

Dynamic Protonation of Titratable Groups in Biomolecules for Molecular Dynamics Simulations

Diplomarbeit

vorgelegt von

Florian Tegeler aus Osnabrück

Göttingen, den 30.05.2008

Department for Theoretical and Computational Biophysics at the Max
Planck Institute for Biophysical Chemistry, Göttingen



Abstract

Properties of biomolecules are often investigated using Molecular Dynamics (MD) simulations whereby atoms and molecules interact for a period of time, thus revealing the time dependent behavior of the system. Whereas the atom positions of the starting structures for such MD runs are usually derived from X-ray crystallography or NMR experiments, properties such as charges or bonds are specified in so called topologies fitted to the starting structure.

During classical MD simulations such a predefined topology is not altered during the complete simulation run. Inherently this requires the protonation states of all titratable groups in bio molecules, e.g. glutamic acid, histidine, lysin etc., to remain unchanged during the full simulation. In this thesis we apply and extend a novel method, so called λ -dynamics, to allow topology switching between different protonation states as a dynamical process. Hereby two Hamiltonians are defined in advance reflecting the protonated and unprotonated configuration. The newly introduced parameter λ is a switching parameter on a linear interpolation of those Hamiltonians treated as a virtual dynamical particle. The forces acting on λ are derived from the derivative of the free energy landscape $\frac{\delta G}{\delta \lambda}$. Hydration free energies for deprotonation are included by empirically altering the energy landscape in G . By accounting for the pH dependence of the hydration free energy in the energy landscape correction, constant pH simulations can be achieved. In an extended version of this method multiple titratable groups can be treated independently to allow molecular dynamics simulations of typical peptides and proteins.

Contents

1	Introduction	7
2	Theory and methods	11
2.1	Free energy difference	12
2.2	λ -dynamics	16
2.3	Single branch free energy calculations	17
2.4	Free energy modifications for a single titratable group	20
2.5	Generalization to multiple titratable groups	22
2.6	Including an additional barrier potential	23
2.7	Technical constraining of the sampling space	25
2.7.1	Entropic potential by circular projection	26
2.8	Heat bath considerations	28
2.9	Error estimation	30
3	Parameter analysis	31
3.1	The test system setup	31
3.2	Temperature and temperature coupling	32
3.3	Mass	32
3.4	Correlation between mass and temperature coupling	33
3.5	Starting λ dependency	34
3.6	Barrier height	36
3.7	Summary	36

4	Results 1 - Single amino acid	39
4.1	Parametrization system setup	39
4.1.1	Free energy	41
4.1.2	Constructing the experimental energy landscape	43
4.2	Glutamic acid titration curve	43
4.2.1	Statistics	45
5	Results 2 - peptide: oligoglutamic acid	51
5.1	Duplicated parametrization system	51
5.2	Test peptide (2 glutamic acids)	54
5.3	Test peptide (3 glutamic acids)	55
6	Summary	65
	Bibliography	69
7	Appendix	75

Chapter 1

Introduction

Most, if not all, biochemical processes on the molecular level are triggered and mediated by proteins. To understand the qualitative and quantitative properties of such processes, the structure, function and process dynamics of the biomolecules have to be investigated. With structural data, e.g. from NMR or X-ray crystallography experiments, an atomistic computer model of the biomolecules involved can be generated. The model system's time dependent behavior is then computed by allowing the atoms and molecules to interact for a period of time. Such *Molecular Dynamics* simulations (MD) have become more common and successful in the last years, not only due to increasing computational resources, but also due to new atomistic models. Although an increasing number of atoms can be modeled, and even combined Quantum Mechanical / Molecular Mechanical (QMMM) [LC03] [RSC05] methods are routinely used today, the simulation of larger systems at a fixed and arbitrarily chosen pH value comprises an active field of research. However, stability and function of proteins strongly depends on the local protonation states of titratable groups, e.g. glutamic acids or histidines, which is governed by the environmental pH ([MGGM⁺85], [Per78], [War79] and [BMP97]).

In standard MD simulations a single topology, describing a fixed construct of atoms and their interactions is used and protonation states are constant during an entire simulation. The pH is hereby only taken into account by generating a starting structure with the most probable protonation arrangement for a given pH. Biological function based on local environmental effects e.g. by closing a reactive site followed by a redistribution of protons, can therefore only be simulated inaccurately in classical MD simulations. It is, hence, necessary to extend the classical MD approach to allow simulations at a given pH, reproducing the

correct average protonation probabilities, while allowing proton reordering by dynamical topology changes.

Nevertheless, performing such constant pH simulations is a relatively new achievement. In 1994 Mertz and Pettitt [MP94] simulated acetic acid by creating a potential function interpolation between the protonated and the deprotonated state over a variable ξ . The general setup is similar to the Hamiltonian interpolation over a twofold Hamiltonian describing an acid's state in the protonated and the deprotonated form as used e.g. in free energy calculations (2.8). The Hamiltonian is extended to reflect the pH dependency by coupling the potentials for each end state to a chemical potential influenced by the pH. From the chosen chemical potential difference the equilibrium value for the interpolating variable ξ (or extent of reaction) is calculated. ξ is then compared with the experimental titration curve to calibrate the interpolation potential. By setting the ξ variable the correct protonation state is achieved over the simulation, effectively running at constant pH.

In 1997 Sham, Chu and Warshel [SCW97] used the method of protein dipoles Langevin dipoles (PDL) treating the protein relaxation in the microscopic framework of the linear response approximation. By the response from the local environment they derived the change in the protonation state. Thus, local effects are much better represented than in a continuum treatment of the electrostatic energies in a macromolecule.

More recent approaches can be generally divided into two main sections: The first set of approaches consists of techniques combining Molecular Dynamics (MD) and Monte Carlo (MC) simulations for sampling the protonation reaction coordinate. In the MD/MC simulations all atoms are calculated using MD and the protonation state represented by the corresponding charges and interaction terms is controlled via MC. The major differences between various constant pH implementations is the choice of solvent representation and the calculation of the free energy change (ΔG) with respect to the switch in protonation state. This ΔG is then used to make the Metropolis [MRR⁺53] decision, whether the new protonation state is accepted or rejected.

Bürigi et al. [BKvG02] use an explicit solvent model and calculate the free energy difference by a short (10ps) thermodynamic integration (TI) run at each MC step. Performing the TI is very costly in computational resource requirements and, moreover, short TI calculations are likely to encounter severe problems in more complex systems due to non-valid equilibrium assumptions. Another approach by Baptista [BMP97] [BTS02] uses static pK_a calculations

based on an implicit Poisson-Boltzmann (PB) model. Antosiewicz et al. [Ant08] and Dlugosz [DA04] extended this model further using continuum electrostatics for implicit solvent simulations. Mongan et al. [MCM04] [MC05] used MC as well, but with a Generalized Born (GB) [DCK04] implicit solvent. A common disadvantage to all MD/MC hybrid simulation methods is the high computational cost either to perform an additional TI for every MC step or to solve the PB or GB equations. Additionally a very fast change in protonation could lead to unforeseeable artifacts during the MD simulation or to a possible failure due to problems in the non-bonded interactions.

In contrast to MD/MC simulations, the techniques presented by Börjesson and Hünenberger [BH01] with explicit solvent and Lee et al. [LJI04] with implicit solvent, dynamically change the protonation state on a continuous titration coordinate in a way similar to the approach of Mertz and Pettitt [MP94]. They circumvent the problem of instantaneous protonation state switches and observe more stable simulations. The method of Lee et al. is based on the λ -dynamics [KI96] approach, whereby the coupling variable λ interpolating between the Hamiltonians describing the two end states is treated as a particle with a virtual mass. This particle is moving in the free energy landscape between the two end states at $\lambda = 0$ and $\lambda = 1$. The theory of λ -dynamics will be outlined in more detail in the theory and methods Chapter 2. Additionally in this continuous pH molecular dynamics method (CPHMD) Lee et al. utilize an artificial barrier to create an end state favorable environment in λ -space. The CPHMD method is extended by Khandogin et al. [KB05] to cope with proton tautomerism as can be found e.g. in histidines by using multi-dimensional λ -variables as virtual particles. This approach was described to accurately model experimental titration curves for simple amino-acids and is therefore a justified candidate for a transfer to the widely used and highly optimized GROMACS [BvdSvD95], [HKvdSL08], [LHvdS01], [SLH⁺05] Molecular Dynamics modeling package.

Nearly all of the previously proposed methods use implicit solvent models which, provide an elegant interface to model external environmental influences on the relevant titratable groups. The implementation of constant pH simulations including explicit solvent based on λ -dynamics requires nontrivial changes in the MD algorithms.

This thesis presents an extended λ -dynamics approach to accurately perform constant pH simulations with explicit solvent of systems including multiple titratable groups. The method is then applied to small test systems consisting

of one to three titratable groups to simulate titration curve experiments.

Chapter 2

Theory and methods

To increase the simulation length and allow larger system sizes including more atoms in molecular dynamics simulations, some approximations, also relevant for later constant pH modifications, are made. The most fundamental one is the classical treatment of atoms, which allows the description of atomic movement by solving Newton's equations of motion. Instead of calculating the potential at each step during the simulation, a simplified potential V is used:

$$m_i \frac{\delta^2 \mathbf{r}_i}{\delta t^2} = \mathbf{F}_i \quad (2.1)$$

$$\mathbf{F}_i = -\frac{\delta V}{\delta \mathbf{r}_i}, \quad (2.2)$$

whereby m_i is the mass, \mathbf{r}_i the position and \mathbf{F}_i the force on atom i . These equations are solved simultaneously in small time steps ($0.5fs$ to $2fs$), which are smaller than the fastest fluctuations in the system, using the Leap-Frog [HGE74] algorithm,

$$\mathbf{v} \left(t + \frac{\Delta t}{2} \right) = \mathbf{v} \left(t - \frac{\Delta t}{2} \right) + \frac{\mathbf{F}(t)}{m} \Delta t \quad (2.3)$$

$$\mathbf{r} \left(t + \frac{\Delta t}{2} \right) = \mathbf{r}(t) + \mathbf{v} \left(t + \frac{\Delta t}{2} \right) \Delta t. \quad (2.4)$$

The classical approach implicitly requires additional assumptions: In the *Born-Oppenheimer approximation* the motion of the electrons can be described independently from the nuclei due to several orders of magnitude difference in mass and velocity. Therefore, the Schrödinger equation as the most general description separates into a time independent part for the electrons and the previously mentioned time dependent potential function V for nuclei motion. In all force fields, V is a fit to the quantum mechanical ground state energy.

