDD has selected two descriptors such that the three conformations can clearly be seen in the corresponding two-dimensional plot 4.4(b). For the application of the algorithms on larger molecules, detailed chemical knowledge could be used to reduce the initial amount of possible descriptors (e.g. regarding the atoms of the backbone only) and, thus, improve the performance of the algorithm.

After the successful application of the HRDD method to the small molecules pentane and alanine dipeptide, we also want to critically examine it. The first point concerns the bounds for the eigenvalues (4.4) and (4.5). As already mentioned above, these barriers affect the height and the "width" of the illustrating tree. However, we cannot give a general definition but recommend to adjust the values individually for the underlying simulation. The second point refers to the strictness of the division of the conformational space. Since each time step of the simulation will be assigned to one conformation, the resulting conformations comprise transition states as well and might be, therefore, artificially enlarged. Anyway, the algorithm reveals the correct number of conformations and their most relevant descriptors.

In chapter 6, we want to use this method to find relevant descriptors for the μ-opioid receptor.

Chapter 5

Direct reweighting strategies in conformation dynamics

5.1 Introduction

During the last years, many efforts have been made to find an efficient method to calculate free energy differences [16] for molecular systems defined in terms of classical mechanics and given potential function. However, as yet, there is still no algorithm that is able to calculate the weighting without producing excess data, approximating functions or knowing the shape of the target function in advance.

Considering a molecular system consisting of N atoms, the probability of finding a state

$$x = (q, p) \in (\Omega, \Gamma) \subset \mathbb{R}^{6N}$$

in a canonical ensemble is given by the Boltzmann distribution $\pi(q, p)$ defined in equation (3.1). In section 3.1 we already showed that this Boltzmann equation can be split into π_q and π_p because of the separableness of the total energy $H(q, p) = E_k(p) + E_p(q)$, compare equation (3.2). While π_p is a rather simple quadratic function, π_q is too complex to be computed analytically. In the following, we want to consider free energy differences in order to approximate this term.

The free energy A of a set $B \subset \Omega$ is defined as

$$A = -\frac{1}{\beta} \ln \left(\int_B \exp\Big(-\beta H(q, p)\Big) \mathrm{d}x \right). \tag{5.1}$$

Due to the fact that the potential energy is determinable only except for a constant and, hence, the free energy as well, we are going to use the difference between two sets, where the unknown factor cancels out. Assuming that $\Omega = \bigcup_k B_k$, the free

energy difference between sets B_i and B_j can be computed by

$$\begin{aligned}\Delta A &= -\frac{1}{\beta}\ln\left(\int_{B_j}\exp\Big(-\beta H(q,p)\Big)\mathrm{d}x\right)+\frac{1}{\beta}\ln\left(\int_{B_i}\exp\Big(-\beta H(q,p)\Big)\mathrm{d}x\right)\\ &= -\frac{1}{\beta}\ln\left(\frac{\int_{B_j}\exp\Big(-\beta H(q,p)\Big)\mathrm{d}x}{\int_{B_i}\exp\Big(-\beta H(q,p)\Big)\mathrm{d}x}\right)\overset{(*)}{=}-\frac{1}{\beta}\ln\left(\frac{\int_{B_j}\exp\Big(-\beta E_p(q)\Big)\,\mathrm{d}q}{\int_{B_i}\exp\Big(-\beta E_p(q)\Big)\,\mathrm{d}q}\right)\\ &= -\frac{1}{\beta}\ln\left(\frac{\int_{B_j}\frac{\exp\left(-\beta E_p(q)\right)}{\int_\Omega\exp\left(-\beta E_p(\tilde{q})\right)\mathrm{d}\tilde{q}}\,\mathrm{d}q}{\int_{B_i}\frac{\exp\left(-\beta E_p(q)\right)}{\int_\Omega\exp\left(-\beta E_p(\tilde{q})\right)\mathrm{d}\tilde{q}}\,\mathrm{d}q}\right)=-\frac{1}{\beta}\ln\left(\frac{w_j}{w_i}\right).\end{aligned} \tag{5.2}$$

In the process, we benefit from the fact that the kinetic energy $E_k(p)$ is equal in each set (compare step $(*)$). The weightings w_k which are indicating the probability to be in set B_k,

$$w_k=\int_{B_k}\exp\Big(-\beta E_p(q)\Big)\left[\int_\Omega\exp\Big(-\beta E_p(\tilde{q})\Big)\,\mathrm{d}\tilde{q}\right]^{-1}\mathrm{d}q,$$

are exponential and, hence, rather difficult to calculate. As long-running deterministic dynamical systems are often chaotic and, consequently, do not converge to the wanted distribution (3.1), stochastic approaches were developed.

5.2 New reweighting strategies

In this section we want to present two new reweighting strategies. Both are based on a decomposition of the state space Ω into non-intersecting sets B_k, $k=1,\dots,b$, with $\Omega=\bigcup_{k=1}^{b}B_k$. Let us assume that inside these subsets B_k, separate samplings have generated $\tilde{n}_k$ sampling points $q_l^{(k)}$, $l=1,\dots,\tilde{n}_k$. This can be achieved by various dynamical models, i.e. the Hybrid Monte Carlo method [30], Nosé-Hoover dynamics [41], Berendsen thermostat [7], Smoluchowski dynamics [46], Langevin dynamics [84] or other [42]. A restriction of the sampling to the according set B_k can be achieved by constraining the proposal step to B_k, or by rejecting any proposed state that does not belong to B_k. By doing so, local approximations of the Boltzmann distribution

$$\pi_q^{(k)}=\frac{\exp\Big(-\beta E_p(q)\Big)}{\int_{B_k}\exp\Big(-\beta E_p(\tilde{q})\Big)\,\mathrm{d}\tilde{q}},\quad k=1,\dots,b, \tag{5.3}$$

on the basis of the decomposition are obtained. Based on these separate samplings we want to establish two novel reweighting strategies named Direct Free Energy Reweighting (DFER) and Estimation of Entropy Differences Reweighting (EEDR). By reweighting the local distributions that we obtained from the separate samplings, we intend to obtain a good approximation of the global Boltzmann distribution (3.1).

5.2.1 Direct Free Energy Reweighting (DFER)

In contrast to the methods for the calculation of free energy differences presented in [16] we introduce a direct ansatz that is feasible without producing excess data. We want to start with a description of the two-phase virtual sampling scheme of this approach.

Jump phase. Starting in state $q_i^{(k)} \in B_k$, a proposal step is made by selecting a target set B_l at random with uniform distribution in $l \in \{1,\dots,b\}$ and by choosing a state $q_j^{(l)} \in B_l$ out of the $\tilde{n}_l$ sampling points at random with uniform distribution in $j \in \{1,\dots,\tilde{n}_l\}$. The proposed state is accepted with probability

$$P_{ij}^{(a)} = \min\left\{1, \frac{\exp\left(-\beta E_p\left(q_i^{(k)}\right)\right)}{\exp\left(-\beta E_p\left(q_j^{(l)}\right)\right)}\right\} \tag{5.4}$$

otherwise the previous sampling point $q_i^{(k)} \in B_k$ is retained. Note that the fraction in the acceptance probability is inverted with regard to the standard Metropolis. This leads to the *inverse balance condition*, to which we will come back later on.

Evaluation phase. The second phase of the virtual sampling scheme consists of the selection of evaluation regions $\mathcal{E}_k \subset B_k$ of equal volume. A small subset of each set B_k is chosen where the sampling of the jump phase is rather dense, e.g. near a local minimum of the potential energy function E_p, or within a region of average potential energy. Finally, the statistical weight of set B_k can be estimated by the inverse of the number of states n_k in set $\mathcal{E}_k$, i.e. $w_k \propto n_k^{-1}$.

A graphic presentation is outlined in Figure 5.1.

In the following, we want to show that the virtual sampling scheme in fact leads to an estimation of free energy differences between the sets B_k. The conditional transition probability P_{ij} from state $q_i^{(k)} \in B_k$ to state $q_j^{(l)} \in B_l$ is defined as a product of the proposition probability $P_{ij}^{(p)}$ and the acceptance probability $P_{ij}^{(a)}$. As the separate samplings correspond to local Boltzmann distributions (5.3) the proposal transition probability yields

$$\frac{P_{ij}^{(p)}}{P_{ji}^{(p)}} = \frac{\pi_q^{(l)}\left(q_j^{(l)}\right)}{\pi_q^{(k)}\left(q_i^{(k)}\right)} = \frac{\exp\left(-\beta E_p\left(q_j^{(l)}\right)\right)}{\exp\left(-\beta E_p\left(q_i^{(k)}\right)\right)} \cdot \frac{w_k^{(i)}}{w_l^{(j)}}.$$

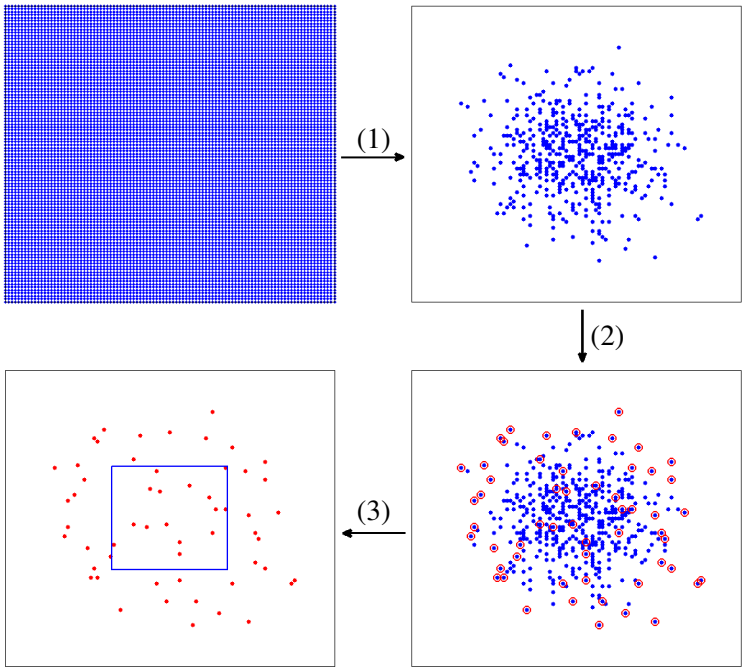


Figure 5.1: Virtual scheme of DFER: Starting in subset B_i, (1) a separate sampling is run, (2) in a jump-sampling the red points are chosen, and (3) an evaluation region is selected. The number of points inside the evaluation region is inversely proportional to the statistical weight of B_i.

The acceptance transition probability is constructed analogously to equation (5.4) of the jump phase

$$\frac{P_{ij}^{(a)}}{P_{ji}^{(a)}} = \frac{\exp\left(-\beta E_p\left(q_i^{(k)}\right)\right)}{\exp\left(-\beta E_p\left(q_j^{(l)}\right)\right)}.$$

Hence, the overall transition probability matrix P results in

$$\frac{P_{ij}}{P_{ji}} = \frac{w_k^{(i)}}{w_l^{(j)}}. \tag{5.5}$$

As this is the reciprocal of the standard balance condition we call it the *inverse balance condition*.

A sampling according to this transition probability matrix is generating a sequence of sampling points whose distribution converges to the dominant left eigenvector γ of P to the eigenvalue 1, see [20], i.e.

$$\gamma^T P = \gamma^T.$$

It is easy to check that P is irreducible and positive, which is necessary for the uniqueness of γ because of the theorem of Frobenius and Perron [51]. By using equation (5.5) the eigenvalue problem can be written as

$$\sum_i \gamma_i \frac{w_k^{(i)}}{w_l^{(j)}} P_{ji} = \gamma_j \quad \forall j .$$

Because of the row-stochastic characteristic of P, i.e. $\sum_i P_{ji} = 1 \; \forall j$, we obtain

$$\sum_i \left(\gamma_i \frac{w_k^{(i)}}{w_l^{(j)}} - \gamma_j \right) P_{ji} = 0 \quad \forall j \text{ and } \forall P_{ji} \in [0,1] .$$

The left eigenvector to the eigenvalue 1 of P results in $\gamma_i \propto \left(w_k^{(i)}\right)^{-1}$. Due to this relation and the equal volume of the evaluation regions, the quantity n_k of all sampling points of the jump phase that are elements of the evaluation region $\mathcal{E}_k$ is proportional to w_k^{-1}.

As trapping is the main reason for an ill-conditioned computation of the global distribution of states [114] we want to offer a modification of the above virtual sampling scheme that lessens trapping effects. To this end, we make use of the disaggregation method based on the ideas of Simon and Ando [91]. As points outside the evaluation regions typically have higher potential energy values than those inside, samplings in the jump phase may be trapped outside the evaluation regions. This applies especially when outlier states with very high potential energy values occur. A trapped sampling may yield only a very small number of states in the evaluation regions, which consequently would lead to a broken statistics. The limit distribution of the jump sampling is given by the dominant left eigenvector γ of a stochastic matrix P. As we are only interested in this part of γ that corresponds to the evaluation regions, we intend to compute the required components of the vector only. This is realized by rejecting states that are outside of the evaluation regions during the jump phase. In this context, rejection means repetition of the current state in the Markov chain if the proposed state is not part of the evaluation region. Assuming that P is the original transition probability matrix, the disaggregated transition probability matrix $\tilde{P}$ looks as follows:

$$\tilde{P}_{ij} = \begin{cases} P_{ij}, & \text{if } i \neq j \\ P_{ii} + \sum\limits_{\iota : q_\iota^{(k)} \notin \mathcal{E}_k} P_{i\iota}, & \text{if } i = j \end{cases}$$

With the identity $\gamma_i P_{ij} = \gamma_j P_{ji}$ one can easily show that $\tilde{P}$ has the same left eigenvector as P, but is restricted to the components that correspond to the evaluation regions.