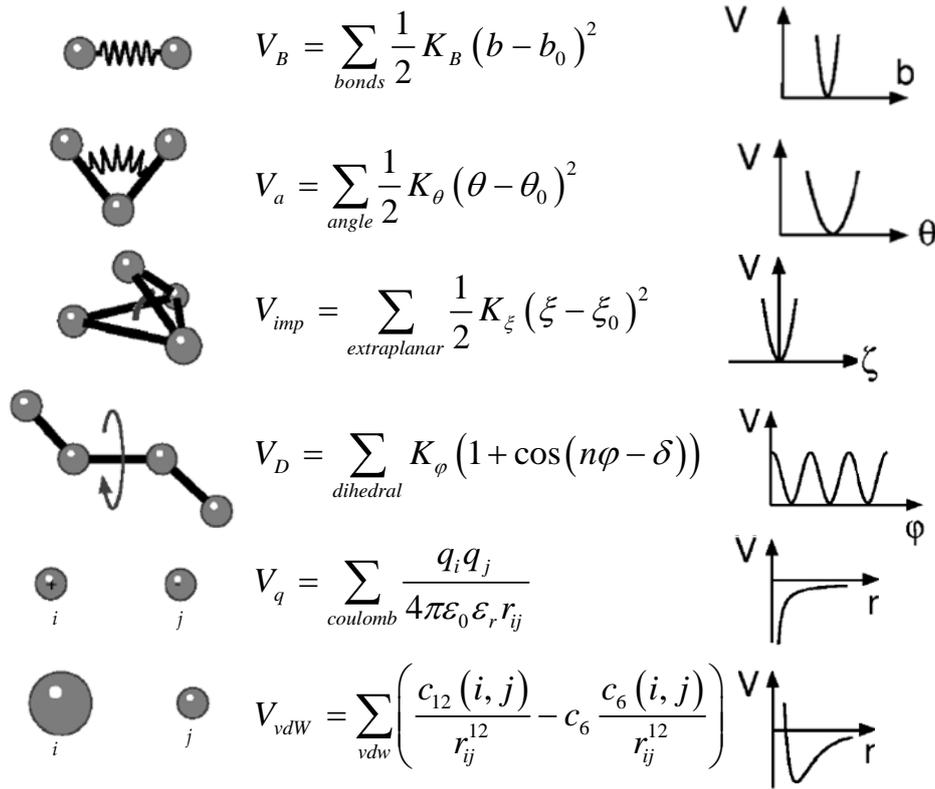


Figure 2.1: Energy terms describing the force field potential. The left column depicts the atomistic interaction and angular constructions. It is followed by the potential description for each interaction and the right column illustrates the term graphically.

Figure 2.1 illustrates the inter- and intramolecular energies including separated terms for bond stretching potentials, bond angles, extraplanar angles, dihedral, Coulomb pair interaction and van der Waals interaction as force field's building blocks.

2.1 Free energy difference

For constant pH simulations the energy difference ΔG between the protonated and deprotonated state enables one to calculate e.g. the pK_a shift of an amino acid transferred from water to a protein. Yet, calculating free energies ΔG from MD simulations using force fields is not straightforward. This is mainly due to weaknesses in the force field model not describing the quantum mechanical effects in bond forming and electron distribution. The harmonic approximation for bonded interactions as shown in Figure 2.1 is not very accurate for vanishing

or newly created particles and does not describe stretched bonds. But most problematic is the modeling of the bonds as harmonic springs with zero energy at the equilibrium length. The inner energies of the molecules influenced by the electrons forming bonds between the atoms are therefore not correctly modeled. By this simplified treatment an offset, different for each molecule, is introduced making the calculation of absolute free energies inapplicable.

Considering a simple proton transfer from an hydronium to an ammonium in vacuum is modeled, the reaction



is modeled with large separation (50nm distance) between the hydronium and the NH_3 so that no interaction influences the free energy analysis. By quantum mechanical calculations performed with the GAUSSIAN [F^TS⁺] software, the enthalpy difference between the NH_3 and the NH_4^+ is calculated to be $\Delta G = -879,7\text{kJ/mol}$ which is in the range of the hydration free energy of a proton of around $\approx 260\text{kcal/mol} = 1088\text{kJ/mol}$ [TTB⁺98]. However, in the molecular mechanical simulation where all quantum mechanical effects concerning electrons are ignored a free energy difference of $\Delta G = 0.0013\text{kJ/mol}$, which is small compared to the inaccuracy of free energy calculation methods, is calculated in such simulations as Figure 2.2 illustrates.

The complete change of inner energy U cannot be calculated as force fields ignore electron influences. The only contribution to the free energy in simulations is the non-bonded interactions e.g. from surrounding water molecules.

Due to this offset problem of inner energy changes in the force fields, free energies are typically calculated as *relative differences* via a thermodynamical cycle. Figure 2.3 illustrates a typical cycle describing the deprotonation of an acid in water and in the protein. This thermodynamical cycle yields meaningful energy differences as the force field introduces the same offset error to all required branches of the cycle, thus canceling each other out. Following

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3 \quad (2.6)$$

either the calculation of ΔG_1 and ΔG_2 or the calculation of ΔG_3 and ΔG_4 is required. $\Delta\Delta G$ expresses the acid's pK_a shift due to the protein environment. This method of calculating pK_a shifts was first proposed by Warshel et al. in [WSK85].

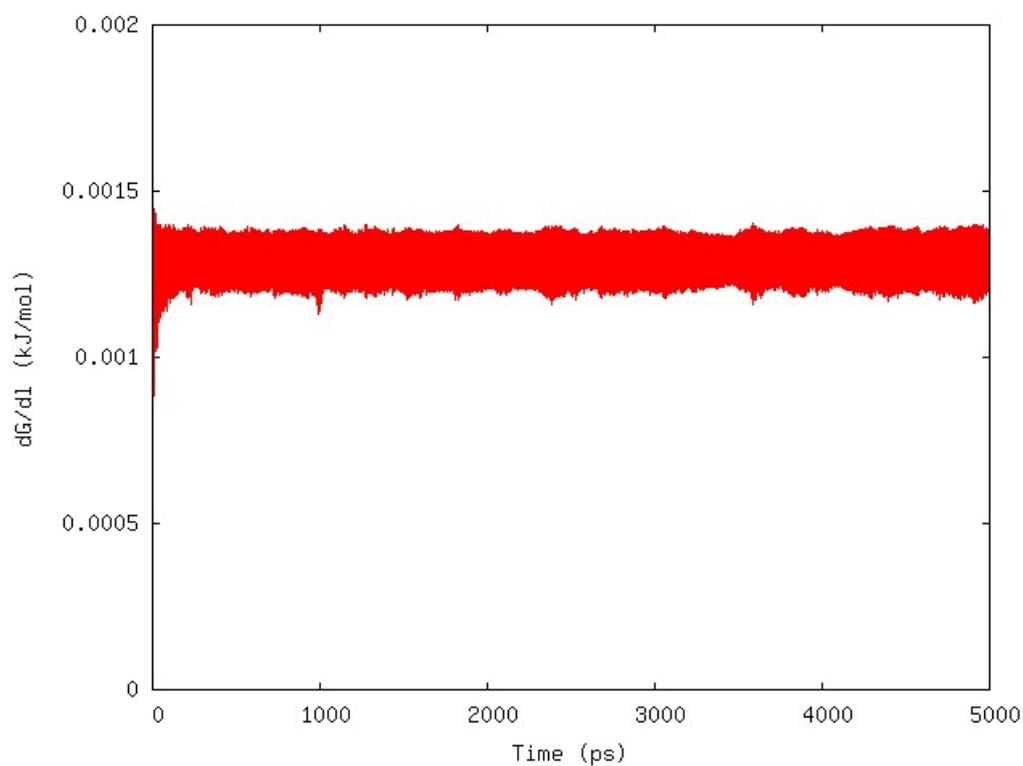


Figure 2.2: $\delta G/\delta\lambda$ for a free energy perturbation of $5000ps$, whereby λ is continuously switched from $\lambda = 0$ to $\lambda = 1$. The λ -dependence expressed in $\delta G/\delta\lambda$ is barely recognizable and integrates to a ΔG smaller than the error of the FEP method.

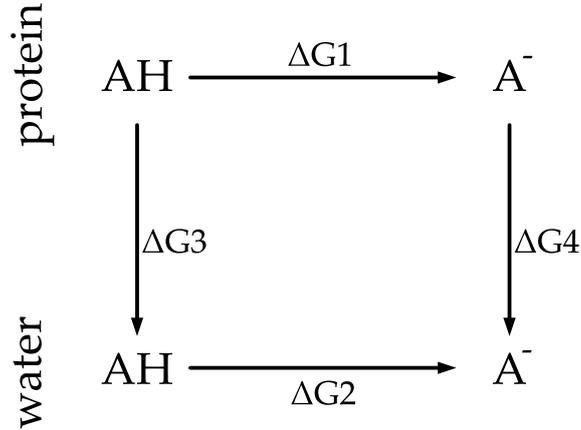


Figure 2.3: Thermodynamical cycle to calculate pK_a shifts. An amino acid is placed in water and in protein environment. By evaluating $\Delta G1$ and $\Delta G2$, the relative difference between the protonated and the deprotonated state can be calculated.

To calculate free energy differences and to evaluate the branches of the thermodynamical cycle in MD simulations either Umbrella Sampling [TV77] or Free Energy Perturbation (FEP) is used. FEP is based on the *Zwanzig* [Zwa54] equation

$$\Delta G(A \rightarrow B) = G_B - G_A = k_B T \ln \left\langle \exp \left\{ -\frac{E_B - E_A}{k_B T} \right\} \right\rangle_A \quad (2.7)$$

which gives the free energy difference between state A and state B .

To realize the perturbation, an interpolated Hamiltonian is used to describe the system. The Hamiltonian of state A is added to the Hamiltonian of state B using a weighting parameter λ , whereby the system is in state A for $\lambda = 0$ and in state B for $\lambda = 1$. The simplest interpolation is linear with $\lambda = 0 \rightarrow 1$:

$$H(\lambda) = H_A \cdot (1 - \lambda) + H_B \cdot \lambda, \quad (2.8)$$

although also other pathways or different functional forms for various force field are possible.

By adding up infinitively small perturbations the energy difference can then be calculated by integrating over λ , such as:

$$\Delta G = G_B - G_A = \int_0^1 \frac{\delta H}{\delta \lambda} \delta \lambda. \quad (2.9)$$

It is important to note that the interpolation pathway is completely arbitrary and has no physical meaning except for the end states. The difference in states could be e.g. a proton distribution, an additional group, an ion created in water etc. A typical problem with the linear Hamiltonian interpolation for charge transfer is a strong barrier potential between both states raised by the artificial interpolation pathway by vanishing one charge and creating a new charge at another position. This is not a realistic representation of the real quantum-mechanical behavior whereby the proton is following a path with different ‘‘hops’’ in between the end states.