In practice, the algorithm is much shorter because we can compute the statistical weights w_k directly and, hence, do not need the transition probability matrix

P. Therefore, the following three steps are necessary for the computation of the statistical weights.

(1) Define an evaluation region $\mathcal{E}_k \subset B_k$ where the sampling is rather dense, e.g. near the minimum, the mean value or the median of the potential energy E_p in the state space B_k. Keep in mind that all evaluation regions have to be of the same size.

(2) Compute the statistical weights as

$$w_k \propto \left(\frac{1}{\tilde{n}_k} \sum_{q_i^{(k)} \in \mathcal{E}_k} \exp\left(\beta E_p\left(q_i^{(k)}\right)\right) \right)^{-1} . \tag{5.6}$$

(3) Finally, normalize the statistical weights to

$$\sum_{k=1}^{b} w_k = 1 .$$

In order to avoid the aforementioned trapping problems, we also want to offer a modification for the direct computation of the statistical weights. Typically, the energy levels of the subsets B_k are varying, i.e. the mean potential energy values $E_p^{(k)} = \langle E_p^{(k)} \rangle$ differ. As a consequence, sets B_k with high potential energy levels can lead to a trapping of the sampling, compare equation (5.4). As we are aiming at an equilibrated sampling among the different sets, we shift the potential energy functions according to their mean values, i.e.

$$\tilde{E}_p\left(q_i^{(k)}\right) = E_p\left(q_i^{(k)}\right) - E_p^{(k)} \quad \text{with} \quad q_i^{(k)} \in B_k .$$

Therefore, we have to rewrite expression (5.6) as

$$w_k \propto \left(\frac{1}{\tilde{n}_k} \sum_{q_i^{(k)} \in \mathcal{E}_k} \exp\left(\beta \tilde{E}_p\left(q_i^{(k)}\right)\right) \right)^{-1} \cdot \exp\left(-\beta E_p^{(k)}\right) .$$

That way, a set B_k with high mean potential energy $E_p^{(k)}$ is assigned a reduced statistical weight w_k.

5.2.2 Estimation of Entropy Differences Reweighting (EEDR)

A simple method for the estimation of entropy differences has been developed by Weber and Andrae [112]. They applied their method to differentiate between varying polymer structures from thermostated molecular dynamics simulations. We want to modify this approach in order to use it for the calculation of statistical weights.

The entropy of a system is a measure of the distribution's "disorder" or multiplicity. In the context of molecular conformation space, a low entropy system

is rather uniform, i.e. sampled conformations look similar, whereas in a high entropy system, sampled conformations are of high variety. In the following, we are interested in the entropy difference

$$\Delta S = \frac{\Delta U - \Delta A}{T}, \tag{5.7}$$

which is defined in terms of the thermodynamic state functions, internal energy U and free energy A. The internal energy difference between two systems with potential energy $E_p^{(i)} : B_i \to \mathbb{R}$ and $E_p^{(j)} : B_j \to \mathbb{R}$ is

$$\begin{aligned}\Delta U &= \frac{\int_{B_j} E_p^{(j)}(q) \exp\left(-\beta E_p^{(j)}(q)\right) \mathrm{d}q}{\int_{B_j} \exp\left(-\beta E_p^{(j)}(q)\right) \mathrm{d}q} - \frac{\int_{B_i} E_p^{(i)}(q) \exp\left(-\beta E_p^{(i)}(q)\right) \mathrm{d}q}{\int_{B_i} \exp\left(-\beta E_p^{(i)}(q)\right) \mathrm{d}q} \\ &= \langle E_p^{(j)} \rangle - \langle E_p^{(i)} \rangle . \end{aligned} \tag{5.8}$$

In practice, the value of the internal energy difference can be approximated by the mean potential energy value over a Boltzmann distributed sampling of position states.

The free energy A and its difference ΔA have already been defined in equations (5.1) and (5.2). In order to get a numerical approximation of the free energy difference we want to use a special Monte Carlo quadrature approach of an integrand $f : \Omega \to \mathbb{R}^+$, see [112],

$$I = \int_\Omega f(x)\, \mathrm{d}x \approx f(\tilde{x})\, \mathrm{vol}(\Omega) \frac{\mathcal{U}_1(\tilde{x})}{\mathcal{U}_\rho(\tilde{x})},$$

where $\mathcal{U}_1(\tilde{x})$ is the number of equally distributed states in an environment of $\tilde{x}$ and $\mathcal{U}_\rho(\tilde{x})$ is the quantity of ρ-distributed states with $\rho \propto f$. Hence, the free energy difference can be written as

$$\begin{aligned}\Delta A &= -\frac{1}{\beta} \ln\left(\frac{\int_{B_j} \exp\left(-\beta E_p^{(j)}(q)\right) \mathrm{d}q}{\int_{B_i} \exp\left(-\beta E_p^{(i)}(q)\right) \mathrm{d}q} \right) \\ &\approx -\frac{1}{\beta} \ln\left(\frac{\exp\left(-\beta E_p^{(j)}\left(q^{(j)}\right)\right) \mathrm{vol}\left(B_j\right) \dfrac{\mathcal{U}_1^{(j)}\left(q^{(j)}\right)}{\mathcal{U}_{\pi_q}^{(j)}\left(q^{(j)}\right)}}{\exp\left(-\beta E_p^{(i)}\left(q^{(i)}\right)\right) \mathrm{vol}\left(B_i\right) \dfrac{\mathcal{U}_1^{(i)}\left(q^{(i)}\right)}{\mathcal{U}_{\pi_q}^{(i)}\left(q^{(i)}\right)}} \right). \end{aligned}$$

Requiring a comparable structure of the position spaces B_i and B_j, which can be expressed by $\mathrm{vol}(B_i)\mathcal{U}_1^{(i)}\left(q^{(i)}\right) = \mathrm{vol}\left(B_j\right)\mathcal{U}_1^{(j)}\left(q^{(j)}\right)$, the estimation of the free energy difference simplifies to

$$\Delta A \approx -\frac{1}{\beta}\ln\left(\frac{\mathcal{U}_{\pi_q}^{(i)}\left(q^{(i)}\right)}{\mathcal{U}_{\pi_q}^{(j)}\left(q^{(j)}\right)}\right) + E_p^{(j)}\left(q^{(j)}\right) - E_p^{(i)}\left(q^{(i)}\right). \tag{5.9}$$

Using equations (5.8) and (5.9), the entropy difference can be written as

$$\Delta S \approx \frac{\left[\langle E_p^{(j)}\rangle - E_p^{(j)}\left(q^{(j)}\right)\right] - \left[\langle E_p^{(i)}\rangle - E_p^{(i)}\left(q^{(i)}\right)\right]}{T} + k_B \ln\left(\frac{\mathcal{U}_{\pi_q}^{(i)}\left(q^{(i)}\right)}{\mathcal{U}_{\pi_q}^{(j)}\left(q^{(j)}\right)}\right).$$

Now, considering representatives $q^{(i)}$, $q^{(j)}$ with mean potential energy value, that is $E_p^{(i)}\left(q^{(i)}\right) = \langle E_p^{(i)}\rangle$ and $E_p^{(j)}\left(q^{(j)}\right) = \langle E_p^{(j)}\rangle$, respectively, the approximation of the entropy difference can be reduced to

$$\Delta S \approx k_B \ln\left(\frac{\mathcal{U}_{\pi_q}^{(i)}\left(q^{(i)}\right)}{\mathcal{U}_{\pi_q}^{(j)}\left(q^{(j)}\right)}\right).$$

As a single state might not provide a good approximation, we take the mean of several representatives with approximately mean potential energy value. Assuming an equal number of representatives in each set, $R_i = R_j = R$, the final approximation yields

$$\Delta S \approx k_B \ln\left(\frac{\sum_{l=1}^{R}\left(\mathcal{U}_{\pi_q}^{(j)}\left(q_l^{(j)}\right)\right)^{-1}}{\sum_{l=1}^{R}\left(\mathcal{U}_{\pi_q}^{(i)}\left(q_l^{(i)}\right)\right)^{-1}}\right).$$

Alternatively, instead of the arithmetic mean, the median can be used. Being able to estimate the entropy difference between two position spaces B_i und B_j, we can apply relation (5.7) to determine the free energy difference ΔA and, using equation (5.2), the statistical weights w_i and w_j.

For a better comprehension of the algorithm, we want to outline it shortly.

(1) In subset B_i of position space Ω, choose R reference points $q_l^{(i)}$, $l = 1,\ldots,R$, with approximately mean potential energy value.

(2) Using some distance measure, count the number $n_p^{(i)}$ of sampling points that are *near*, i.e. within a certain distance around each reference point $q_l^{(i)}$, and compute its inverse

$$\left(\mathcal{U}_{\pi_q}^{(i)}\left(q_l^{(i)}\right)\right)^{-1} \approx \frac{1}{n_p^{(i)}/\tilde{n}_i + 1} =: \frac{1}{n_l^{(i)} + 1}$$

with $\tilde{n}_i$ being the total number of sampling points in subset B_i.

(3) Compute the entropy

$$S_i = k_B \ln\left(\frac{1}{R}\sum_{l=1}^{R}\left(\mathcal{U}_{\pi_q}^{(i)}\left(q_l^{(i)}\right)\right)^{-1}\right),$$

the free energy $A_i = \langle E_p^{(i)}\rangle - T\,S_i$

and the statistical weights

$$w_i = w_{i-1}\cdot\exp\left(-\beta\left(A_i - A_{i-1}\right)\right) \tag{5.10}$$

with $w_1 = 1$. If necessary, the free energy values have to be ordered by size before calculating the statistical weights.

(4) Finally, the statistical weights have to be normalized to

$$\sum_{i=1}^{b} w_i = 1\,.$$

5.3 Application

After having derived and explained new strategies for calculating statistical weights on the basis of separate samplings, we want to prove the performance of the novel methods with the help of two examples.

5.3.1 An abstract example: a tetrahedral molecule

The first example is meant to be a simple model, as it is rather abstract and not to be found in practice. Assume a molecule consisting of four atoms is arranged like a regular tetrahedron, see Figure 5.2. Vertices $Q_i = \left(Q_{ix}, Q_{iy}, Q_{iz}\right)^T$, $i = 1,\ldots,4$, represent the atoms, and the edges $r_{ij} = \left\|Q_i - Q_j\right\|$, $i \neq j$, form the bonds. Using generalized coordinates $q = \left(Q_1^T, Q_2^T, Q_3^T, Q_4^T\right)^T \in \Omega \subset \mathbb{R}^{12}$ and, analogously, generalized momenta $p \in \mathbb{R}^{12}$, we can define the energy functions

$$E_k(p) = \frac{1}{2}p^T M^{-1} p$$

$$E_p(q) = \sum_{i=1}^{3}\sum_{j=i+1}^{4} k\left(r_{ij}(q) - r\right)^2$$

with force constant k and bond length r. Note that this is an extremely simplified potential energy function, which in practice would be a sum of different energy terms describing the various intra- and intermolecular forces acting on the molecule.

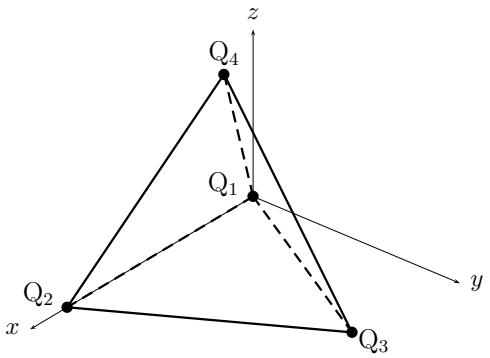


Figure 5.2: This regular tetrahedron serves as a model for a simple example to prove the performance of the new reweighting strategies.

subset	exact	DFER	EEDR
B_1	0	0.00007	0
B_2	0.5	0.50001	0.45025
B_3	0	0	0
B_4	0	0	0
B_5	0	0	0
B_6	0	0	0
B_7	0.5	0.49987	0.54958
B_8	0	0.00005	0.00017

Table 5.1: Comparison of the results of the statistical weighting of the conformations of the tetrahedral molecule model: exact vs. Direct Free Energy Reweighting (DFER) vs. Estimation of Entropy Differences Reweighting (EEDR).

Clearly, this artificial potential energy function is minimal for $r_{ij} = r \; \forall \, i \neq j$. Hence, two metastable conformations can be detected which only differ in the orientation of the apex, Q_4, of the tetrahedron. During a molecular simulation, the tetrahedron may not be able to change into another conformation because of the high energy barriers. Therefore, we divided state space Ω into eight subsets B_k, $k = 1, \ldots, 8$, ran the separate samplings and applied the reweighting strategies. The results are listed in Table 5.1.

Obviously the Direct Free Energy Reweighting induces a better result than the Estimation of Entropy Differences Reweighting. But we have to remind that the tetrahedral molecule is an abstract example and that we used a very simple function of the potential energy which is not very realistic.