2.2 λ -dynamics

Another approach using the Hamiltonian interpolation 2.8 was proposed by Kong et al. [KI96]: Instead of controlling λ and integrating over the free energy landscape’s derivative, λ is used as a virtual particle defined by a fictitious constant mass m unequal zero, a position coordinate λ in the interval $[0 : 1]$ and a velocity $v = \delta\lambda/\delta t$ acting in λ space. This is expressed by an extended Hamiltonian:

$$H_{extended}(\lambda_i) = H_{part}(\lambda_i) + \sum_{i=1}^n \frac{m_i}{2} \dot{\lambda}_i^2 + U^*(\lambda_i) \quad (2.10)$$

A λ -dependent potential term $U^*(\lambda_i)$ will serve as an umbrella or biasing potential to limit the range of λ_i (see [KI96]). The $H_{part}(\lambda_i)$ represents a partitioned Hamiltonian of the general form

$$H_{part}(\lambda_i) = H_R(\lambda_i) + H_P(\lambda_i) + H_{Env} \quad (2.11)$$

with H_R and H_P being the Hamiltonians describing the reactant and product state of the perturbed atoms.

In contrast to normal free energy perturbation, the quantity $\frac{\delta H}{\delta \lambda}$ is not used to compute the free energy difference ΔG , but is used as a driving force for the virtual λ -particle. The forces acting on the λ -particle are consequently defined as:

$$F_{\lambda_i} = -\frac{1}{m_i} \frac{\delta G}{\delta \lambda_i} \quad (2.12)$$

The λ -coordinate is not a reaction coordinate as it does not correspond to a physical mechanism. However, only the end points represent thermodynamical states thus the actual pathway over the interpolated states is irrelevant.

This λ -dynamics approach allows to switch dynamically between two pre-defined topologies. By applying λ -dynamics on a topology set e.g. representing an unprotonated and a protonated amino acid, the system can dynamically adapt to the environmentally favored protonation state. The λ -particle governed by the free energy landscape of the system would feel forces directing it to the energetically optimal state.

2.3 Single branch free energy calculations

The virtual λ particle moves on the free energy landscape between two systems and therefore requires a correct description of the free energy differences between these systems. But, as outlined before, free energy calculations based on MD simulations using force fields can only yield *free energy differences* calculating a thermodynamical cycle where the branches compared encounter the same offset effects.

To circumvent this offset problem, empirical data is used as a replacement for the second branch in the thermodynamical cycle effectively compensating the incorrect force field description of electrons and corresponding minimal bond energies. After applying these corrections to the force field, the λ -dynamics will reproduce the correct statistical weights for the product and reactant states. This ratio between protonated and deprotonated acid is pH dependent which is included in the corrections as well.

For the theoretical analysis the thermodynamical cycle depicted in Figure 2.4 is used: If one is interested in an amino acid's pK_a shift expressed by a difference in free energy τ , one usually performs a free energy perturbation to derive ΔG_{prot} and a second FEP to derive ΔG_{ref} . This allows the computation of the protein environment's influence τ on the ratio of protonated and deprotonated acid (the pK_a shift) by calculating $\tau = \Delta \Delta G = \Delta G_{prot} - \Delta G_{ref}$.

The calculation is valid, because all other branches are canceling out each other except for τ . This can be easily seen following the red arrows traversing the

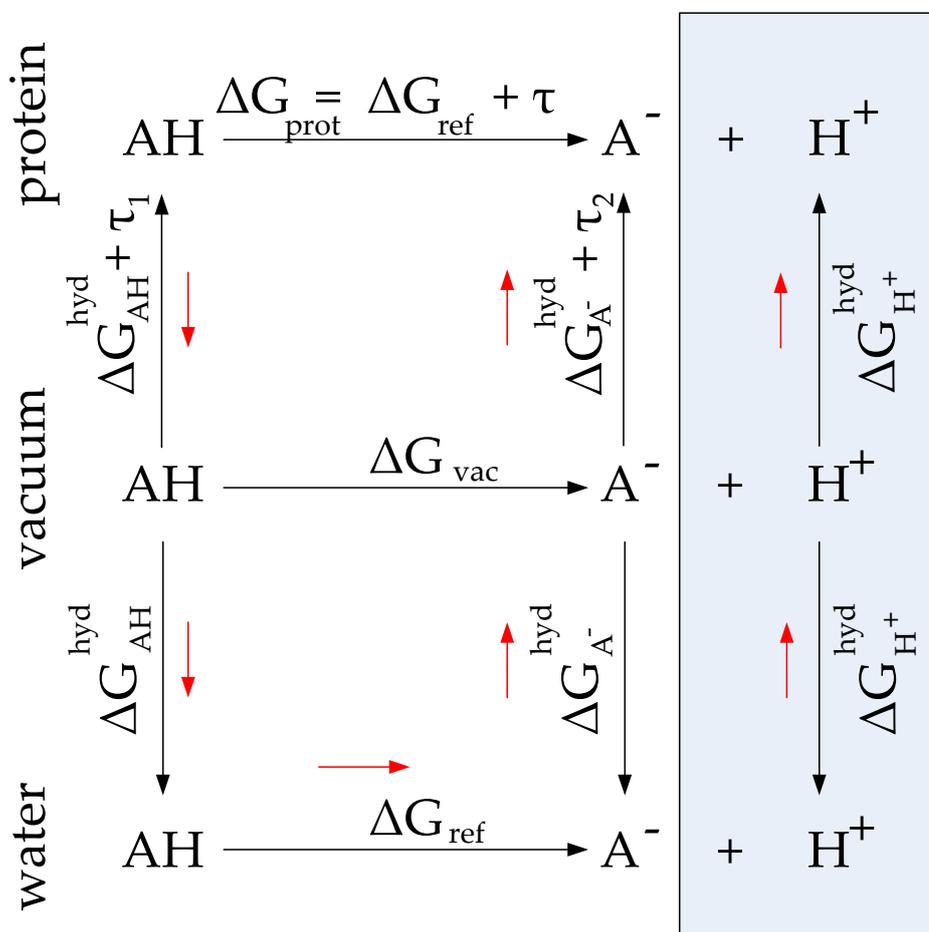


Figure 2.4: Fundamental thermodynamical cycle for λ -dynamics calculations at constant pH in the ΔG_{prot} branch. It is related to the classical cycle depicted in Figure 2.3 but extending it for the vacuum description. Walking through the cycle following the red arrows also yields ΔG_{prot} and as will be shown, directly correspond to ΔG_{ref} plus the energy difference τ from the environment. A key concept for constant pH simulations is the separation of the hydration free energy $\Delta G_{H^+}^{hyd}$ (emphasized in blue) of the proton from the remaining contributions during deprotonation. This pH dependent hydration free energy ($\Delta G_{H^+}^{hyd}$) is altered to allow simulations at different pH.

thermodynamical cycle (Figure 2.4).

1. The first step is the transfer of the protonated acid (AH) from the protein environment to the vacuum state. The energy difference between these states is $-\left(\Delta G_{AH}^{hyd} + \tau_1\right)$ representing the energy that was required to first hydrate the protonated acid and τ_1 to embed it in the protein.
2. In the next step the protonated acid is solvated in water which requires an energy of ΔG_{AH}^{hyd} .
3. In the water environment the acid deprotonates yielding an unprotonated acid (A^-) and a solvated hydrogen (H^+).
4. Transferring the deprotonated acid back into vacuum and dehydrating the proton includes two energy contributions: $\Delta G_{A^-}^{hyd}$ for the acid and $-\left(\Delta G_{H^+}^{hyd}\right)$ for the hydrogen.
5. The final step is the transfer of the deprotonated acid from vacuum back to the protein environment requiring an energy of $\Delta G_{A^-}^{hyd} + \tau_2$. τ_2 describes the influence of the protein compared to the water reference system. The energy required to solvate the hydrogen $\Delta G_{H^+}^{hyd}$ is assumed to not depend on any protein environment but on the pH what is true for dilute solutions. Only the processes highlighted in Figure 2.4 by the blue box contribute to the pH dependency and are therefore split off from the rest of the system.

The following equation expresses the calculation of ΔG_{prot} with the discussed thermodynamical cycle Figure 2.4:

$$\begin{aligned} \Delta G_{prot} &= -\left(\Delta G_{AH}^{hyd} + \tau_1\right) + \Delta G_{AH}^{hyd} + \Delta G_{ref} - \\ &\Delta G_{A^-}^{hyd} + \left(-\Delta G_{H^+}^{hyd}\right) + \left(\Delta G_{A^-}^{hyd} + \tau_2\right) + \Delta G_{H^+}^{hyd} \end{aligned} \quad (2.13)$$

with $\tau = \tau_2 - \tau_1$:

$$\Delta G_{prot} = -\tau_1 + \Delta G_{ref} + \tau_2 = \Delta G_{ref} + \tau \quad (2.14)$$

In the simulation no hydrogen solvating in the water box is modeled but the pH dependent effect of $\Delta G_{H^+}^{hyd}$ on the free energy difference $\Delta G_{A^-}^{hyd} + \tau_2 + \Delta G_{H^+}^{hyd}$ is accounted for by an analytical correction function.

Although τ_1 and τ_2 are not known in advance a free λ -dynamics simulation can be performed to evaluate ΔG_{prot} . The protein environmental effects τ_1 and τ_2 are influencing the deprotonation of the acid by $\tau = \tau_2 - \tau_1$ compared to ΔG_{ref} . The force field offset errors originating from neglecting the electrons in the bond descriptions are identically introduced in both branches of the simulation: In calculating ΔG_{ref} and in calculating ΔG_{prot} .

2.4 Free energy modifications for a single titratable group

Correcting the free energy landscape by empirical data requires a set of additional parameters to reproduce the correct ΔG_{prot} from single branch simulations. By deriving these parameters and correction functions from the calculation of ΔG_{ref} the cycle can be closed and the correct free energy difference between the end states depending on a given pH is computed.

The first step is the separation of the hydrogen's solvation free energy $\Delta G_{H^+}^{hyd}$ from the rest of the deprotonation process. It is assumed, that only this term has a dependency of pH. This can be obtained from the experimental standard reaction free energy describing the solvation of the hydrogen and the deprotonation free energy of the amino acid expressed in the pK_a :

$$\Delta G_{exp} = k_B T (pK_a - pH) \ln(10) \quad (2.15)$$

Thus, the proton disappears from the box during the deprotonation step and is accounted for by computing the hydration free energy analytically via a function $\kappa(\lambda, pH)$ which can be used to model the pH influence on the system. Because only the end state free energies at $\lambda = 0$ and $\lambda = 1$ are physically relevant, a simple linear function can be chosen:

$$\kappa(\lambda, pH) = c(pH)\lambda \quad (2.16)$$

whereby $c(pH)$ is a constant depending on a given pH value. $\kappa(\lambda, pH)$ can be calculated from equation 2.15 which can be expressed as

$$\Delta G_{exp} = \kappa(\lambda = 1, pH) - \kappa(\lambda = 0, pH) \quad (2.17)$$

$$\Rightarrow \kappa(\lambda, pH) = \ln(10) \cdot k_B T (pK_a - pH) \cdot \lambda \quad (2.18)$$

The second step is the treatment of the force field offset errors by a function $f(\lambda)$ reproducing the correct free energy landscape expressed by

$$\Delta G_{exp} = \Delta G_{ref} + \kappa(\lambda, pH) + f(\lambda). \quad (2.19)$$

Physically relevant are only the end states and $f(\lambda)$ must correct for the difference in free energy ΔG_{ref} . Additionally, a flat energy landscape would be most favorable by introducing any barriers at unknown positions affecting transitions rates between $\lambda = 0$ and $\lambda = 1$ states. To generate such a flat energy landscape not only the energy difference ΔG_{ref} must be known but the functional form of the λ dependence as well.

With *Linear Response Theory* [CP88] [HMV78] Simonson outlines [Sim02] that dielectric response of a solvated protein to a perturbed charge is linear which yields to a parabolic free energy functional form. The response is related to the equilibrium fluctuations (which are assumed to be gaussian) of the unperturbed system. A perturbing charge density λp_p contributes therefore to the Hamiltonian by

$$\Delta H = -\lambda \mathbf{f} \cdot \mathbf{P} \quad (2.20)$$

whereby \mathbf{f} is the field raised by p_p and $\mathbf{E} = -4\pi\mathbf{P}$ describes the field due to the remaining charge density \mathbf{P} . With $\delta\mathbf{P} = \mathbf{P} - \langle\mathbf{P}\rangle_0$ and $\langle\delta\mathbf{P}\rangle_\lambda$ as the Boltzmann average over the perturbed system, the response can be explained as the mean microscopic density of polarization charge induced by the perturbing field $\lambda\mathbf{f}$. Overall the free energy derivative is linear with respect to $\lambda\mathbf{f}$ [Sim02]:

$$\frac{\delta G}{\delta \lambda} = \left\langle \frac{\delta \Delta H}{\delta \lambda} \right\rangle_\lambda = -\mathbf{f} \cdot \langle \mathbf{P} \rangle_\lambda \quad (2.21)$$

$$= -\mathbf{f} \langle \mathbf{P} \rangle_0 - \lambda \mathbf{f} \cdot a \mathbf{f} (1 + O(\beta \Delta H)) \quad (2.22)$$

with $\beta = 1/k_B T$ and $O(..)$ representing quantities of first order or more in $\beta \Delta H$. If the fluctuations are gaussian, which can be assumed for most systems [Sim02] higher orders are zero and the free energy derivative is linear. The expression a is a dielectric susceptibility operator not depending on λ .

With a linear derivative the free energy is of parabolic form. Influences of the environment or systems of higher complexity will create non-parabolic

free energy landscapes which can be nevertheless handled by the parabolic compensation due to the relevance of the end states only.

The function $f(\lambda)$ is in the parabolic compensation approach derived from a reference thermodynamic integration run resulting in a linear regression for $\delta G/\delta\lambda$, e.g. $\delta G_{ref}/\delta\lambda = a + b \cdot \lambda$.

To make the energy landscape plain and compensate for the force field artifacts the correction function $f(\lambda)$ is chosen as:

$$\frac{\delta f}{\delta\lambda} = a_1 + b_1\lambda = -a + (-b)\lambda \quad (2.23)$$

$$\Rightarrow f(\lambda) = -a_1\lambda - \frac{1}{2}b_1\lambda^2 \quad (2.24)$$

so that $a_1 = -a$ and $b_1 = -b$. Without the $\kappa(\lambda, pH)$ function $\Delta G_{exp} = \Delta G_{ref} + f(\lambda = 1)$ results now into a plain energy landscape which would correspond to a $pK_a = pH$ simulation reflected by $\kappa(\lambda, pH = pK_a) = 0 \forall \lambda$.

For the simulation the new $\delta G_{prot}^{correct}/\delta\lambda$ is calculated as

$$\frac{\delta G_{prot}^{correct}}{\delta\lambda} = \frac{\delta G_{prot}}{\delta\lambda} + \frac{\delta\kappa}{\delta\lambda} + \frac{\delta f}{\delta\lambda} \quad (2.25)$$

$$\frac{\delta G_{prot}^{correct}}{\delta\lambda} = \frac{\delta G_{prot}}{\delta\lambda} + \ln(10) \cdot k_B T (pK_a - pH) + a_1 + b_1\lambda \quad (2.26)$$

By the experimentally determined pK_a value the $\kappa(\lambda, pH)$ function represents - together with the force field artifact compensating function $f(\lambda)$ - the second branch of the thermodynamical cycle (ΔG_{ref} in Figure 2.4).

2.5 Generalization to multiple titratable groups

Extending the single λ -dynamics approach to multiple (n) λ -coordinates requires independent interpolation for each λ -coordinate. This is expressed with an extended Hamiltonian similar to equation 2.10:

$$H_{ext} = \sum_{i=1}^n ((1 - \lambda_i)H_{Reactant}^i + \lambda_i H_{Product}^i) + \sum_{i=1}^n \frac{m_i}{2} \lambda_i^2 + H_{env} \quad (2.27)$$

whereby $H_{Reactant}^i$ ($H_{Product}^i$) is the Hamiltonian describing the reactant (product) groups for the λ -coordinate i . The environmental Hamiltonian H_{env} is

split off. The same notation is implicitly used before in Equation 2.8 because the linear interpolation in a single λ -coordinate reads as follows:

$$H_{ext} = (1 - \lambda)H_A + \lambda H_B \quad (2.28)$$

$$= ((1 - \lambda)(H_{Reactant} + H_{env}) + \lambda(H_{Product} + H_{env})) \quad (2.29)$$

$$= (1 - \lambda)H_{Reactant} + \lambda H_{Product} + H_{env} \quad (2.30)$$

With the linear Hamiltonian interpolation also the Coulomb interaction between two atoms i and j of which charge varies with λ_i and λ_j is linear:

$$V_c = \frac{f}{\epsilon_r r_{ij}} [((1 - \lambda_i)q_i^A + \lambda_i q_i^B) ((1 - \lambda_j)q_j^A + \lambda_j q_j^B)] \quad (2.31)$$

Coulomb forces are sufficient to describe interactions in our approach, because we do not relate any van der Waals parameter to the variable proton in the titratable group. The forces driving the system along λ_i and λ_j are:

$$-\frac{\delta V}{\delta \lambda_i} \Big|_{\lambda_j} = -\frac{f}{\epsilon_r r_{ij}} \left[(q_i^B - q_i^A) ((1 - \lambda_j)q_j^A + \lambda_j q_j^B) \Big|_{\lambda_j} \right] \quad (2.32)$$

$$-\frac{\delta V}{\delta \lambda_j} \Big|_{\lambda_i} = -\frac{f}{\epsilon_r r_{ij}} \left[(q_j^B - q_j^A) ((1 - \lambda_i)q_i^A + \lambda_i q_i^B) \Big|_{\lambda_j} \right] \quad (2.33)$$

with $f = \frac{1}{4\pi\epsilon_0}$. In the simulation the forces on each λ are computed consecutively while keeping the remaining λ values constant. This treatment as constant values allows a straightforward implementation with low computational complexity.

2.6 Including an additional barrier potential

An additional barrier potential $U(\lambda)$ of height h which optimizes the transitions between the states at $\lambda = 0$ and $\lambda = 1$ and reduces the times the λ -particle stays in between the end states is added. The Hamiltonian can be extended again to

$$H_{ext} = \sum_{i=1}^n ((1 - \lambda_i)H_{Reactant}^i + \lambda_i H_{Product}^i) + \sum_{i=1}^n \frac{m_i}{2} \dot{\lambda}_i^2 + \sum_{i=1}^n U(\lambda_i) + H_{env}. \quad (2.34)$$

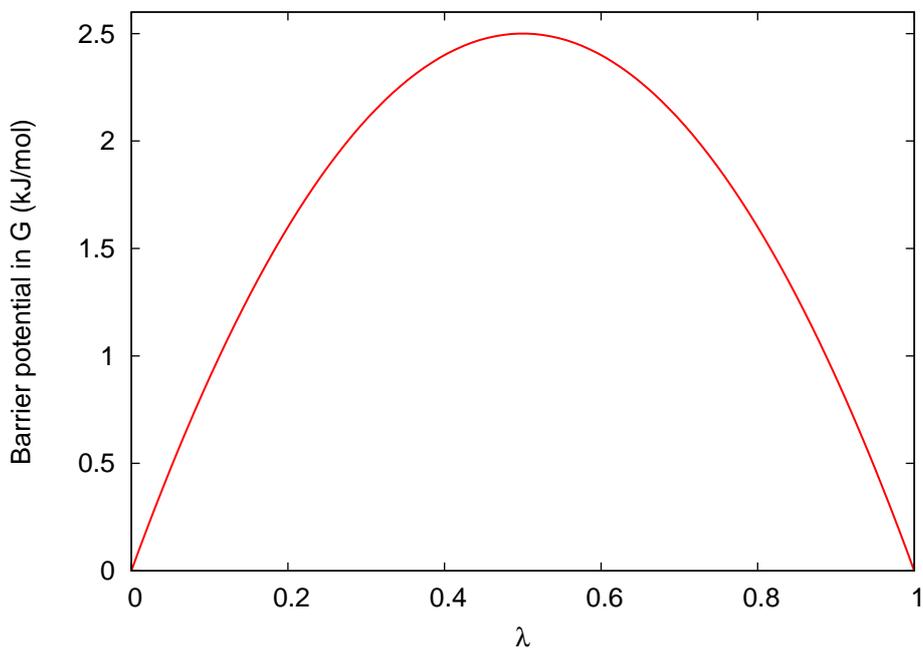


Figure 2.5: Barrier potential of 2.5 kJ/mol. The barrier is applied to the free energy landscape and not altering the thermodynamically relevant end states.

The barrier should not effect the end states and should act as a simple linear function in the ΔG space so it is defined as:

$$U(\lambda) = -\frac{1}{2}c \cdot \lambda + \frac{1}{2}c \cdot \lambda^2 \quad (2.35)$$

With

$$c = \frac{2h}{\lambda^2 - \lambda} \quad (2.36)$$

follows for the barrier height at $\lambda = 0.5$

$$c = -8 \cdot h \quad (2.37)$$

Therefore the barrier is expressed by:

$$U(\lambda) = 4h \cdot \lambda - 4h \cdot \lambda^2 \quad (2.38)$$

$$\Rightarrow \frac{\delta U}{\delta \lambda} = 4h - 8h\lambda \quad (2.39)$$

as illustrated in Figure 2.5 for $h = 2.5 \text{ kJ/mol}$.