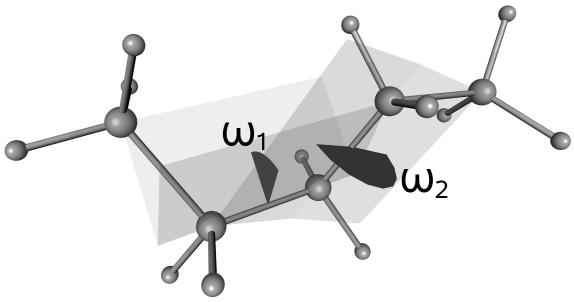


Figure 5.3: The torsion angles ω_1 between carbon atoms C_1, C_2, C_3, C_4 and ω_2 between the carbon atoms C_2, C_3, C_4, C_5 define the metastable conformations of the alkane pentane.

5.3.2 Pentane

To provide a more realistic example, we want to derive the statistical weights of the alkane pentane. As the conformational distribution of pentane at a temperature of $T = 300\,\mathrm{K}$ has been determined in a robust manner before, see [87], we decided to use this molecule for evaluation purposes. Pentane is able to adopt nine metastable conformations, defined by the two torsion angles ω_1 and ω_2 between the carbon atoms C_1, C_2, C_3, C_4 and C_2, C_3, C_4, C_5, respectively, see Figure 5.3. At a torsion angle of ±180° (trans) for both ω_1 and ω_2, the geometry is energetically most favorable. Furthermore, local minima of the torsion potential can be found at torsion angle values of 60° (gauche$^-$) and -60° (gauche$^+$). We will use the same abbreviations as in section 4.4, namely trans=t, gauche$^-$=g$^-$ and gauche$^+$=g$^+$. Because of the conformational symmetry of pentane, one would expect to find approximately equal weights for corresponding conformations, e.g. $w(\mathrm{t/g^-}) \approx w(\mathrm{g^-/t})$.

To obtain a dataset for reweighting, we first parameterized pentane according to the Merck molecular force field (MMFF) [37], a force field optimized for small, drug-like molecules. The pentane molecule was modeled in vacuo at a temperature of $T = 300\,\mathrm{K}$. Nine geometries corresponding to the conformations listed in Table 5.2 were generated using the molecule editor from the visualization software Amira [85], subsequently serving as starting points for nine separate Hybrid Monte Carlo (HMC) [30] samplings. HMC was set up to sample five chains with 1000 HMC steps each per starting geometry. For each HMC step, a 60 fs molecular dynamics trajectory was calculated to generate a trial state. The average acceptance probability of the HMC runs was more than 90%. The convergence was monitored using the Gelman-Rubin acceptance criterion [33] with a threshold value of 1.2. The potential function was modified so that the sampling was restrained to the current local minimum, assuring a thorough sampling of even the energetically unfavorable conformations.

conformation	"exact"	DFER	EEDR
t/t	0.473	0.464	0.440
g^-/t	0.120	0.100	0.103
g^+/t	0.132	0.127	0.130
t/g^-	0.117	0.102	0.103
t/g^+	0.132	0.131	0.134
g^+/g^+	0.013	0.040	0.044
g^-/g^-	0.012	0.034	0.044
g^+/g^-	<0.005	0.001	0.001
g^-/g^+	<0.005	0.001	0.001

Table 5.2: Comparison of the results of the statistical weighting of the conformations of pentane at a temperature of 300 K: "exact" [87] vs. Direct Free Energy Reweighting (DFER) vs. Estimation of Entropy Differences (EEDR).

The results of the two novel reweighting strategies are presented in Table 5.2, together with the "exact" values from [87]. For this example, the alkane pentane, it is not easy to determine the better strategy. Both reweighting methods approximate the "exact" weights rather well. When comparing the weights with the Euclidean and the infinity norm, however, the Direct Free Energy Reweighting provides a marginally better result.

5.4 Conclusion

After having described two novel direct reweighting strategies for the use in conformation dynamics in section 5.2, we could prove their applicability in section 5.3. Both algorithms provide a good approximation of statistical weights in a direct way. While the Direct Free Energy Reweighting (DFER) determines the weights only on the basis of the potential energy values of the separate samplings, and, furthermore, evaluates only one equation, namely (5.6), the Estimation of Entropy Differences Reweighting (EEDR) requires additional data to compute distances between conformations (e.g. dihedral angle values or intramolecular distances) and handles more than one equation. Therefore, the DFER is faster and requires less memory. Additionally, the DFER provided marginally better results in our two small examples.

In both algorithms we mentioned certain evaluation regions and representatives, respectively. The values for the size of the evaluation region and for the quantity of representative points have to be chosen individually, adapted for the present case, so that the optimal result can be reached. In the process, one has to regard that both, too large or too small evaluation regions, and too few or too many representative points, respectively, may lead to a bad statistics.

In contrast to existing methods [16], we do not have to generate overlapping distributions in order to weight them against each other. Our novel direct strategies offer a fast and efficient possibility for the computation of statistical weights without producing excess data or approximating functions.

In the next chapter, we want to apply the reweighting strategies to parallel running trajectories for the exploration of the conformation space of the μ-opioid receptor.

Chapter 6

The μ-opioid receptor, analysis

6.1 Introduction

After having presented the theory of molecular simulation in chapter 3, a coarse graining method for the dimension reduction of the state space in chapter 4, and direct reweighting strategies in conformation dynamics in chapter 5, we want to analyze the μ-opioid receptor (MOR) using these techniques in this chapter, and check two of the four hypotheses proposed in section 2.4 concerning the reasons for the better behavior of the μ-opioid receptor in inflamed tissue, namely

(1) The differing Coulomb interactions enable the analgesic to find its way into the binding pocket of the MOR better in the inflamed tissue.

(2) The conformation of the MOR in the inflamed tissue is more suitable for the analgesic than in healthy tissue.

With regard to the first hypothesis, we want to investigate the electrostatic potential of the receptor and run short simulations with a small molecule near to the binding pocket of the MOR to analyze the way of binding, see section 6.2. For the study of the second hypothesis in section 6.3, a whole conformation dynamics analysis has to be performed, starting from a meshless discretization of the state space, running trajectories in parallel, and evaluate these trajectories with direct reweighting strategies. This is a very difficult task since it is not easy to generate different starting configurations for parallel running trajectories for proteins like the μ-opioid receptor. Anyhow, we want to emphasize, that we do not expect to find only one reason for the better behavior of the MOR in inflamed tissue, but that there is an interaction of different factors. As already mentioned in chapter 2.4, both, the intracellular pH value and the chemical properties of the ligand seem to have an influence on the better effect of the analgesics in inflamed tissue.

During the study of the μ-opioid receptor, many different molecular simulations have been run. As already described in chapter 2, we modeled the μ-opioid

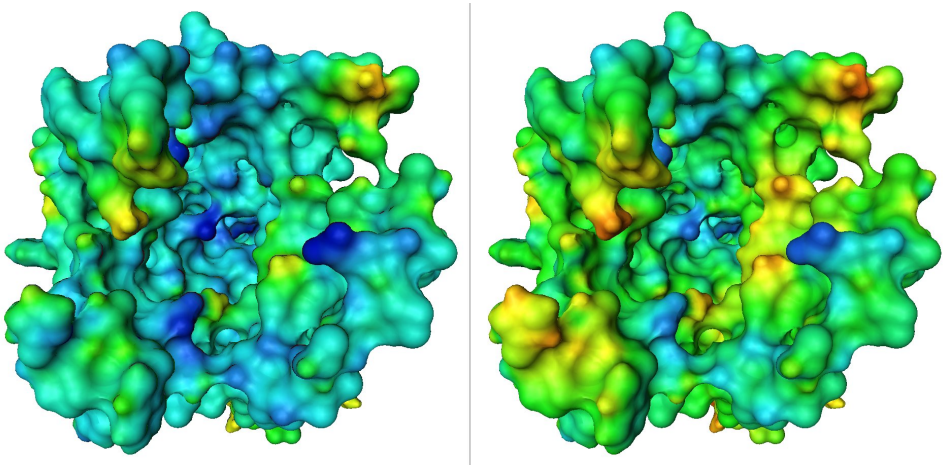


Figure 6.1: Electrostatic potential for the deprotonated (left) and the protonated (right) μ-opioid receptor, generated and visualized with Amira. View into the binding pocket. The colormap ranges from blue (low/negative charge) to red (high/positive charge).

receptor in a simulation box with the lipid bilayer DPPC and the spc water model, altogether more than 31000 atoms. For the molecular modeling and simulation, we chose the software package GROMACS with the force field GROMOS 53A6 combined with the Berger Lipids. Some of the simulations, especially the parallel running trajectories over several nanoseconds, were run on the HLRN (North-German Supercomputing Alliance) as they are computationally too intensive for normal computers.

6.2 Way of binding

In this section we want to analyze the first hypothesis, i.e. the Coulomb interactions between the μ-opioid receptor and the ligand, the positively charged morphine. In a first step, we use the visualization software Amira [85] to approximate and visualize the electrostatic potential. Electrostatic potential maps, also known as molecular electrical potential surfaces, illustrate the distribution of charge of a molecule and enable us to analyze the interaction of molecules. With the help of Open Babel [72] to convert the initially generated molecular model into the correct file format, and MATLAB [68] to add the correct charges out of the GROMACS topology file created in the beginning of the modeling phase, we generated and visualized the electrostatic potential with Amira for both protonation states (deprotonated in healthy tissue, protonated in inflamed tissue). In Figure 6.1, the electrostatic potential maps for the deprotonated and the protonated μ-opioid receptor are depicted with a view into the binding pocket. Thereby, the color refers to the charge distribution on the molecular surface, from low/negative (blue) to high/positive (red). In

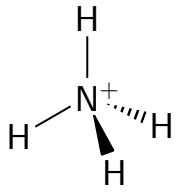
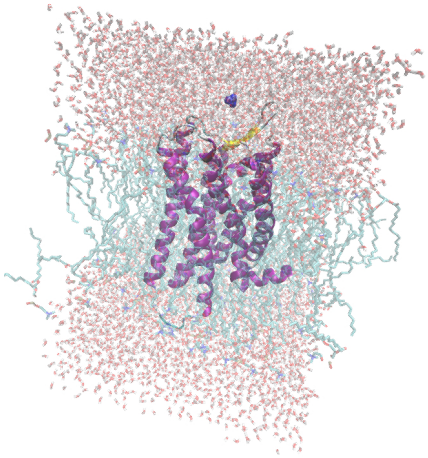


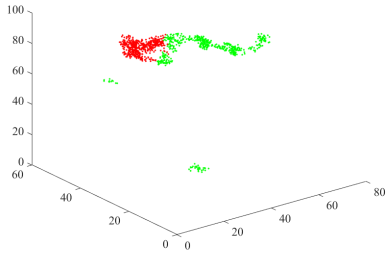
Figure 6.2: Chemical structure of ammonium NH_4^+.

the case of the deprotonation, the negative charge (blue) is widely spread, i.e. the entropy is very high, so that the (positively charged) morphine would have many possibilities to interact (possibly allosteric binding). In contrast, in the case of the protonated receptor there are only a few negatively charged amino acids, especially nearby and in the binding pocket, so that the (positively charged) morphine probably finds the way into the binding pocket faster. This does not imply that the binding of morphine is stronger or more lasting, which still has to be verified by different analysis techniques, but these results are still strengthening the assumption that the pH value is a big influencing factor in the investigation of the different behavior of morphine in healthy and inflamed tissue.

In the next step, we want to investigate the Coulomb interactions by setting up a simulation with a suitable ligand, computing the entropy of this ligand and calculating the Coulomb energies between ligand and receptor. Since we do not want to study the interactions in detail, and for the sake of convenience, instead of the complex molecule morphine, we chose a small positively charged molecule for further investigations: ammonium NH_4^+. Using MarvinSketch [1], the molecule NH_4 was drawn and exported as mol2-file. This mol2-file was translated in a pdb-file using Open Babel. With the help of ACPYPE [21] the neutral NH_4 was positively charged and parametrized. In Figure 6.2, the graphic formula of ammonium is shown. In the simulation box modeled in section 2.2 with the μ-opioid receptor embedded in the DPPC lipid bilayer and solvated with water, we replaced one of the water molecules by the ammonium (knowing that the ammonium has two more atoms than a water molecule). For this, we visualized the simulation box with Amira and chose a suitable water molecule above the binding pocket that we replaced by the ammonium. In Figure 6.3(a), the ammonium is drawn in purple within the simulation box of the μ-opioid receptor. Since the ammonium is positively charged, we had to add one negative counterion to the system for neutrality. Putting high position restraints on the ammonium in order to prevent a movement away from the current position, first a NVT equilibration and then a NPT equilibration was run. Thereby, we had to choose very small time steps (0.02 fs instead of 2 fs) so that the ammonium did not disturb the surrounding water molecules. After the equilibration, we could run the molecular dynamics simulation. We chose a time step of $\Delta t = 2\,\text{fs}$, and a number of total steps of $n_{steps} = 500000$ which corresponds to a total simulation time of 1 ns. The first molecular dynamics simulation



(a) Ammonium (purple) in the simulation box of MOR, placed above the binding pocket. Visualization with VMD.



(b) Coordinates of the nitrogen of NH_4^+ in a MD simulation (green: deprotonated, red: protonated), visualized with MATLAB.

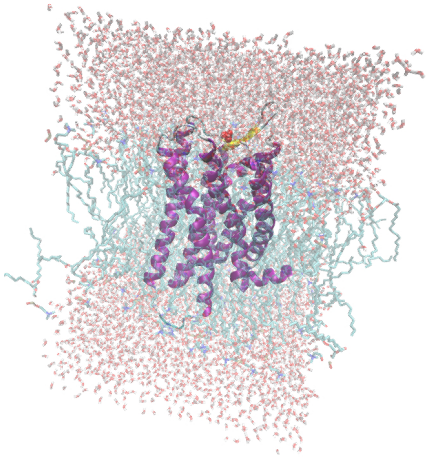
Figure 6.3: A simulation with the small positively charged molecule ammonium NH_4^+ shall help to investigate the Coulomb interactions between receptor and ligand. In healthy tissue (deprotonated system), the ammonium drifts away from the starting position; in inflamed tissue (protonated system), the ligand moves only in a small radius around the binding pocket.

was very promising: in the protonated system, the ammonium was moving in a small radius around its starting position, while in the deprotonated system, the ammonium was drifting far away from the starting point. To visualize this result, see Figure 6.3(b), we used MATLAB to draw the position coordinates of the nitrogen of NH_4^+ that have been written out during the MD simulation every 1000 steps. For statistical reasons, we started further simulations with different starting positions for the ammonium, see Figure 6.4(a), but in contrast to the first simulation, the ammonium in the protonated system was moving away from the binding pocket, while in the deprotonated system the ammonium stayed in a small radius to the starting point, see Figure 6.4(b).

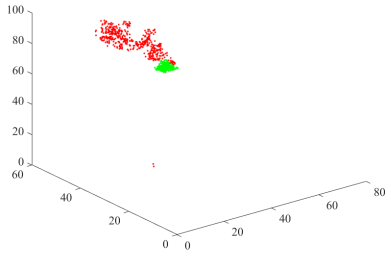
To confirm this visual result, we calculated the entropy S by discretizing the simulation box into $10 \times 10 \times 10 = 1000 = b$ boxes and counting the number n_i of sampling points (of the nitrogen of the ammonium) in each box

$$S = -\sum_{i=1}^{b} P_i \ln(P_i) \tag{6.1}$$

with $P_i = n_i/n$ and a total number of sampling points n. In the protonated system, the computed entropy of the ammonium is higher than in the deprotonated system,



(a) Ammonium (red) in the simulation box of MOR, placed above the binding pocket. Visualization with VMD.



(b) Coordinates of the nitrogen of NH_4^+ in a MD simulation (green: deprotonated, red: protonated), visualized with MATLAB.

Figure 6.4: Another simulation with the small positively charged molecule ammonium NH_4^+ to investigate the Coulomb interactions between receptor and ligand. In this case, the ligand moves away in the protonated system (inflamed tissue) while it stays nearby the binding pocket in the deprotonated system (healthy tissue).

except for the first MD simulation, compare Table 6.1. This emphasizes the assumption that there is more disorder in the protonated system, that the ammonium is moving much more.

For the analysis of the Coulomb interactions between the MOR and ammonium, we extracted the Coulomb energies from the MD simulations and visualized them with MATLAB, see Figure 6.5. Negative Coulomb energies mean attractive forces between the molecules, positive Coulomb energies stand for repulsive forces. It is rather obvious that the ammonium has almost no electrostatic interaction with the μ-opioid receptor in the protonated system. Most of the time, the Coulomb energy values for the protonated receptor are on the base line, only if

Table 6.1: The entropy of the ammonium in four different simulations over 1 ns in the MOR system.

	deprotonated	protonated
1	2.7262	2.0164
2	0.7224	2.6373
3	2.5131	2.8924
4	0.9378	3.0067

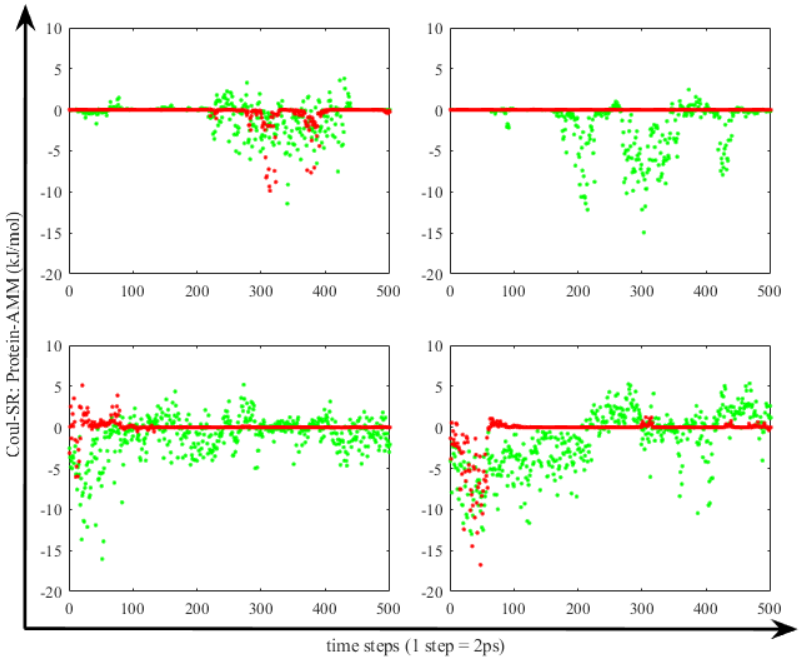


Figure 6.5: Coulomb energies between MOR and NH_4^+ for four MD simulations with different starting points (green: deprotonated, red: protonated), visualized with MATLAB. Negative Coulomb energies mean attractive forces, positive Coulomb energies stand for repulsive forces.

started close to the binding pocket (the two lower plots), attractive and repulsive forces can be observed. In contrast to this, in the deprotonated system the ammonium is permanently interacting with the protein. However, this result is not contradicting to chemical properties since, in the protonated system, the μ-opioid receptor is positively charged, implying a repulsion of positively charged molecules. After having falsified the first hypothesis, we want to investigate the second possible reason for the better effect of morphine in inflamed tissue, the conformations of the μ-opioid receptor. To analyze this assumption, we have to run parallel trajectories based on different starting configurations for both the protonated and the deprotonated system to, hopefully, explore the whole conformational space.

6.3 Analysis of conformations

As already described in chapter 3, the best way to sample the conformational space of a complex molecule is to decompose the state space and run separate samplings. In chapter 3.3 we discussed the difficulties to decompose the state space without knowing it in advance. The well-established presampling at high temperature fails

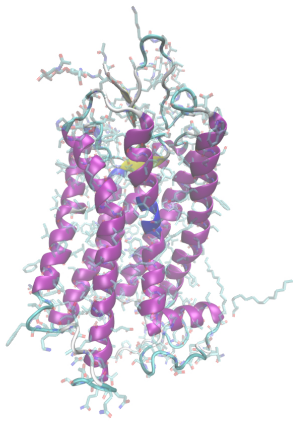


Figure 6.6: VMD snapshot of the receptor rhodopsin, created on the basis of the crystal structure found by Okada et al. in 2004 [73].

for proteins, since proteins are denaturing at high temperature and the system drifts apart. By simulating at low temperature and starting in only one conformation, the trajectory will probably not overcome energy barriers to find other conformations, even not over a long time.

Therefore, cluster data from rhodopsin shall be used to generate suitable starting structures. Like the μ-opioid receptor, rhodopsin belongs to the family of rhodopsin-like G-protein-coupled receptors and possesses seven transmembrane helices, see Figure 6.6. The researchers Pérez-Hernández and Noé from Freie Universität Berlin are working on the dynamics of rhodopsin [76] and gained a huge amount of simulation data. Thereby, approximately 8000 clusters and the belonging centers, configurations with lowest potential energy inside the clusters, could be identified. We transferred the cluster data to the μ-opioid receptor by using a homology modeling based on a multiple sequence alignment. The in this way generated configurations can be used as starting points for the parallel running MD trajectories to explore the conformational space.

Since we do not want to use all cluster centers of rhodopsin, we applied the in chapter 4.3 presented method *Hierarchical Relevant Descriptor Detector* (HRDD) to find the most relevant clusters and their descriptors which we can use to generate starting configurations for the μ-opioid receptor. The main idea of this method is coarse graining in space, i.e. describing relevant motions (conformational changes) by using only a few degrees of freedom (the most relevant descriptors). Possible descriptors can be dihedral angles, torsion angles or bond lengths, for instance. Based on a (long-term) trajectory or another sequence of sampling points, the conformational space represented by the sampled data is discretized uniformly into intervals (bins) for all given descriptors. For each descriptor the transition matrix is calculated and the second largest eigenvalues are computed. The largest of all second largest eigenvalues, necessarily close to 1, characterizes the descriptor with

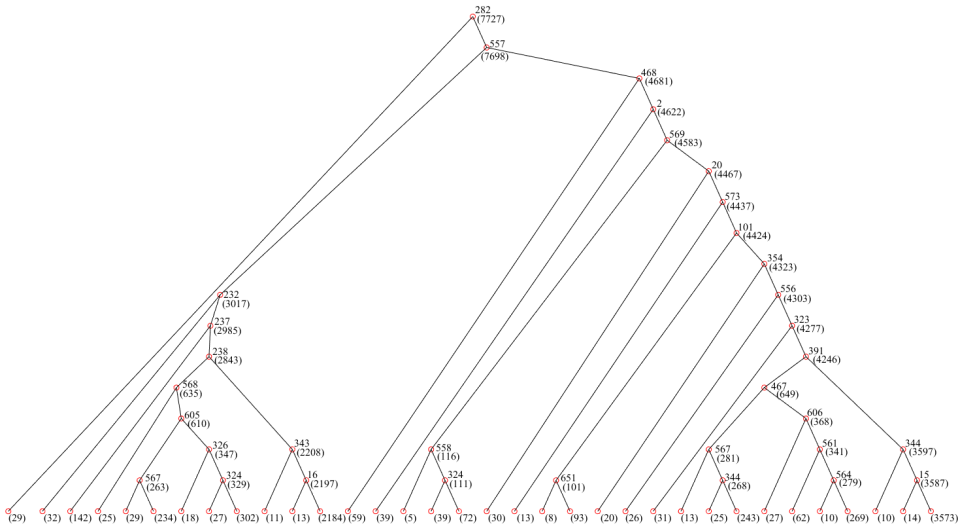


Figure 6.7: The rooted tree as the result of the HRDD method applied to the dihedral angles of the main chain of rhodopsin, extracted from the cluster data found by Pérez-Hernández and Noé [76]. The numbers beside each node refer to the relevant descriptors, i.e. the relevant dihedral angles, the numbers in brackets refer to the numbers of cluster data points that belong to this cluster.

the most relevant motional change. The actual number of metastable states characterized by this descriptor is defined by the number of all eigenvalues of the corresponding transition matrix that are within a tolerance close to 1. The algorithm is recursive, so that the detected clusters are analyzed for further metastabilities characterized by other descriptors.

For the application of the method to the cluster centers of rhodopsin, we chose the dihedral angles of the main chain of rhodopsin to be the selectable descriptors. With the help of VMD, the dihedral angles ϕ ($C_{n\text{-}1}$-N-Cα-C) and ψ (N-Cα-C-N_{n+1}) for each amino acid of the main chain were read out of the files of the almost 8000 cluster centers of rhodopsin found by Pérez-Hernández and Noé. We implemented the method HRDD in MATLAB. Depending on the eigenvalue tolerances, different numbers of clusters are revealed. If the values are too large, the number of clusters is too large, and the aim of dimension reduction cannot be reached. If the tolerance values are too small, relevant dynamical information of the molecule might get lost, and the descriptors would draw an insufficient conclusion. Since the clusters found by HRDD shall become the future starting structures for the μ-opioid receptor, we chose the eigenvalue tolerances with respect to the desired number of starting configurations. Regarding equation (4.4), we selected a tolerance value of 0.045, and for the tolerance value of equation (4.5) we chose 0.05, compare also Table 4.1. In Figure 6.7, the results of the method HRDD applied to the cluster data of rhodopsin are presented in form of a rooted tree. The number beside each

```
>rhodopsin 1U19
MNGTEGPNFYVPFSNKTGVVRSPFEAPQYYLAEPWQFSMLAAYMFLLIMLGFPINFL
TLYVTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEG
FFATLGGEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPP
LVGWSRYIPEGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFFCYGQLVF
TVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQGSDFGP
IFMTIPAFFAKTSAVYNPVIYIMMNKQFRNCMVTTLCCGKN
```

Figure 6.8: The FASTA format for rhodopsin (RCSB PDB 1U19).