The final calculation for the $\Delta G_{prot}^{correct}$ is done by:

$$\frac{\delta G_{prot}^{correct}}{\delta \lambda} = \frac{\delta G_{prot}}{\delta \lambda} + \ln(10) \cdot k_B T (pK_a - pH) + a_1 + b_1 \lambda + 4h - 8h\lambda \quad (2.40)$$

This approach can even be extended to an **dynamical barrier height adaptation**. Hereby the last n transitions are observed and the average time τ between the transitions is calculated. If τ is larger than a previously given control variable ξ , the barrier is lowered to optimize sampling and allow more transitions. If τ is smaller than ξ the barrier is increased again to lower rate of transitions. This method could be used to achieve realistic transition rates if the average time between transitions ξ is much smaller than the total simulation time. But even if this is not the case, the adaptive barrier potential can optimize sampling by allowing the λ -particle to move away from an end state where it was hold by a too strong barrier potential.

2.7 Technical constraining of the sampling space

Performing extended λ -dynamics simulations requires additional constraintment of the space λ can sample. The following three requirements are crucial for a physically correct λ -dynamics run:

1. The transition rates should be under control
2. Interpolated states describing unphysical configurations e.g. at $\lambda = 0.5$ should be minimized
3. The λ -particle should be constrained to the interval $[0 : 1]$ and do not be allowed to sample energetically favored but unphysical states e.g. at $\lambda = 4$.

While items 1 and 2 can be solved using an additional barrier of appropriate height, item 3 requires an additional solution. A typical problematic scenario could be the following: The system consist of one protonated glutamic acid surrounded by a high amount of NA^+ ions and the topologies are distinct only in the protonation state of the acid. In reality the glutamic acid would quickly deprotonate and the total charge in the system would not change significantly. With an unconstrained λ the glutamic acid could provide a negative charge large enough to compensate for all positive ions in the system.

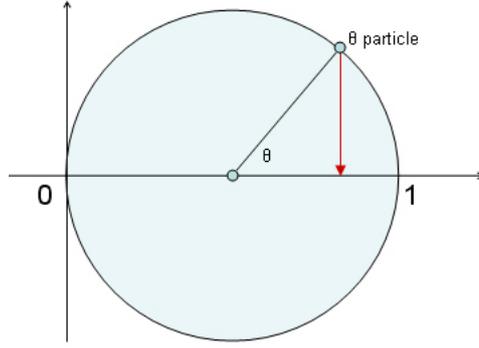


Figure 2.6: θ -dynamics as constraint solution. By projecting the forces $-\delta V/\delta\lambda$ to θ -space the MD is actually performed on the circle and λ is calculated via $\lambda = r \cos(\theta) + 0.5$ by projecting it back.

To circumvent moving over the defined end states and sampling of unphysical states, λ is re-defined as a circular projection from a newly introduced variable θ . λ is expressed by $\lambda = r \cos(\theta) + 0.5$ as shown in Figure 2.6. By this modification the actual MD simulation is performed in *theta*-space with θ being the new dynamical variable. λ is reduced to the projection of θ to the Hamiltonian interpolation axis.

The forces acting on λ have to be projected to the circle as well. The new Leap-Frog algorithm (see 2.3) reads as

$$F_\theta = -\frac{\delta V}{\delta\theta} = -\frac{\delta V}{\delta\lambda} \frac{\delta\lambda}{\delta\theta} = +\frac{\delta V}{\delta\lambda} r \sin(\theta) \quad (2.41)$$

$$v_\theta \left(t + \frac{\Delta t}{2} \right) = v_\theta \left(t - \frac{\Delta t}{2} \right) + \frac{F_\theta(t)}{m_\theta} \Delta t \quad (2.42)$$

$$\theta \left(t + \frac{\Delta t}{2} \right) = \theta(t) + v_\theta \left(t + \frac{\Delta t}{2} \right) \Delta t. \quad (2.43)$$

The final λ update reads as

$$\lambda(t + \Delta t) = r \cos(\theta(t + \Delta t)) + 0.5. \quad (2.44)$$

2.7.1 Entropic potential by circular projection

By using such a circular projection while calculating and updating λ , an additional underlying entropic term is introduced which can be calculated as follows:

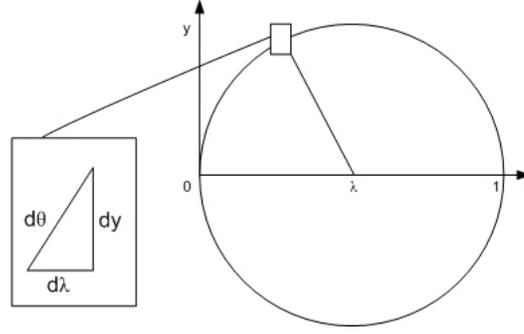


Figure 2.7: Calculation of the state density by calculating $d\theta$ in an element $d\lambda dy$ of the circle.

It is assumed that the particle moving on the circular space is influenced by an external potential V e.g. originating from the particle's environment in the simulation. Hereby it is important to note that holds $V(\theta) = V(-\theta)$ which reduces the problem to calculating the entropic effects of a half-circle.

The entropic contribution originates from the different number of states separated by $d\theta$ in a fragment $d\lambda$ depending on the position as illustrated in Figure 2.7. This could also be expressed as the density in a fragment $d\lambda$. At $\lambda = 0$ and $\lambda = 1$ the density reaches its maximum while $\lambda = 0.5$ is least occupied. States around the endpoints $\lambda = 0$ and $\lambda = 1$ are therefore entropically favored. The density of states g_θ can be described with the Boltzmann inverse [Sop96] [RPMP03] by

$$g_\theta d\theta = \exp(-\beta V(\theta)) d\theta \quad (2.45)$$

with $\beta = 1/k_B T$ as the inverse temperature. From the definition of λ the expression for $d\theta$ can be derived:

$$\lambda = \frac{1}{2} + \frac{1}{2} \cos(\theta) \quad (2.46)$$

$$\theta = \arccos(2\lambda - 1) \quad (2.47)$$

$$\Rightarrow d\theta = -\frac{d\lambda}{\sqrt{\lambda(1-\lambda)}} \quad (2.48)$$

To calculate the free energy contribution $F(\lambda) = -k_B T \ln(g_\lambda)$, the average distribution function g_λ is required:

$$g_\lambda d\lambda = g_{\theta(\lambda)} \frac{d\theta}{d\lambda} d\lambda = -\exp(-\beta V(\arccos(2\lambda - 1))) \frac{d\lambda}{\sqrt{\lambda(1-\lambda)}} \quad (2.49)$$

$$\Rightarrow F(\lambda) = -k_B T \ln(g_\lambda) = -k_B T \ln \left(\frac{\exp(-\beta V(\arccos(2\lambda - 1)))}{\sqrt{\lambda(1-\lambda)}} \right) \quad (2.50)$$

$$= -k_B T \left[-\beta V(\arccos(2\lambda - 1)) + \ln(\sqrt{\lambda(1-\lambda)}) \right] \quad (2.51)$$

$$= -k_B T \left[-\beta V(\arccos(2\lambda - 1)) + \frac{1}{2} \ln(\lambda(1-\lambda)) \right] \quad (2.52)$$

If there is even no external potential ($V(\theta) = 0$), the free energy surface influencing the λ -particle is given with respect to the following entropic term:

$$F(\lambda) = -k_B T \frac{1}{2} \ln(\lambda(1-\lambda)) \quad (2.53)$$

Figure 2.8 is showing a plot of the free energy landscape originating from the entropic term. As can be seen, both end states are not changed with respect to each other. Nevertheless, a potential of a few kJ/mol is created by the circular projection favoring the end states which is a desired probability of the extended λ -dynamics approach. Even with the parabolic compensation function $f(\lambda)$ the free energy landscape is not perfectly flat when using the circular constraining.

Figure 2.9 underlines the influence of additional potential from the circle. The end states are significantly stabilized.

2.8 Heat bath considerations

During the simulation the λ -particle is coupled to a heat bath first to prevent energy building up in the λ -space that could be transferred to the real system and second to allow the particle to have a minimal velocity.

In general there is a set of possible heat baths and coupling algorithms available:

If the virtual particle is **coupled to the heat bath of the system** energy can flow from the λ depending virtual space to the real space which would lead to artifacts in the long run. Nevertheless changing the system by e.g. creating charge or moving charge from one position to another is always conjuncted with transferring energy from or to the system. It is assumed that this modification

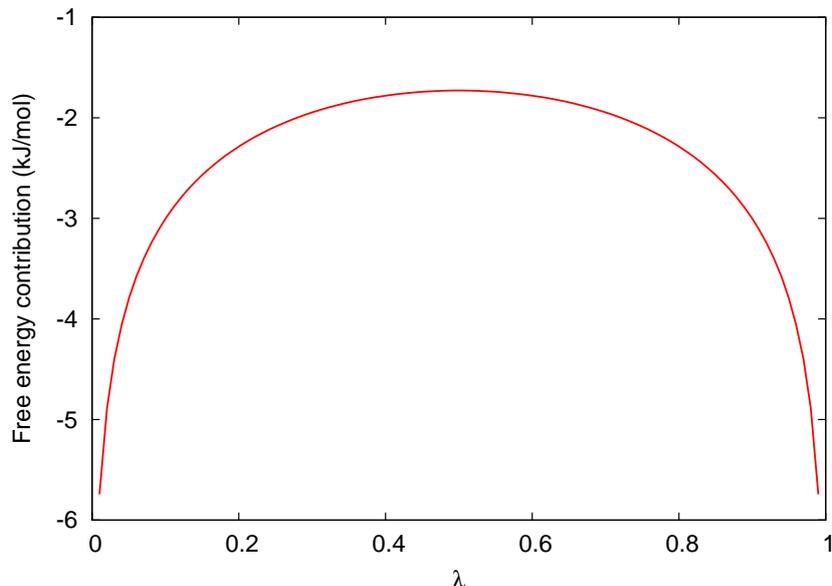


Figure 2.8: Helmholtz free energy potential affecting the λ -particle due to entropic effects without any overlaying external potential $V(\lambda)$

occurs slow enough to allow the thermostat to absorb or deliver sufficient energy and to keep the system in equilibrium. In fact, proper parameter selection and implementation should allow the system to fulfill the overall assumption of performing molecular dynamics simulations in a thermodynamical equilibrium.