[dihedral_restraints]

; ai	aj	ak	al	type	phi	dphi	kfac
2466	2468	2470	2476	1	57.000	0	1000

Figure 6.9: An example for the entry of dihedral restraints in the topology file. For all entries, we chose the values 1 for *type*, 0 for *dphi* and 1000 for *kfac*. This entry belongs to the first cluster with only one descriptor, compare Figure 6.7. The atom indices refer to the atoms C of Pro309, N, CA and C of Glu310, and hence, are characterizing the dihedral angle ϕ between the amino acids proline and glutamate, which has a mean value of 57° in this cluster.

node refers to the descriptor, i.e. the dihedral angle of the main chain, the number in brackets refers to the number of cluster data points that belong to this cluster. On the basis of these 34 clusters and their most relevant descriptors, we set up the starting configurations for the μ-opioid receptor using a multiple sequence alignment. In preparation, we translated the chain of amino acids of both the rhodopsin and the μ-opioid receptor in FASTA format which is originated from the FASTA software [60]. In Figure 6.8, the FASTA sequence is shown for rhodopsin. Applying the COBALT alignment tool [75], the FASTA sequences of rhodopsin and MOR were aligned, so that we could transfer the most relevant descriptors and their average values found by HRDD for rhodopsin to the μ-opioid receptor. This was achieved by adding dihedral restraints in the GROMACS topology file for each cluster, i.e. by setting the values for all relevant descriptors in the respective cluster to the average value. The section of dihedral restraints includes the corresponding atom numbers of the dihedral angles ϕ (C_{n-1}-N-Cα-C) and ψ (N-Cα-C-N_{n+1}), respectively, the average values in the relating cluster and a force factor that constrains the molecule to stay in this configuration, see Figure 6.9. Running a short energy minimization with a restraining of the dihedral angles, finally, led us to the desired starting configurations of the μ-opioid receptor.

The simulations were run at the HLRN (North-German Supercomputing Alliance) as they are computationally too intensive for normal computers. For each of the 34 starting configurations in the protonated and in the deprotonated state the same script was started (68 separate simulations in all). First, an energy minimiza-

tion was performed, using the conjugate gradient integrator, an initial step size of 0.01 ps and a convergence tolerance of 20 kJ/(mol nm) for the maximum force, i.e. the minimization is converged if the maximum force is smaller than this tolerance value. Second, an energy equilibration was conducted at temperature $T = 293\,\mathrm{K}$, with a time step of $\Delta t = 1\,\mathrm{fs}$ and a total number of steps of 1000000 which equals 1 ns. Third and last, the molecular dynamics simulation was run at $T = 293\,\mathrm{K}$ with a time step of $\Delta t = 1\,\mathrm{fs}$ and a total amount of 6000000 steps which corresponds to 6 ns simulation time for each trajectory. All coordinates and other necessary data were written out every 500 time steps, providing data of 12001 frames.

To estimate the global distribution from these local samplings, we applied the in chapter 5 presented reweighting strategies. In preparation, we saved the values of the potential energy in each frame and each trajectory. By selecting only the protein in *gmx trjconv* and rerunning the simulation, we could extract the energy values of the μ-opioid receptor from the whole system, and save them using the GROMACS command *gmx energy*. Furthermore, the values of the dihedral angles were written out with VMD, i.e. ϕ (C-N-Cα-C) and ψ (N-Cα-C-N) for every amino acid of the backbone of the μ-opioid receptor for each trajectory and over all frames. Altogether, 564 dihedral angles over 12001 frames were stored for each of the 68 trajectories. To use the collected data for the two reweighting strategies from chapter 5, that we implemented in MATLAB, we converted it into a MATLAB readable format to calculate the statistical weights in this direct way.

The first method, the Direct Free Energy Reweighting (DFER), requires the values of the potential energy only. On the basis of this data, the mean values of the potential energy were calculated, the equally sized evaluation regions $\mathcal{E}_k$ selected and the number $\tilde{n}_k$ of jump-sampling points $q_i^{(k)}$ inside these evaluation regions counted. Using equation (5.6) and normalizing this vector, the statistical weights w_k were calculated. Each evaluation region $\mathcal{E}_k$ was built around the mean value of the potential energies of the corresponding cluster with a fixed size of

$$s = 2 \cdot sz = 2 \cdot \mathrm{Var}_{k=1,\ldots,b}\left(\sigma_{i=1,\ldots,\tilde{n}_k}\left(E_p\left(q_i^k\right)\right)\right),$$

$\mathrm{Var} = \sigma^2$, i.e. each evaluation region $\mathcal{E}_k$ consists of sampling points with a potential energy value between $[\langle E_p^{(k)}\rangle - sz, \langle E_p^{(k)}\rangle + sz]$. Since the algorithm requires the values of the potential energy only, and evaluates only one equation, it is rather fast, the execution needs less than a second.

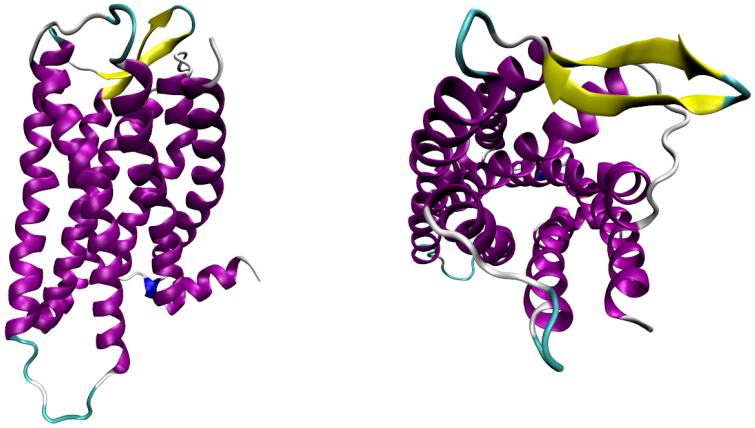
For the implementation of the second strategy, the Estimation of Entropy Differences Reweighting (EEDR), the statistical weights w_i were computed by using the free energy differences $A_i - A_{i-1}$, compare equation (5.10). To estimate the entropy, a certain number R of reference points $q_l^{(i)}$, $l = 1\ldots,R$, with approximately mean potential energy value has to be selected, and the number $n_p^{(i)}$ of sampling points that are near, i.e. within a certain distance around each reference point, counted. For the attribute near, we had to select some distance measure. In case of

the μ-opioid receptor, we set the number of reference points to $R = 50$ and chose the root mean square distance between the dihedral angles as distance measure. In contrast to the DFER strategy, the necessary data for the EEDR comprises not only the values of the potential energy, but also the values of the dihedral angles of the backbone, and several equations have to be evaluated. With a running time of several minutes, the EEDR strategy is much slower than the DFER.

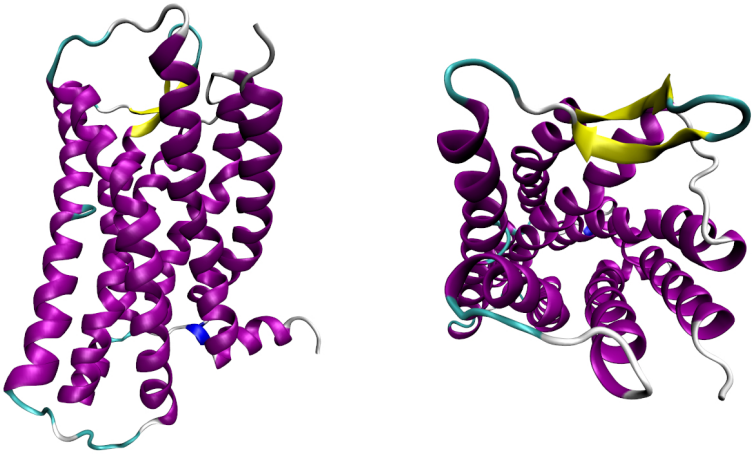
In a first attempt, neither of the two algorithms was successful, since the potential energy values were (absolutely) too high for the use in the exponential functions in equations (5.6) and (5.10), in ranges of -10^4, so that the calculations resulted in *NaN* for the statistical weights. Since a shifting of the energy values according to their mean values, see the modification of DFER in section 5.2, did not lead to the desired results, we applied a scaling of energies knowing full well that this is not physically correct, but we want to consider the energies as a scaling in this way. We tested several values for the scaling factor and discovered that the statistical weights converged to nearly uniform distribution the lower the scaling factor was. For this reason, we chose the highest possible scaling factor $s_f = 0.1$. This worked for both protonation states and both algorithms. In Table 6.2, the statistical weights of each cluster and each protonation state are opposed for both reweighting strategies. It is obvious, that on the one hand, the two reweighting algorithms produce similar results, and that on the other hand, the statistical weights are very different in the deprotonated and protonated state. While, in the deprotonated state, there are several clusters with a statistical weight higher than 10%, in the protonated state, there is only one cluster with a statistical weight higher than 10% but this one has a very high statistical weight of $\approx 85\%$, i.e. in the inflamed tissue, there is one strong favored conformation while in the healthy tissue, several conformations are probable. This result shows parallels to the analysis of the electrostatic potential, see Figure 6.1. There, the deprotonated receptor was negatively charged at several positions, offering many possibilities to interact with the positively charged morphine, while the protonated receptor showed only a few negatively charged amino acids, especially nearby and in the binding pocket, so that the analgesic probably finds its way into the binding pocket easier. Although the results of the reweighting shall be considered carefully, especially because of the essential scaling of energies previously, they suggest that differing conformations of the μ-opioid receptor might be a reason for the better effect of analgesics in healthy tissue, i.e. hypothesis (2) in section 2.4 might be true. In Figure 6.10, the two most probable conformations are opposed for the deprotonated and the protonated system. In the protonated system, the transmembrane helix 6 (TM6) is displaced outwards, which was described by Huang et al. [45] as a characteristic for the receptor in the active state, necessary for the binding of the G protein. Additionally, we can observe that the way into the binding pocket is more open in the most probable cluster of the protonated system.

#	deprotonated		protonated	
	DFER	EEDR	DFER	EEDR
1	0.016	0.017	0.846	0.848
2	0.059	0.069	0.000	0.000
3	0.001	0.002	0.000	0.000
4	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000
6	0.134	0.131	0.000	0.000
7	0.000	0.000	0.000	0.000
8	0.000	0.000	0.001	0.001
9	0.000	0.000	0.000	0.000
10	0.001	0.001	0.000	0.000
11	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000
13	0.000	0.000	0.001	0.001
14	0.000	0.000	0.000	0.000
15	0.163	0.163	0.000	0.000
16	0.000	0.000	0.000	0.000
17	0.000	0.000	0.000	0.000
18	0.013	0.011	0.000	0.000
19	0.001	0.002	0.001	0.001
20	0.000	0.000	0.000	0.000
21	0.010	0.009	0.029	0.039
22	0.000	0.000	0.000	0.000
23	0.000	0.000	0.022	0.016
24	0.011	0.011	0.004	0.005
25	0.000	0.000	0.094	0.088
26	0.002	0.003	0.000	0.000
27	0.256	0.252	0.000	0.000
28	0.000	0.000	0.000	0.000
29	0.000	0.000	0.000	0.000
30	0.000	0.000	0.000	0.000
31	0.005	0.006	0.000	0.000
32	0.119	0.131	0.000	0.000
33	0.000	0.000	0.001	0.001
34	0.208	0.194	0.000	0.000

Table 6.2: The results of the statistical weighting of the conformations of the μ-opioid receptor at $T = 293\,\mathrm{K}$. Both, for the deprotonated and the protonated state, the two strategies Direct Free Energy Reweighting (DFER) and Estimation of Entropy Differences Reweighting (EEDR) are opposed to each other.



(a) Cluster 27 of the deprotonated system with a statistical weight of $w_{27} \approx 25\%$.



(b) Cluster 1 of the protonated system with a statistical weight of $w_1 \approx 85\%$.

Figure 6.10: The two most probable conformations of the μ-opioid receptor for the deprotonated and the protonated system based on separate samplings in 34 clusters and a following reweighting. Visualization with VMD, secondary structure, (left) front view, (right) view into the binding pocket. In the protonated system (b), the transmembrane helix 6 (TM6) is displaced outwards, which is described by Huang et al. [45] as a characteristic for the receptor in the active state. Additionally, we can observe that the way into the binding pocket is more open in (b).

6.4 Conclusion and discussion

The μ-opioid receptor is an important transmembrane protein that is responsible for the perception and alleviation of pain. Because of the increasing misuse and overdosing of analgesics during the last years and the threat of serious side effects, the investigation of the μ-opioid receptor (MOR) became more and more important. By understanding the responsible mechanisms that lead to the activation of the receptor in the inflamed tissue, new analgesics with no or only few side effects can be developed, hopefully.

One of the major challenges is the extraction of the MOR out of the organism, and the complete presentation as crystal structure. The molecular model of the μ-opioid receptor in this thesis was built based on the crystal structure found in 2012 by Manglik et al. [65]. For their studies, they used the μ-opioid receptor from the Mus musculus, the house mouse, emphasizing that it differs only in four residues in the resolved part of the structure from the human MOR, none with contacts in the ligand-binding pocket. However, Seok et al. stated in 2013 [89] that mouse models are not useful to study human inflammatory diseases, starting a controversial debate. Two years later, Takao and Miyakawa [101] analyzed the same data but with different assumptions and methods and found exceptionally significant correlations between mouse and human models. Nevertheless, the use of mice models in the research of human diseases requires the careful attention of differences and similarities between mice and men [77].

Another source of error is the incompleteness of the crystal structure of the μ-opioid receptor on which the molecular model in this thesis is based. Single missing atoms of existing residues could be added using the GROMACS software, but it was not possible to append whole amino acids to the receptor afterwards, so that the molecular model of the MOR stays incomplete. Some of the missing residues that were not located in the experiment but that were listed with their sequence number could be used for the alignment with rhodopsin since only the residue names and not the position of single atoms is required therefor.