To minimize energy transfer into or from the system **a separate heat bath for the virtual λ** was chosen. Having only one particle in the heat bath is more complex for the thermostat than having a large number because no ensembles can be generated and e.g. no averaging is possible. If only one particle is in the heat bath and this particle is kept at a constant temperature **using the Berendsen thermostat** [BPvG⁺84] (see Appendix as well), the particles velocity is scaled down every coupled step to the velocity defined as a temperature in the thermostat. The system is not able to stabilize in a favorable position but the λ -particle moves out with constant velocity.

Instead an **Andersen thermostat** [And80] was chosen, which is proven to generate correct velocity distributions and correct thermodynamical ensembles but was not designed to be used with only one particle. Nevertheless, selecting a new velocity from a gaussian distribution allows the system to oscillate

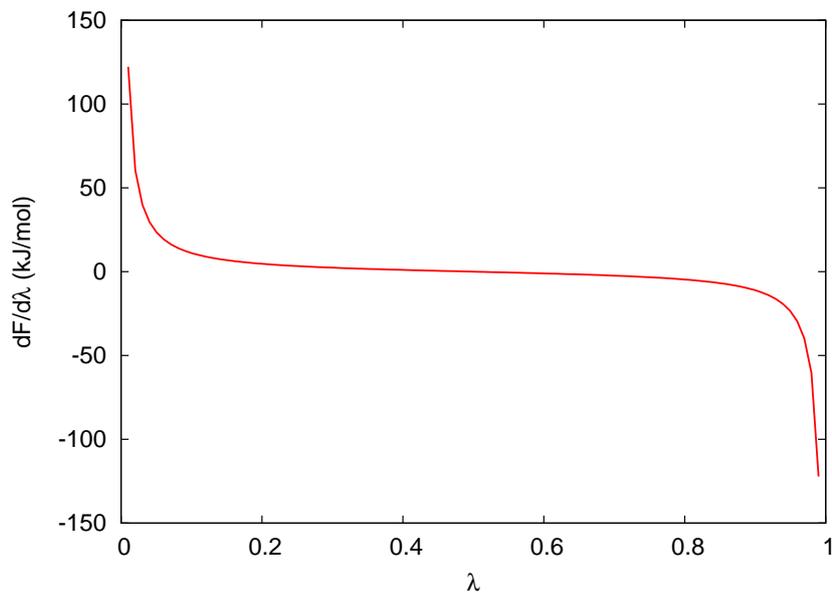


Figure 2.9: Derivative of the free energy surface. The end points are significantly stabilized.

around the favored position and no large momentum can be build up destroying small local environmental influences. By using the Andersen thermostat the λ particle is sampling the protonation space by a Monte-Carlo run.

2.9 Error estimation

The analysis of a constant pH simulation often requires to derive the ratio of protonated and deprotonated acid. This is expressed by the average over all λ in time $\langle \lambda \rangle$. As a trajectory, the λ variables depend on the previous so one measures a series of dependent points. This is taken into account by evaluating the autocorrelation function yielding the time τ in which the autocorrelation function has declined to $1/e$. With the standard deviation σ the error can be calculated as

$$err = \frac{\sigma}{\sqrt{T/\tau}} \quad (2.54)$$

whereby T is the total length of the simulation and T/τ is the number of independent blocks.

Chapter 3

Parameter analysis

The λ -dynamics code has some fundamental parameters not touched by any constant pH approach or later modifications of the code. The mass of the virtual particle can be set for each run, the same holds for the heat bath temperature, the coupling parameter and the starting λ . In this chapter the influence of these various parameters on the development of λ over time is analyzed and evaluated using a simple molecule as a test system.

3.1 The test system setup

The test system shown in Figure 3.1 consists of an aromatic ring as found e.g. in Histidine and two additional dummy atoms (depicted in purple in Figure 3.1). The dummy atoms are fixed point charges only having Coulomb interaction with the rest of the system. They are fixed in the plane of three defined atoms and keeps therefore its position relative to the aromatic ring. The protonatable sites of the imidazole side chains correspond to a configuration with a protonation of γ_1 (HISA) and a protonation of ϵ_2 (HISB).

By setting charges of $+0.3e$ and $-0.3e$ on the second dummy (and vice versa) a well defined potential can be applied to the freely moving charge and the system can switch between HISA and HISB to adapt to the charge configuration set.

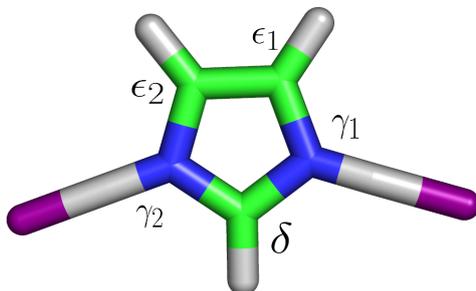


Figure 3.1: The test system setup consisting of an aromatic ring with additional dummy atoms (purple)

3.2 Temperature and temperature coupling

As noted in the Section 2.8 the heat bath is realized using a separate Andersen thermostat [And80] for controlling the λ particle alone. The temperature influences the sampling properties of the system: By using higher temperatures on the one hand sampling of the the free energy surface is increased but on the other hand fast and frequent protonation state changes are generating other problems as high energy transfer to the system and the danger of bringing the system out of equilibrium. To exclude such problems the temperature is always chosen to be the identical to the systems temperature.

3.3 Mass

The virtual particle's mass determines via its inertia how fast the system reacts on fluctuations and how much energy is required to overcome potential barriers. The first analysis performed was on the effect of increasing mass in the test system. The simulations were started at $\lambda = 0$ in a configuration favoring the other protonation state at $\lambda = 1$. Therefore it is expected that the system quickly reacts on the applied force and evolves to the correct protonation state. As shown in Fig. 3.2 an increasing mass slows down the reaction and smoothes the movement of the virtual particle λ .

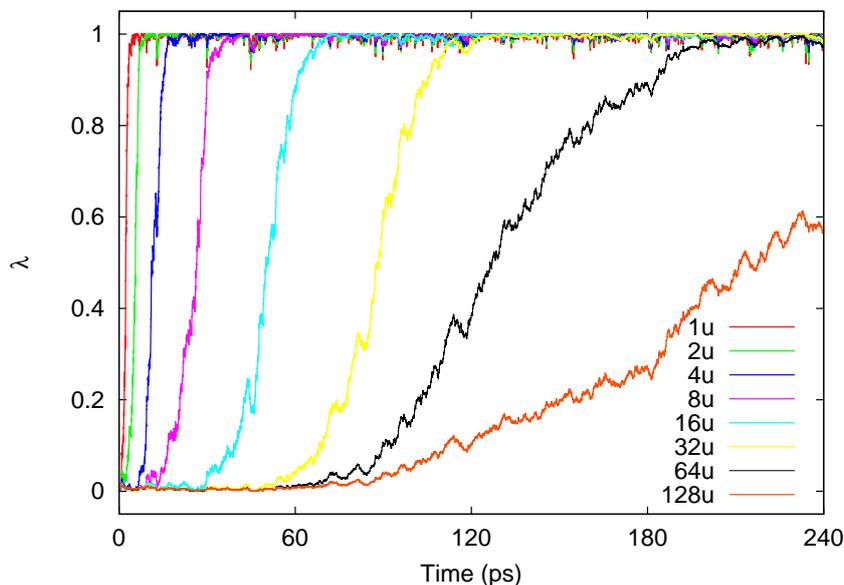


Figure 3.2: λ -development for different masses in a strong potential

3.4 Correlation between mass and temperature coupling

The behavior of the system and the influence of the mass is strongly related to the thermocoupling shown above. The more often the virtual particle's velocity is updated by the thermostat, the less important is the mass due to the destruction of inertia at each coupling event.

If there is only loose thermocoupling to the heat bath, energy exchange with the heat bath could not take place and in an isolated system the λ particle would harmonically swing on the circle around the end point (here $\lambda = 1$) without damping. Figure 3.3 illustrates this behavior with a probability of connecting to the heat bath of $p = 0.002$ each step. If the λ -controlled group is embedded into a larger protein structure as e.g. a Histidine in a large biomolecule, energy transfer to the environment damps the oscillation of the λ -particle around the end point but increases the energies of the surrounding atoms.

Temperature coupling ten times more often (around every 10th step) than in Fig. 3.2 destroys the momentum of the virtual particle more often so the

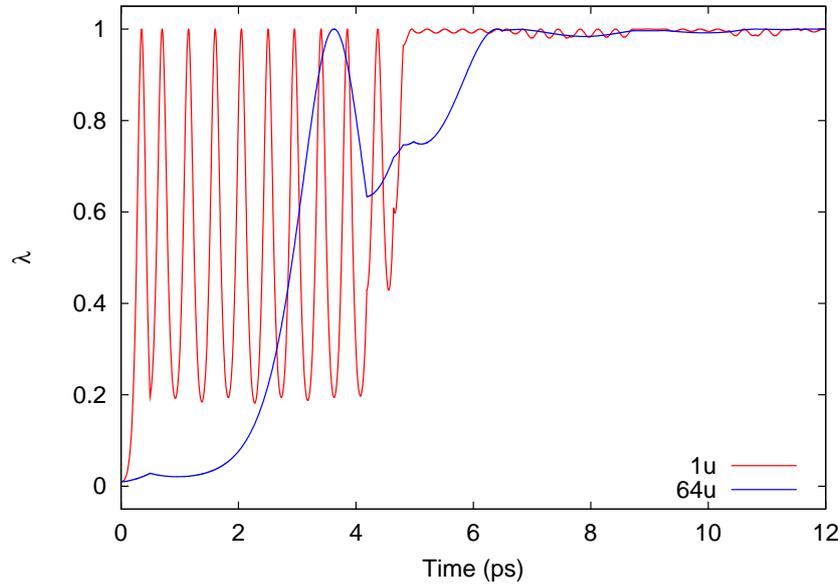


Figure 3.3: Position of the λ -particle for different masses in a strong potential with very loose coupling

development into a certain direction is slowed down as shown in Fig. 3.4.

As was shown loose coupling (τ very small) allows the λ -particle to build up momentum and sample more space but will also produces harmonic oscillation around an end-point due to the circular constraint construction. Strong coupling changes any momentum and transform the MD simulation in λ -space to a pure Monte-Carlo run.

3.5 Starting λ dependency

For a stable λ -dynamics based MD simulation the starting value of λ should be irrelevant after a certain equilibration time. As can be seen in Fig. 3.5 the test system with simple steep potential gradient evolves quickly into the correct state. If additional barriers would be imposed the system could stay longer in the beginning state.

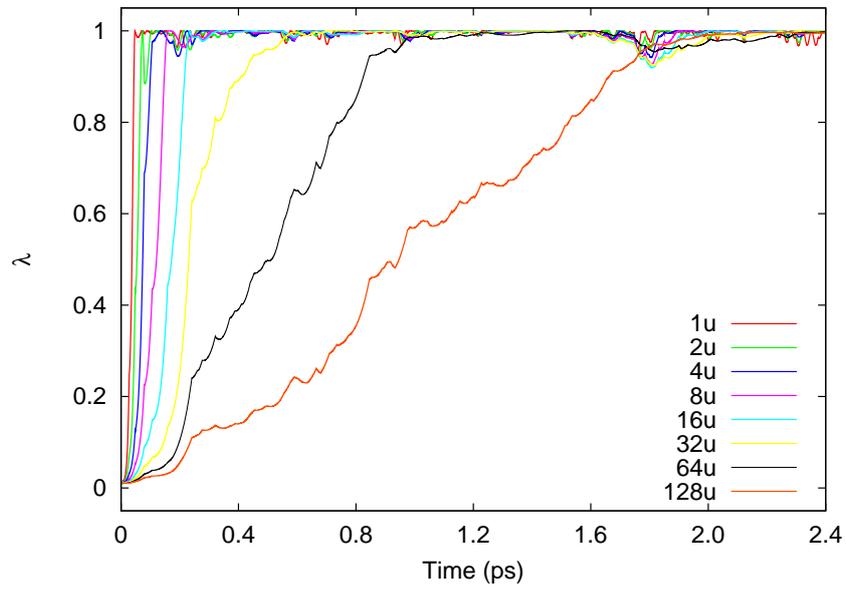


Figure 3.4: Position of the λ -particle for different masses in a strong potential with strong coupling

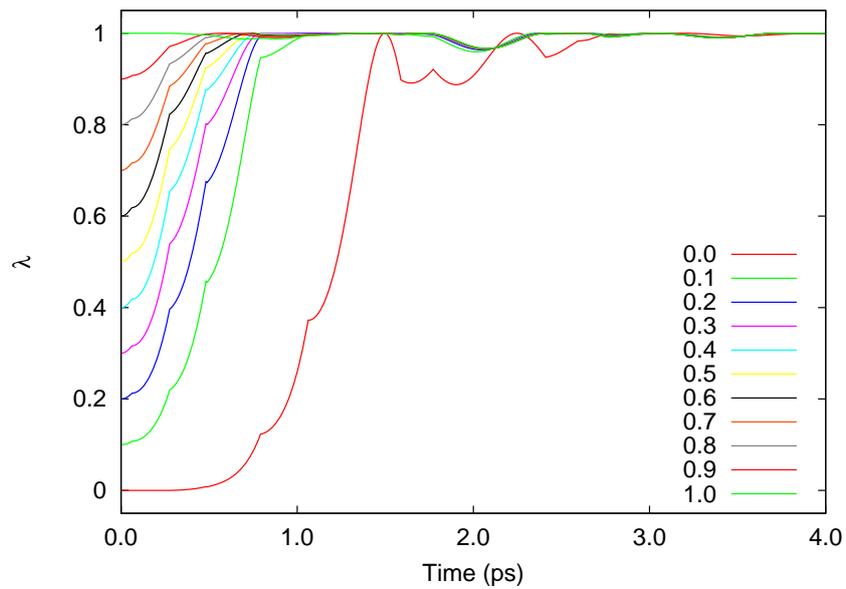


Figure 3.5: From every starting λ the system evolves to the preferred state

3.6 Barrier height

The barrier separating both end-states should be high enough to allow sufficient sampling of the end-states (e.g. λ in 90% of the time around 0 or 1) and low enough to allow transitions from one to the other state. It would be best if the average time the λ -particle stays at one end-state and the transitions rates between both states would reflect correct experimental measurable values. Due to the fact that NMR shows typical protonation/de-protonation times in a range of μs to ms , sampling in MD simulations would be very bad while running only a few ns with such high barriers. Therefore the barrier is reduced to few kJ/mol to allow the system to change protonation state more often and optimize sampling of λ . An optimal algorithm to determine the barrier height or even let the the barrier height dynamically change is under development.

3.7 Summary

To allow a stable λ -dynamics simulation, the parameters must be chosen to not destroy the required properties as regular transitions, fast reaction on local environment changes and minimal energy transfer to the rest of the system. By selecting the temperature for the λ -particle's heat bath identical to the temperature of the complete system, the energy transfer is minimized. As was shown the influences of the mass and the temperature coupling are highly related. Overall the total energy of the virtual particle should be low to reduce again the energy transferable to the local environment. This can be reached by selecting a small mass and relatively strong coupling. The parameters used for all following simulations if not stated otherwise are shown in Table 3.1. The barrier height is highly system dependent. For a good analysis of a simulation using the λ -dynamics approach, every run should be controlled whether the virtual particle leaves it's starting state. As a starting value $2.5kJ/mol$ is chosen following [LJI04]. Additionally there should be some well observable transitions from one state to the other to ensure full sampling of the λ space.

Parameter	Starting value
Temperature coupling: ν	4
Temperature of the λ -particle	300K
Mass	5u
Barrier height	2.5kJ/mol
Starting λ	irrelevant

Table 3.1: Values for the various parameters introduced by the technical realization of the λ -dynamics. These build a starting set for simulations with unknown energy landscapes.

Chapter 4

Results 1 - Single amino acid

The λ -dynamics controlled switching between two protonation states of a single amino acid allows a proof of principle and gives a first insight into which parameters are required and how they affect the λ -particle's behavior.

Glutamic acid as shown in Figure 4.1 was chosen to be parameterized and analyzed. Glutamic acid (3-letter coding GLU, 1-letter coding E) is a neurotransmitter and a standard building block for proteins. It is often added to casual food or fast food as a flavor enhancer. Due to its simplicity glutamic acid is a suitable test candidate for the extended λ -dynamics algorithms.

Glutamic acid has a pK_a of 4.25 and therefore a titration curve as depicted in Figure 4.2.

4.1 Parametrization system setup

The system consists of a glutamic acid including the standard amino-acid backbone. There are three titratable sites: The backbone and the symmetric COO-group, so the hydrogen can bind to any of the two oxygens. Parallely performed analysis have shown that simulating only the side chain, the functional group COOH, result in similar values for the free energy difference between the protonated and the deprotonated state. Therefore the acid is modeled with only one titratable site to reduce complexity and placed in a dodecahedral box (see Appendix) filled with 4231 water molecules. The system was energy minimized in 400 steps and equilibrated for 1 ns which should be sufficient for a system with such low conformational complexity. Energy minimization and the water equilibration were performed with position restrains on the amino

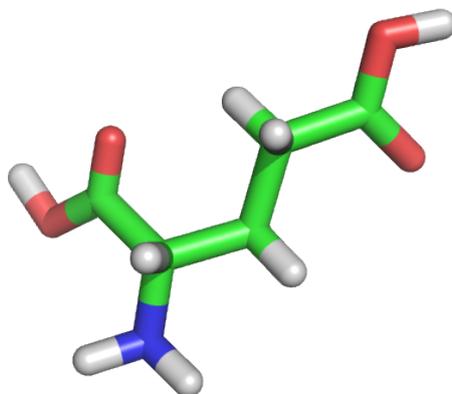


Figure 4.1: Glutamic acid in its protonated form with neutral end groups (N-terminus: NH_2 , C-terminus: $COOH$).

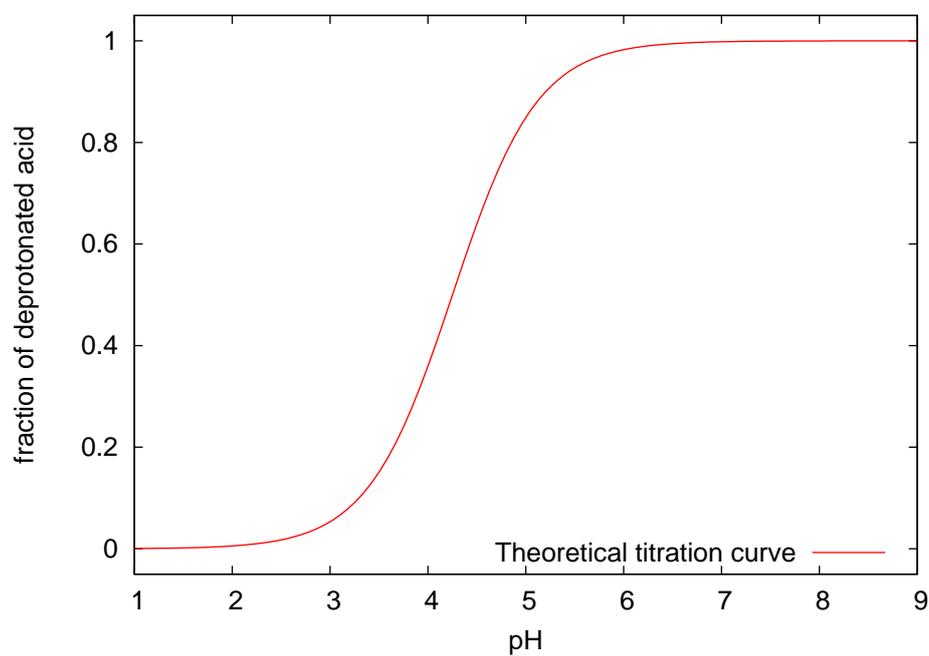


Figure 4.2: Theoretical titration curve for glutamic acid with experimentally determined pK_a of 4.25. The fraction of deprotonated acid α is calculated from the pK_a by $\alpha = \frac{1}{10^{pK_a - pH} + 1}$.

Direction	Reg. const. ($\frac{kJ}{mol \cdot \lambda}$)	Reg. coeff. ($\frac{kJ}{mol \cdot \lambda^2}$)	ΔG ($\frac{kJ}{mol}$)
Forward	98.596 ± 0.0529	-633.2 ± 0.092	-218.00 ± 0.1
Backward	-53.94 ± 0.0530	$+633.0 \pm 0.092$	-218.44 ± 0.1

Table 4.1: Results of linear regressions on the $\delta G/\delta \lambda$ data generated by free energy perturbations and the corresponding free energy difference calculated by integrating over λ .

acid. The topology consists of the protonated GLUH structure in state *A* and the deprotonated form GLU in state *B*. Due to the different protonation states a total charge shift of $-1e$ was introduced between the two topologies. This total charge difference is not yet accounted for, so the systems reaction is depending on the long range charge interaction scheme (e.g. Cutoff, Reaction Field, Particle Mesh Ewald). All following simulations were performed using the ReactionField method and the GROMOS96 53A6 forcefield.