Before we could start with the analysis of the μ-opioid receptor, we had to create a molecular model. Naturally, the parametrization of the receptor is only an approximation of the real behavior of the MOR. From many possible force fields, we, hopefully, chose the most suitable. Anyhow, the simulation box represents only a minor part of the environment of the μ-opioid receptor, so that interactions with other (membrane) proteins or ion channels cannot be taken into account. The setting of the inflamed tissue was implemented by changing the protonation state, assuming that the pH value is the only differing characteristic in the inflamed tissue. For this, additional hydrogens were attached to each amino acid histidine in the molecular structure of the MOR.

In the first part of this chapter, we analyzed the Coulomb interactions between ligand and receptor (MOR). The comparison of the electrostatic potential maps of

the μ-opioid receptor in healthy and inflamed tissue lead to the assumption that a positively charged ligand, e.g. morphine, finds its way into the binding pocket faster. Though, several simulations with the small, positively charged molecule ammonium showed that, in the protonated system, the entropy is higher, i.e. the ammonium is moving much more, and there are almost no electrostatic interactions between the ammonium and the MOR. This observation can be justified by the chemical properties since a positively charged ligand is repulsed by a positively charged receptor. With these results, hypothesis (1) stated in section 2.4 could be falsified.

In the second part of this chapter, we analyzed hypothesis (2), i.e. the conformations of the receptor in healthy and inflamed tissue. The generation of starting structures for the separate samplings is based on cluster data of rhodopsin, using the in chapter 4.3 introduced method of the Hierarchical Relevant Descriptor Detector and a sequence alignment. Although both receptors belong to the same family, the results of rhodopsin cannot be transferred one-to-one to the MOR. Anyhow, since the starting structures are only a tool to start separate trajectories independent from the data of rhodopsin, transcription errors are insignificant. With 34 separate trajectories of 6 ns for each protonation state we sampled (parts of) the conformation space. Since there is no convergence criterion for the parallel running molecular dynamics simulations, and since the starting points are probably not distributed over the whole conformation space, the 34 trajectories did probably not sample the whole conformation space. However, we assumed that the trajectories covered the most important parts of the conformation space, and applied the in chapter 5 presented direct reweighting strategies, as we cannot easily derive the global distribution from the local distributions. The results are very promising: in the healthy tissue, there were five clusters with a statistical weight higher than 10%, but all less than 26%, comparable to the electrostatic potential map where the deprotonated receptor was negatively charged at several positions, offering many possibilities to interact with the ligand. In contrast, in the inflamed tissue, there is only one cluster with a very high statistical weight of 85%, similar to the few negatively charged amino acids in the electrostatic potential map of the protonated receptor. A visualization of the two most probable conformations of the deprotonated and the protonated receptor showed the typical outward displacement of the transmembrane helix TM6 and the more opened binding pocket in the inflamed tissue. Unfortunately, we had to scale the values of the potential energy for the calculations of the statistical weights because they were (absolutely) too high. For this reason, and also because all calculations during the analysis of the μ-opioid receptor are susceptible to errors, the results have to be considered carefully. Anyhow, following these results, the reason for the better effect of analgesics in inflamed tissue is (at least partly) due to the different conformations of the MOR.

Chapter 7

Outlook

The investigation of pain relieving drugs that act in the inflamed tissue only and, thus, have no side effects, gets more and more important [36, 110, 80]. The idea of polarizing opioid ligands in order to decrease the blood-brain barrier permeability [97, 107, 96] failed because of peripheral side effects or unfeasible blood-brain barrier impermeability at higher doses. Another approach involved carriers that were engineered for the target delivery of drugs [44, 15], but was unsuccessful due to allergic reactions to proteins, instability in the circulation and other reasons. Besides the promising approach of Spahn et al. [94], where the novel agent NFEPP selectively activates the opioid receptors in the inflamed tissue only, González-Rodríguez et al. [34] found a way to selectively release morphine in injured tissue only, by covalently attaching it to a hyperbranched polyglycerol using a cleavable linker. The approach of Spahn et al. is unprecedented since it is based on structural changes of ligand and receptor under pathological (dynamic) conditions instead of considering physiological (static) conformations only, and shall build the basis for a new class of pain relieving drugs, hopefully. Both novel agents are already successfully tested in vitro and in vivo, and may serve as prototypes for analgesics without side effects in the future.

The effect of the pH value on the μ-opioid receptor was discussed in this thesis. As another approach, the influence of the temperature could be investigated since inflamed tissue has a higher temperature than healthy tissue. Temperature-dependent drugs are also important for the application in targeted temperature management, i.e. lowering of the body temperature to prevent brain injury after an accident or cardiac arrest. Looking a step further into the future, with temperature-dependent drugs, even a human hibernation might be possible [57].

For the validation of our computational results, chemical experiments can be conducted. With the ATR-IR method (attenuated total reflection infrared spectroscopy), liquid samples can be examined. By projecting the IR radiation on a prism of silicium, the resulting evanescent wave typically goes 1 μm deep into the

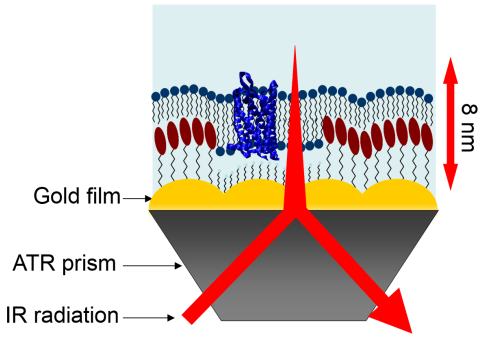


Figure 7.1: ATR-IR with SEIRA (with kind permission of Dr. Jacek Kozuch).

sample. The surface enhancement SEIRA (surface-enhanced infrared absorption spectroscopy) [4] enables a depth of only 8 nm by coating the prism with a nano-structured gold film of 50 nm thickness, and, thus, allows the study of monolayers, compare Figure 7.1. The IR radiation generates an electrical field above the gold film that is nearly perpendicular within the first 8 nm, leading to an enhancement of intensity and a perpendicular polarization. In this way, vertical vibrations are measured with a higher intensity, so that the orientation of the sample can be evaluated. By using the gold film for the binding of the protein within the membrane, the protein is in a preferably native environment and firmly bound to the surface, so that only small amounts of the protein are necessary and modifications at the same mono-film are possible, like changing the pH value or adding ligands. The results of this experiment can be illustrated with a spectrum that describes the secondary structure in terms of characteristic bands, so called amide vibrations. By inducing local structural changes, the spectrum shows several sharp peaks, representing changes of single bonds, e.g. (de)protonation of acid/basic groups, but also changes in the secondary structure. The presented experiments were planned in cooperation with the Institute for Chemistry of Technische Universität Berlin, but the collaborators did not succeed in extracting and purifying the required μ-opioid receptor.

The redundancy of extensively chemical processing is one of the reasons why computer based drug design is becoming increasingly important. Furthermore, computational methods can help to save time and expense for synthesizing each possible compound by calculating binding affinities and finding the optimal structure with the computer [50]. Another incentive is the desired reduction of animal testing, at least in the initial phase of the drug design. Despite the expected ongoing increase of computing power in the future which facilitates molecular simulations of even larger molecules and time scales, the *big data* has to be stored, processed and analyzed, however. Coarse graining methods as presented in chapter 4 will play an important role, and new strategies in the theory of Markov State Models (MSM), that we introduced in chapter 3.3 as conformation dynamics, are developed [113].

For instance, by combining Markov State Models with reaction-diffusion (RD), simulations at large timescales but with observations in atomic detail might soon be possible [29].

Essential for the development of the novel agent NFEPP is the close collaboration of mathematicians, biologists, chemists and clinicians. The combination of their knowledge and competence facilitates the efficient translation of the results from in silico, in vitro and in vivo studies. Hopefully, the novel opioid ligand will succeed in clinical studies, and soon, be used as a pain relieving drug without side effects, helping millions of people all over the world, in this way.

Bibliography

[1] *Marvin 15.1.26*, ChemAxon (http://www.chemaxon.com), 2015. (Cited on page 53.)

[2] R. Al-Hasani and M. R. Bruchas, *Molecular Mechanisms of Opioid Receptor-dependent Signaling and Behavior*, Anesthesiology **115** (2011), no. 6, 1363–1381. (Cited on page 12.)

[3] B. J. Alder and T. E. Wainwright, *Phase Transition for a Hard Sphere System*, The Journal of Chemical Physics **27** (1957), no. 5, 1208–1209. (Cited on page 25.)

[4] K. Ataka, S. T. Stripp, and J. Heberle, *Surface-enhanced infrared absorption spectroscopy (SEIRAS) to probe monolayers of membrane proteins*, Biochimica et Biophysica Acta (BBA) - Biomembranes **1828** (2013), no. 10, 2283–2293. (Cited on page 68.)

[5] E. Barth, K. Kuczera, B. Leimkuhler, and R. D. Skeel, *Algorithms for Constrained Molecular Dynamics*, Journal of Computational Chemistry **16** (1995), no. 10, 1192–1209. (Cited on page 25.)

[6] A. L. Beberg, D. L. Ensign, G. Jayachandran, S. Khaliq, and V. S. Pande, *Folding@home: Lessons from eight years of volunteer distributed computing*, 2009 IEEE International Symposium on Parallel Distributed Processing, 2009, pp. 1–8. (Cited on page 25.)

[7] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *Molecular dynamics with coupling to an external bath*, The Journal of Chemical Physics **81** (1984), no. 8, 3684–3690. (Cited on page 38.)

[8] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, *Interaction Models for Water in Relation to Protein Hydration*, Intermolecular Forces (B. Pullman, ed.), vol. 14, The Jerusalem Symposia on Quantum

Chemistry and Biochemistry, Springer-Verlag, 1981, pp. 331–342. (Cited on page 14.)

[9] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*, Computer Physics Communications **91** (1995), no. 1-3, 43–56. (Cited on pages 13, 25, and 31.)

[10] O. Berger, O. Edholm, and F. Jähnig, *Molecular Dynamics Simulations of a Fluid Bilayer of Dipalmitoylphosphatidylcholine at Full Hydration, Constant Pressure, and Constant Temperature*, Biophysical Journal **72** (1997), no. 5, 2002–2013. (Cited on page 13.)

[11] D. Bonchev and D. H. Rouvray (eds.), *Chemical graph theory: introduction and fundamentals*, vol. 1, Gordon and Breach Science Publishers, 1991. (Cited on page 26.)

[12] G. R. Bowman, K. A.Beauchamp, G. Boxer, and V. S.Pande, *Progress and challenges in the automated construction of Markov state models for full protein systems*, The Journal of Chemical Physics **131** (2009), no. 12, 124101. (Cited on page 25.)

[13] N.-V. Buchete and G. Hummer, *Coarse Master Equations for Peptide Folding Dynamics*, The Journal of Physical Chemistry B **112** (2008), no. 19, 6057–6069. (Cited on page 25.)

[14] G. Bussi, D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*, The Journal of Chemical Physics **126** (2007), no. 1, 014101. (Cited on page 31.)

[15] Z. Cheng, A. Al Zaki, J. Z. Hui, V. R. Muzykantov, and A. Tsourkas, *Multifunctional Nanoparticles: Cost Versus Benefit of Adding Targeting and Imaging Capabilities*, Science **338** (2012), no. 6109, 903–910. (Cited on page 67.)

[16] C. Chipot and A. Pohorille (eds.), *Free Energy Calculations – Theory and Applications in Chemistry and Biology*, Springer-Verlag, Berlin, 2007. (Cited on pages 37, 39, and 49.)

[17] J. D. Chodera and F. Noé, *Probability distributions of molecular observables computed from Markov models. II. Uncertainties in observables and their time-evolution*, The Journal of Chemical Physics **133** (2010), no. 10, 105102. (Cited on page 25.)

[18] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, *Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics*, The Journal of Chemical Physics **126** (2007), no. 15, 155101. (Cited on page 25.)

[19] J. D. Chodera, W. C. Swope, J. W. Pitera, and K. A. Dill, *Long-Time Protein Folding Dynamics from Short-Time Molecular Dynamics Simulations*, Multiscale Modeling & Simulation **5** (2006), no. 4, 1214–1226. (Cited on page 25.)

[20] K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, Springer-Verlag, Berlin, 1967. (Cited on pages 20 and 40.)

[21] A. W. S. da Silva and W. F. Vranken, *ACPYPE - AnteChamber PYthon Parser interfacE*, BMC Research Notes **5** (2012), no. 1, 367. (Cited on page 53.)

[22] M. Dehmer, K. Varmuza, and D. Bonchev (eds.), *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*, vol. 2, Wiley-VCH Verlag & Co, Weinheim, 2012. (Cited on page 26.)

[23] M. Dellnitz and A. Hohmann, *A subdivision algorithm for the computation of unstable manifolds and global attractors*, Numerische Mathematik **75** (1997), no. 3, 293–317. (Cited on page 22.)

[24] M. Dellnitz and O. Junge, *On the Approximation of Complicated Dynamical Behavior*, SIAM Journal on Numerical Analysis **36** (1999), no. 2, 491–515. (Cited on page 22.)

[25] P. Deuflhard, *From Molecular Dynamics to Conformational Dynamics in Drug Design*, Trends in Nonlinear Analysis (M. Kirkilionis, S. Krömker, R. Rannacher, and F. Tomi, eds.), Springer, 2003, pp. 269–287. (Cited on page 22.)

[26] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte, *Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains*, Linear Algebra and its Applications **315** (2000), no. 1, 39–59. (Cited on page 22.)

[27] P. Deuflhard and C. Schütte, *Molecular Conformation Dynamics and Computational Drug Design*, Applied Mathematics Entering the 21st Century (J. Hill and R. Moore, eds.), vol. 116, 2004, pp. 91–119. (Cited on page 22.)

[28] P. Deuflhard and M. Weber, *Robust Perron cluster analysis in conformation dynamics*, Linear Algebra and its Applications **398** (2005), 161–184. (Cited on pages 22 and 25.)

[29] M. Dibak, M. J. del Razo, D. De Sancho, C. Schütte, and F. Noé, *MSM/RD: Coupling Markov state models of molecular kinetics with reaction-diffusion simulations*, arXiv preprint arXiv:1712.08149 (2017). (Cited on page 69.)

[30] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Physics Letters B **195** (1987), no. 2, 216–222. (Cited on pages 21, 31, 38, and 47.)

[31] K. Fackeldey, M. Klimm, and M. Weber, *A coarse graining method for the dimension reduction of the state space of biomolecules*, Journal of Mathematical Chemistry **50** (2012), no. 9, 2623–2635. (Cited on page 9.)

[32] A. Fischer, C. Schütte, P. Deuflhard, and F. Cordes, *Hierarchical Uncoupling-Coupling of Metastable Conformations*, vol. 24, pp. 235–259, Springer-Verlag, Berlin, 2002. (Cited on pages 25 and 31.)

[33] A. Gelman and D. B. Rubin, *Inference from Iterative Simulation Using Multiple Sequences*, Statistical Science **7** (1992), no. 4, 457–472. (Cited on pages 21 and 47.)

[34] S. González-Rodríguez, M. A. Quadir, S. Gupta, K. A. Walker, X. Zhang, V. Spahn, D. Labuz, A. Rodriguez-Gaztelumendi, M. Schmelz, J. Joseph, M. K. Parr, H. Machelska, R. Haag, and C. Stein, *Polyglycerol-opioid conjugate produces analgesia devoid of side effects*, eLife **6** (2017). (Cited on page 67.)

[35] M. Griebel, S. Knapek, and G. Zumbusch, *Numerical Simulation in Molecular Dynamics*, Texts in Computational Science and Engineering, vol. 5, Springer-Verlag, Berlin, 2007. (Cited on page 25.)

[36] T. Grosser, C. J. Woolf, and G. A. FitzGerald, *Time for nonaddictive relief of pain*, Science **355** (2017), no. 6329, 1026–1027. (Cited on page 67.)

[37] T. A. Halgren, *Merck molecular force field. I-V.*, Journal of Computational Chemistry **17** (1996), no. 5-6, 490–641. (Cited on pages 31 and 47.)

[38] G. S. Hammond, *A Correlation of Reaction Rates*, Journal of American Chemical Society **77** (1955), no. 2, 334–338. (Cited on page 26.)

[39] W. Häuser, F. Bock, P. Engeser, T. Tölle, A. Willweber-Strumpf, and F. Petzke, *Clinical Practice Guideline: Long-Term Opioid Use in Non-Cancer Pain.*, Deutsches Ärzteblatt International **111** (2014), no. 43, 732–740. (Cited on page 7.)

[40] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation*, Journal of Chemical Theory and Computation **4** (2008), no. 3, 435–447. (Cited on pages 25 and 31.)

[41] W. G. Hoover, *Time Reversibility, Computer Simulation and Chaos*, World Scientific, 1999. (Cited on page 38.)

[42] W. G. Hoover, A. Kenichiro, C. G. Hoover, and S. V. deGroot, *Time-reversible deterministic thermostats*, Physica D: Nonlinear Phenomena **187** (2004), no. 1-4, 253–267. (Cited on page 38.)

[43] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, *Comparison of multiple AMBER force fields and development of improved protein backbone parameters*, Proteins **65** (2006), no. 3, 712–725. (Cited on page 31.)

[44] S. Hua and P. J. Cabot, *Targeted Nanoparticles that Mimic Immune Cells in Pain Control Inducing Analgesic and Anti-inflammatory Actions: A Potential Novel Treatment of Acute and Chronic Pain Conditions*, Pain Physician **16** (2013), no. 3, 199–216. (Cited on page 67.)

[45] W. Huang, A. Manglik, A. J. Venkatakrishnan, T. Laeremans, E. N. Feinberg, A. L. Sanborn, H. E. Kato, K. E. Livingston, T. S. Thorsen, R. C. Kling, S. Granier, P. Gmeiner, S. M. Husbands, J. R. Traynor, W. I. Weis, J. Steyaert, R. O. Dror, and B. K. Kobilka, *Structural insights into μ-opioid receptor activation*, Nature **524** (2015), no. 7565, 315–321. (Cited on pages 61 and 63.)

[46] W. Huisinga, *Metastability of Markovian systems: A transfer operator based approach in application to molecular dynamics*, Doctoral thesis, Freie Universität Berlin, 2001. (Cited on page 38.)

[47] W. Humphrey, A. Dalke, and K. Schulten, *VMD: Visual Molecular Dynamics*, Journal of Molecular Graphics **14** (1996), no. 1, 33–38. (Cited on page 13.)

[48] D. G. Isom, V. Sridharan, R. Baker, S. T. Clement, D. M. Smalley, and H. G. Dohlman, *Protons as Second Messenger Regulators of G Protein Signaling*, Molecular Cell **51** (2013), no. 4, 531–538. (Cited on page 17.)

[49] U. Issberner, P. W. Reeh, and K. H. Steen, *Pain due to tissue acidosis: a mechanism for inflammatory and ischemic myalgia?*, Neuroscience Letters **208** (1996), no. 3, 191–194. (Cited on page 8.)

[50] A. Jain, *Computer aided drug design*, Journal of Physics: Conference Series **884** (2017), no. 1. (Cited on page 68.)

[51] C. R. Johnson and R. A. Horn, *Matrix Analysis*, Cambridge University Press, Cambridge, 1990. (Cited on pages 28 and 41.)

[52] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids*, Journal of the American Chemical Society **118** (1996), no. 45, 11225–11236. (Cited on page 31.)

[53] C. Kandt, W. L. Ash, and D. P. Tielemann, *Setting up and running molecular dynamics simulations of membrane proteins*, Methods **41** (2007), no. 4, 475–488. (Cited on page 13.)

[54] M. Klimm, A. Bujotzek, and M. Weber, *Direct Reweighting Strategies in Conformation Dynamics*, MATCH Communications in Mathematical and in Computer Chemistry **65** (2011), no. 2, 333–346. (Cited on page 9.)

[55] B. Kozlíková, M. Krone, N. Lindow, M. Falk, M. Baaden, D. Baum, I. Viola, J. Parulek, and H.-C. Hege, *Visualization of Biomolecular Structures: State of the Art*, Computer Graphics Forum (2016). (Cited on page 15.)

[56] A. R. Leach, *Molecular Modelling: Principles and Applications*, 2 ed., ch. 5, Pearson Education Limited, 2001. (Cited on page 32.)

[57] C. C. Lee, *Is Human Hibernation Possible?*, Annual Review of Medicine **59** (2008), no. 1, 177–186. (Cited on page 67.)

[58] J.-G. Li, C. Chen, J. Yin, K. Rice, Y. Zhang, D. Matecka, J. K. de Riel, R. L. DesJarlais, and L.-Y. Liu-Chen, *Asp147 in the third transmembrane helix of the rat μopioid receptor forms ion-pairing with morphine and naltrexone*, Life Sciences **65** (1999), no. 2, 175–185. (Cited on page 15.)

[59] E. Lindahl, B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis*, Journal of Molecular Modeling **7** (2001), no. 8, 306–317. (Cited on page 31.)

[60] D. J. Lipman and W. R. Pearson, *Rapid and Sensitive Protein Similarity Searches*, Science **227** (1985), no. 4693, 1435–1441. (Cited on page 59.)

[61] S. P. Lloyd, *Least Squares Quantization in PCM*, IEEE Transactions on Information Theory **28** (1982), no. 2, 129–137. (Cited on page 22.)

[62] E. Lyman and D. M. Zuckerman, *Ensemble-Based Convergence Analysis of Biomolecular Trajectories*, Biophysical Journal **91** (2006), no. 1, 164–172. (Cited on page 25.)

[63] A. O. Lyubartsev and A. L. Rabinovich, *Force Field Development for Lipid Membrane Simulations*, Biochimica et Biophysica Acta - Biomembranes **1858** (2016), no. 10, 2483–2497. (Cited on page 13.)

[64] L. Manchikanti and A. Singh, *Therapeutic Opioids: A Ten-Year Perspective on the Complexities and Complications of the Escalating Use, Abuse, and Nonmedical Use of Opioids*, Pain Physician **11** (2008), 63–88. (Cited on page 7.)

[65] A. Manglik, A. C. Kruse, T. S. Kobilka, F. S. Thian, J. M. Mathiesen, R. K. Sunahara, L. Pardo, W. I. Weis, B. K. Kobilka, and S. Granier, *Crystal structure of the μ-opioid receptor bound to a morphinan antagonist*, Nature **485** (2012), no. 7398, 321–326. (Cited on pages 13 and 64.)

[66] S. J. Marrink, A. H. de Vries, and A. E. Mark, *Coarse Grained Model for Semiquantitative Lipid Simulations*, The Journal of Physical Chemistry B **108** (2004), no. 2, 750–760. (Cited on page 13.)

[67] S. Marsili, G. F. Signorini, R. Chelli, M. Marchi, and P. Procacci, *ORAC: A molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level*, Journal of Computational Chemistry **31** (2010), no. 5, 1106–1116. (Cited on page 25.)

[68] The Mathworks, Inc., Natick, Massachusetts, *Matlab version r2014a*, 2014. (Cited on page 52.)

[69] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of State Calculations by Fast Computing Machines*, The Journal of Chemical Physics **21** (1953), no. 6, 1087–1092. (Cited on page 21.)

[70] L. Monticelli, E. J. Sorin, D. P. Tieleman, V. S. Pande, and G. Colombo, *Molecular simulation of multistate peptide dynamics: A comparison between microsecond timescale sampling and multiple shorter trajectories*, Journal of Computational Chemistry **29** (2008), no. 11, 1740–1752. (Cited on page 25.)

[71] D. Nockemann, *Coupling between opioid receptors and potassium channels in pain reduction: relevance for species differences and efficacy*, Doctoral thesis, Freie Universität Berlin, 2012. (Cited on page 12.)

[72] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, *Open Babel: An open chemical toolbox*, Journal of Cheminformatics **3** (2011), no. 1, 33. (Cited on page 52.)

[73] T. Okada, M. Sugihara, A. N. Bondar, M. Elstner, P. Entel, and V. Buss, *The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure*, Journal of Molecular Biology **342** (2004), no. 2, 571–583. (Cited on page 57.)

[74] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren, *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6*, Journal of Computational Chemistry **25** (2004), no. 13, 1656–1676. (Cited on page 13.)

[75] J. S. Papadopoulos and R. Agarwala, *COBALT: constraint-based alignment tool for multiple protein sequences*, Bioinformatics **23** (2007), no. 9, 1073–1079. (Cited on page 59.)

[76] G. Pérez-Hernández and F. Noé, *Opsin Activation Mechanism*, in preparation. (Cited on pages 57 and 58.)

[77] R. L. Perlman, *Mouse models of human disease: An evolutionary perspective*, Evolution, Medicine, and Public Health **2016** (2016), no. 1, 170–176. (Cited on page 64.)

[78] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, *Scalable molecular dynamics with NAMD*, Journal of Computational Chemistry **26** (2005), no. 16, 1781–1802. (Cited on page 25.)

[79] D. Poger, W. F. van Gunsteren, and A. E. Mark, *A new force field for simulating phosphatidylcholine bilayers*, Journal of Computational Chemistry **31** (2010), no. 6, 1117–1125. (Cited on page 13.)

[80] S. Puig and H. B. Gutstein, *Opioids: keeping the good, eliminating the bad*, Nature Medicine **23** (2017), no. 3, 272–273. (Cited on page 67.)

[81] M. M. Reif, P. H. Hünenberger, and C. Oostenbrink, *New Interaction Parameters for Charged Amino Acid Side Chains in the GROMOS Force Field*, Journal of Chemical Theory and Computation **8** (2012), no. 10, 3705–3723. (Cited on page 13.)

[82] S. Renfrey, C. Downton, and J. Featherstone, *The painful reality*, Nature Reviews Drug Discovery **2** (2003), no. 3, 175–176. (Cited on page 7.)

[83] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, *An overview of the Amber biomolecular simulation package*, Wiley Interdisciplinary Reviews: Computational Molecular Science **3** (2013), no. 2, 198–210. (Cited on page 13.)

[84] T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary guide*, Springer-Verlag, New York, 2002. (Cited on page 38.)

[85] J. Schmidt-Ehrenberg, D. Baum, and H. C. Hege, *Visualizing dynamic molecular conformations*, Proceedings of the conference on Visualization '02, IEEE Computer Society, 2002, pp. 235–242. (Cited on pages 47 and 52.)

[86] L. D. Schuler, X. Daura, and W. F. van Gunsteren, *An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase*, Journal of Computational Chemistry **22** (2001), no. 11, 1205–1218. (Cited on page 13.)

[87] C. Schütte, *Conformational Dynamics: Modelling, Theory, Algorithm and Application to Biomolecules*, Habilitation thesis, Freie Universität Berlin, 1999. (Cited on pages 22, 25, 27, 31, 47, and 48.)

[88] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, *A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo*, Journal of Computational Physics **151** (1999), no. 1, 146–168. (Cited on page 25.)

[89] J. Seok, H. S. Warren, A. G. Cuenca, M. N. Mindrinos, H. V. Baker, W. Xu, D. R. Richards, G. P. McDonald-Smith, H. Gao, L. Hennessy, C. C. Finnerty, C. M. López, S. Honari, E. E. Moore, J. P. Minei, J. Cuschieri, P. E. Bankey, J. L. Johnson, J. Sperry, A. B. Nathens, T. R. Billiar, M. A. West, M. G. Jeschke, M. B. Klein, R. L. Gamelli, N. S. Gibran, B. H. Brownstein, C. Miller-Graziano, S. E. Calvano, P. H. Mason, J. P. Cobb, L. G. Rahme, S. F. Lowry, R. V. Maier, L. L. Moldawer, D. N. Herndon, R. W. Davis, W. Xiao, R. G. Tompkins, and the Inflammation and Host Response to Injury, Large Scale Collaborative Research Program, *Genomic responses in mouse models poorly mimic human inflammatory diseases*, Proceedings of the National Academy of Sciences of the United States of America (PNAS) **110** (2013), no. 9, 3507–3512. (Cited on page 64.)

[90] J. Shim, A. Coop, and A. D. MacKerell Jr., *Molecular Details of the Activation of the μOpioid Receptor*, The Journal of Physical Chemistry B **117** (2013), no. 26, 7907–7917. (Cited on page 15.)

[91] H. A. Simon and A. Ando, *Aggregation of variables in dynamic systems*, Econometrica **29** (1961), no. 2, 111–138. (Cited on page 41.)

[92] L. J. Smith, X. Daura, and W. F. van Gunsteren, *Assessing equilibration and convergence in biomolecular simulations*, Proteins: Structure, Function, and Bioinformatics **48** (2002), no. 3, 487–496. (Cited on page 25.)

[93] P. E. Smith, *The alanine dipeptide free energy surface in solution*, The Journal of Chemical Physics **111** (1999), no. 12, 5568–5579. (Cited on pages 31 and 32.)

[94] V. Spahn, G. Del Vecchio, D. Labuz, A. Rodriguez-Gaztelumendi, N. Massaly, J. Temp, V. Durmaz, P. Sabri, M. Reidelbach, H. Machelska, M. Weber, and C. Stein, *A nontoxic pain killer designed by modeling of pathological receptor conformations*, Science **355** (2017), no. 6328, 966–969. (Cited on pages 3, 8, 9, 12, 17, and 67.)

[95] C. Stein, *Opioid Receptors*, Annual Review of Medicine **67** (2016), no. 1, 433–451. (Cited on page 12.)

[96] C. Stein and S. Küchler, *Targeting inflammation and wound healing by opioids*, Trends in Pharmacological Sciences **34** (2013), no. 6, 303–312. (Cited on page 67.)

[97] C. Stein and H. Machelska, *Modulation of Peripheral Sensory Neurons by the Immune System: Implications for Pain Therapy*, Pharmacological Reviews **63** (2011), no. 4, 860–881. (Cited on page 67.)

[98] C. Stein, M. Weber, O. Scharkoi, and P. Deuflhard, *Method and system for identifying compounds that bind and preferably activate a target opioid receptor in a ph-dependent manner*, European Patent Application, Bulletin 2013/28, 2013, EP2613277. (Cited on page 9.)

[99] C. Stein, C. Zöllner, M. Weber, and O. Scharkoi, *Fentanyl derivatives as ph-dependent opioid receptor agonists*, European Patent Application, Bulletin 2013/08, 2013, EP2559685. (Cited on page 9.)

[100] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters*, The Journal of Chemical Physics **76** (1982), no. 1, 637–649. (Cited on page 20.)

[101] K. Takao and T. Miyakawa, *Genomic responses in mouse models greatly mimic human inflammatory diseases*, Proceedings of the National Academy of Sciences of the United States of America (PNAS) **112** (2015), no. 4, 1167–1172. (Cited on page 64.)

[102] R. L. Thurlkill, G. R. Grimsley, J. M. Scholtz, and C. N. Pace, *pK values of the ionizable groups of proteins*, Protein Science **15** (2006), no. 5, 1214–1218. (Cited on page 16.)

[103] D. P. Tieleman and H. J. C. Berendsen, *Molecular dynamics simulations of a fully hydrated dipalmitoylphosphatidylcholine bilayer with different macroscopic boundary conditions and parameters*, The Journal of Chemical Physics **105** (1996), no. 11, 4871–4880. (Cited on page 13.)

[104] R. Todeschini and V. Consonni, *Molecular descriptors for chemoinformatics*, vol. 41, John Wiley & Sons, 2009. (Cited on page 26.)

[105] M. E. Tuckerman and B. J. Berne, *Molecular dynamics in systems with multiple time scales*, The Journal of Chemical Physics **95** (1991), no. 11, 8362–8364. (Cited on page 25.)

[106] M. E. Tuckerman, B. J. Berne, and A. Rossi, *Molecular dynamics algorithm for multiple time scales: Systems with disparate masses*, The Journal of Chemical Physics **94** (1991), no. 2, 1465–1469. (Cited on page 25.)

[107] N. Vadivelu, S. Mitra, and R. L. Hines, *Peripheral opioid receptor agonists for analgesia: a comprehensive review*, Journal of Opioid Management **7** (2011), no. 1, 55–68. (Cited on page 67.)

[108] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, *GROMACS: Fast, Flexible and Free*, Journal of Computational Chemistry **26** (2005), no. 16, 1701–1718. (Cited on page 31.)

[109] W. F. van Gunsteren and H. J. C. Berendsen, *Algorithms for brownian dynamics*, Molecular Physics **45** (1982), no. 3, 637–647. (Cited on page 25.)

[110] G. Del Vecchio, V. Spahn, and C. Stein, *Novel Opioid Analgesics and Side Effects*, ACS Chemical Neuroscience **8** (2017), no. 8, 1638–1640. (Cited on page 67.)

[111] M. Weber, *Meshless Methods in Conformation Dynamics*, Doctoral thesis, Freie Universität Berlin, 2006. (Cited on page 22.)

[112] M. Weber and K. Andrae, *A simple method for the estimation of entropy differences*, MATCH Communications in Mathematical and in Computer Chemistry **63** (2010), no. 2, 319–332. (Cited on pages 42 and 43.)

[113] M. Weber, K. Fackeldey, and C. Schütte, *Set-free Markov State Model Building*, The Journal of Chemical Physics **146** (2017), no. 12. (Cited on page 68.)

[114] M. Weber, S. Kube, L. Walter, and P. Deuflhard, *Stable Computation of Probability Densities for Metastable Dynamical Systems*, Multiscale Modeling & Simulation **6** (2007), no. 2, 396–416. (Cited on page 41.)

[115] J. Woodcock, *A Difficult Balance – Pain Management, Drug Safety, and the FDA*, The New England Journal of Medicine **361** (2009), no. 22, 2105–2107. (Cited on page 7.)

[116] G. A. Worth, F. Nardi, and R. C. Wade, *Use of Multiple Molecular Dynamics Trajectories To Study Biomolecules in Solution: The YTGP Peptide*, The Journal of Physical Chemistry B **102** (1998), no. 32, 6260–6272. (Cited on page 25.)

[117] A. B. Yongye and K. Martínez-Mayorga, *Molecular Aspects of Opioid Receptors and Opioid Receptor Painkillers*, ch. 3, pp. 43–62, InTech, 2012. (Cited on page 12.)

[118] C. Zöllner and C. Stein, *Opioids*, Handbook of Experimental Pharmacology **177** (2007), 31–63. (Cited on page 7.)

Acknowledgements

When I started working at the Zuse Institute more than nine years ago, I was sure, that I would never do a doctor's degree. And now, I am here, writing the acknowledgements for my doctoral thesis. The biggest role in this development certainly played my group leader and supervisor Marcus Weber who believed in me and supported me, even when I told him the third time that I am pregnant. I would like to thank him for including me in his working group, for sharing knowledge, discussing results and handing out advices. Furthermore, I would like to thank all former and present members of our working group at ZIB that have accompanied me a short way, especially Konstantin "Max" Fackeldey for always cheering me up, Vedat Durmaz for introducing me to molecular simulation with GROMACS, Amira and VMD, Alexander Bujotzek for fruitful collaboration, and Marthe Solleder for helping me with the simulations at HLRN.

I am grateful to the *School of Analytical Sciences Adlershof (SALSA)*, not only for their financial support but also for informative lectures, lively discussions, social events, helpful workshops and for demanding interim reports – these were very helpful for writing the thesis in the end. Representative for the organizing team, I want to thank Katharina Schultens for her assistance in all organizational matters and her sympathy concerning compatibility of family and career.

Besides these two institutions, I also want to mention the Charité Universitätsmedizin Berlin and the Technische Universität Berlin which gave me a little insight into the "real" world outside of the computer. I would like to thank Prof. Dr. Stein from the Department of Anesthesiology and Critical Care Medicine for giving me the opportunity to present my results in his working group and for providing me with lots of literature to study, as well as Dr. Viola Spahn for explaining me the function of the μ-opioid receptor in simple words. Furthermore, I want to thank Prof. Dr. Hildebrandt and Dr. Jacek Kozuch from the Institute for Chemistry of the TU Berlin for introducing me to their experimental set-ups and illustrating chemical processes.

However, my deepest gratitude goes to my family. I thank my sister Stefanie and my parents Sabine and Harald Zech for being always there for me and for supporting me in all my decisions. Finally, I want to thank my husband Max Klimm and my three children Merle, Mailo and Malte for everything – you are the best in my life!

Zusammenfassung

In der vorliegenden Dissertation werden neu erarbeitete Methoden der Konformationsdynamik für die computergestützte Untersuchung des μ-Opioidrezeptors verwendet. Dieser Rezeptor ist ein transmembranes Protein, welches insbesondere für die Wahrnehmung und die Linderung von Schmerz wichtig ist. Eine Behandlung mit den momentan verfügbaren Schmerzmitteln ist jedoch mit verschiedenen, teilweise schwerwiegenden Nebenwirkungen verbunden. Daher ist es erstrebenswert, ein Schmerzmittel zu entwickeln, welches nur im entzündeten Gewebe wirkt und somit nebenwirkungsarm bzw. -frei ist. Vorausgegangene Untersuchungen des Rezeptors bezüglich der Wirkung von Schmerzmitteln zeigten bereits, dass Analgetika eine bessere Wirkung im entzündeten Gewebe erzielen, welches insbesondere durch einen niedrigen pH-Wert gekennzeichnet ist. In dieser Doktorarbeit werden mögliche Gründe für das unterschiedliche Verhalten des Rezeptors im gesunden und entzündeten Gewebe erörtert.

Im ersten Teil der Arbeit beschäftigen wir uns mit der Konformationsdynamik. Nach einer Einführung in die Theorie der molekularen Simulation werden neue Methoden vorgestellt, mit denen die Rechenleistung in der Konformationsdynamik stark verbessert werden kann. Zunächst wollen wir die hohe Dimensionalität des Konformationsraums durch Coarse Graining Methoden reduzieren. Die erste Methode clustert die Bewegungen von Partikeln nach bestimmten Kriterien, dabei handelt es sich um ein Coarse Graining der Zeit basierend auf Markov State Models. Die zweite Methode erzielt eine Beschreibung von relevanten Bewegungen durch wenige Deskriptoren anstatt des hochdimensionalen Zustandsvektors. Im nächsten Schritt stellen wir zwei neue Strategien für eine direkte Berechnung von statistischen Gewichten metastabiler Mengen vor, die auf der Schätzung von Differenzen der freien Energie basieren. Der erste Ansatz erfordert eine inverse Detailed Balance Bedingung, der zweite approximiert Entropiedifferenzen.

Im zweiten Teil erfolgt die computergestützte Analyse des μ-Opioidrezeptors unter Verwendung der vorgestellten Methoden. Die zuvor aufgestellten Hypothesen, warum Analgetika im entzündeten Gewebe besser wirken, sollen geprüft werden. Im ersten Abschnitt untersuchen wir die Coulomb-Interaktionen zwischen Rezeptor und Ligand und erkennen, dass diese im entzündeten Gewebe einen geringeren Einfluss haben als im gesunden Gewebe. Im zweiten Abschnitt analysieren wir die Konformationen des μ-Opioidrezeptors im gesunden und entzündeten Gewebe mit dem Resultat, dass es im entzündeten Gewebe eine klar präferierte Konformation gibt, wohingegen es im gesunden Gewebe mehrere ähnlich wahrscheinliche Konformationen gibt. Auch wenn diese theoretischen Berechnungen nicht praktische Untersuchungen ersetzen können, geben unsere Resultate Hinweise darauf, dass die bessere Wirksamkeit von Schmerzmitteln im entzündeten Gewebe unter anderem in der Konformationsänderung begründet liegt.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich alle Hilfsmittel und Hilfen angegeben habe, und dass ich die Arbeit auf dieser Grundlage selbstständig verfasst habe. Ich versichere, dass die Arbeit nicht schon einmal in einem früheren Promotionsverfahren eingereicht wurde.