4.1.1 Free energy

The free energy landscape between the two physically relevant end states was analyzed by thermodynamic integration based on 21 discrete steps with 5ns simulation time each. Additionally a 5ns slow growth simulation was performed and convergence was analyzed. Due to the simplicity of the system equilibration times are extremely short and the 5ns multi-conformational TI shows no significant differences ($< 0.5kJ/mol$) in ΔG after 500ps or 5ns. The results for the 5ns slow growth forward and backward simulations are illustrated in Table 4.1 and the $\delta G/\delta \lambda$ data from this simulation is shown in Figure 4.3:

The error in free energies calculated with FEP methods is usually derived by performing the simulation forward and backward and analyzing the hysteresis. The error corresponds to the difference in forward and backward free energies. With a free energy difference of $\Delta G_{Forward} - \Delta G_{Backward} = 0.44kJ/mol$ the error is below the systematical error of every FEP analysis. It can be therefore stated that a 5ns simulation is sufficient to accurately measure the underlying free energy surface for a glutamic acid in water so that the following regression function is accepted as a basis for further calculations:

$$\frac{\delta G}{\delta \lambda} = 98.59 \frac{kJ}{mol} - 633.2 \frac{kJ}{mol} \cdot \lambda \quad (4.1)$$

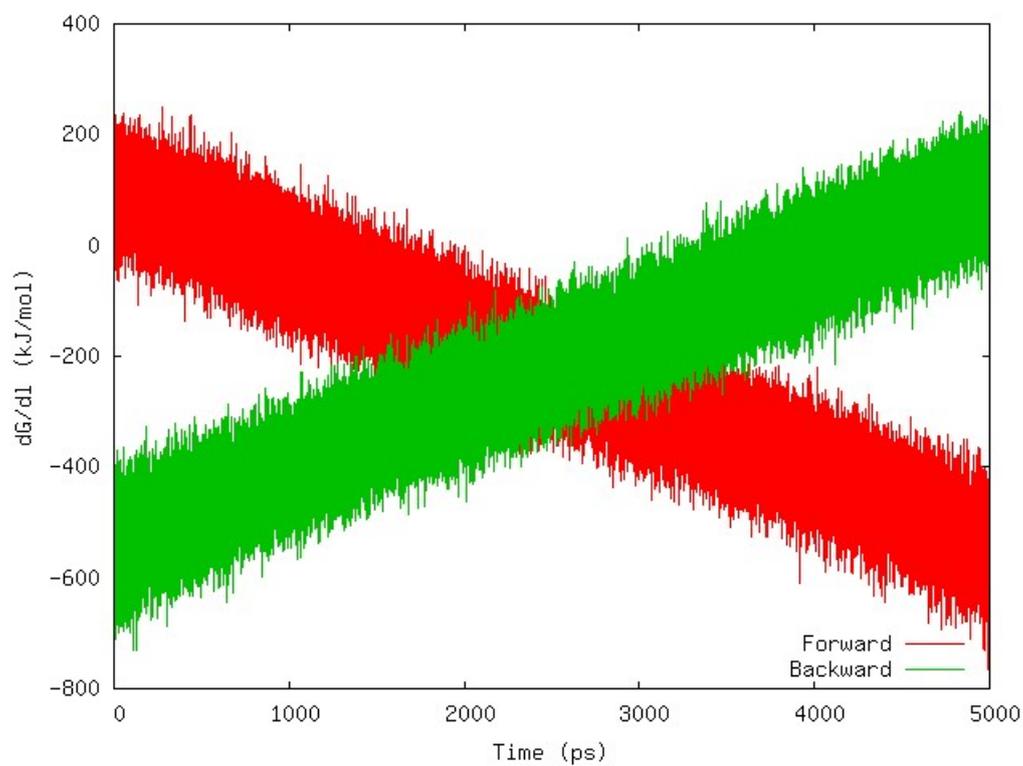


Figure 4.3: Free Energy Perturbation simulations forward and backward, each $5ns$. The linear behavior as predicted by Linear Response Theory (see 2.4) can be seen.

It is necessary to stress out the fact that a sufficient box size is mandatory for obtaining reasonable results. A simple glutamic acid in a water box including 2000 water molecules was found to not suffice. From up to around 3000 water molecules in a box (tests until 30000 water molecules) the same values for ΔG is obtained which is converged. FEP convergence is shown by the very small difference in the slow growth FEP of only 0.44 kJ/mol.

4.1.2 Constructing the experimental energy landscape

With the regression derived from the free energy perturbation, $f(\lambda)$ (2.23) can directly be computed to

$$\frac{\delta f(\lambda)}{\delta \lambda} = -98.59 \frac{\text{kJ}}{\text{mol}} + 633.2 \frac{\text{kJ}}{\text{mol}} \cdot \lambda \quad (4.2)$$

Now $\kappa(\lambda, pH)$ needs to be calculated which yields following equation 2.18 for glutamic acid with $pK_a = 4.25$:

$$\frac{\delta \kappa(pH, \lambda)}{\delta \lambda} = 2.3 \cdot k_B T (4.25 - pH) \quad (4.3)$$

4.2 Glutamic acid titration curve

The first test for reproducing the correct experimental free energy landscape is the calculation of glutamic acid's titration curve. Constant pH simulations with 500ps equilibration time and 4500ps simulation time were performed for 9 points between $pH = 1.0$ and $pH = 9.0$. For each simulation the average λ over the last 4500ps was computed and the error was estimated as outlined in Section 2.9. Figure 4.4 shows the simulated titration curve (red) compared to the theoretical curve (green). On the one hand the error estimation function seems to underestimate the error and on the other hand the sampling is not too good. After 5ns simulation time the theoretical titration curve is qualitatively reproduced (especially the sigmoidal behavior) but some points e.g. at $pH = 5.25$ and $pH = 5.75$ are quite far off.

For better sampling of the end states at $pH = 1$ and $pH = 9$ increasing the radius of the λ -particles constraintment circle could be a helpful option. By

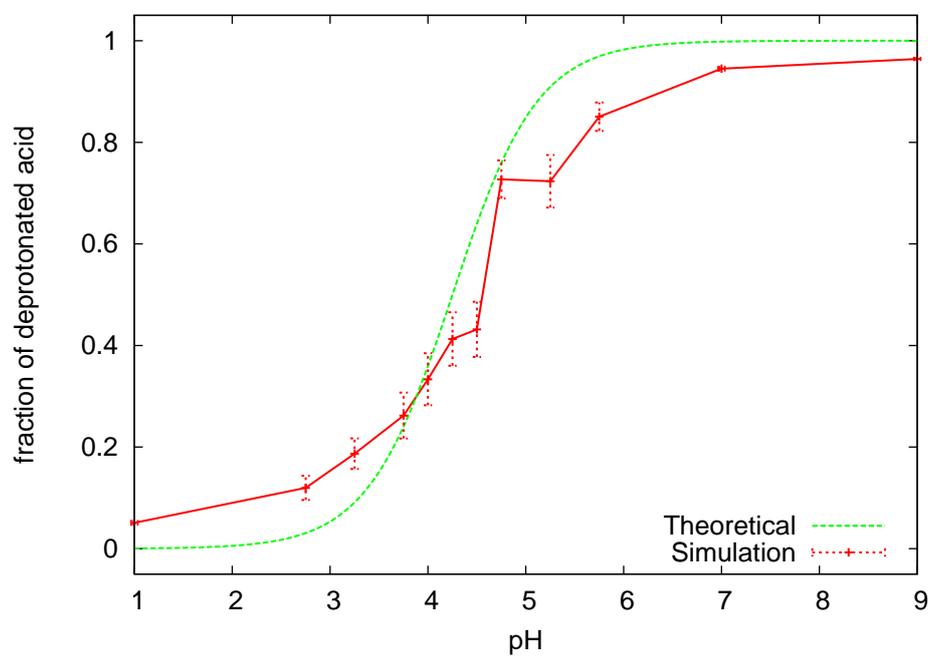


Figure 4.4: Titration curve for glutamic acid in a water box.

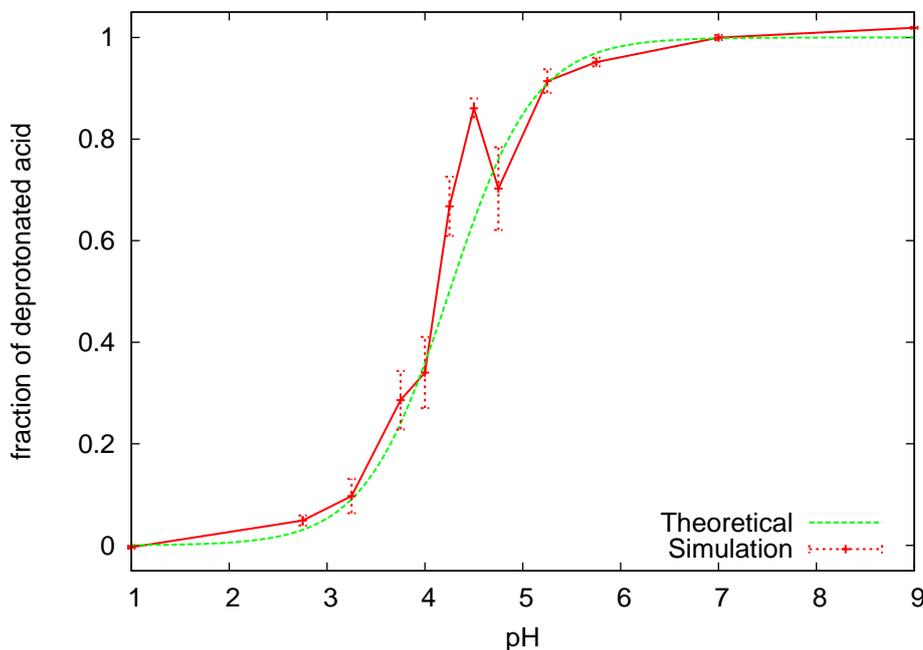


Figure 4.5: Titration curve for glutamic acid in a water box. The radius for the circular projection was increased from $r = 0.50$ to $r = 0.55$

thermal fluctuations λ could otherwise never stabilize at the correct values of ≈ 0 and ≈ 1 but will always stay a bit above and below these values.

As can be seen in Figure 4.5 the end states around $\lambda = 0$ and $\lambda = 1$ are better reached at a radius of $r = 0.55$ instead of $r = 0.50$. For the following simulations always a radius of $r = 0.55$ is used.

4.2.1 Statistics

To enhance the sampling by increased simulation time and more computational resources can be achieved by two different approaches. Performing one simulation much longer to derive better statistics or to start multiple simulations from the same starting structure simply by changing the starting velocity distribution for all atoms. In Figure 4.6 a single $30ns$ run (green) is compared to the averaged titration curve (red) over six independent $5ns$ simulations. Especially in the critical area around $pH \approx pK_a$ the simulated titration curve is reproducing the theoretical curve more accurately.

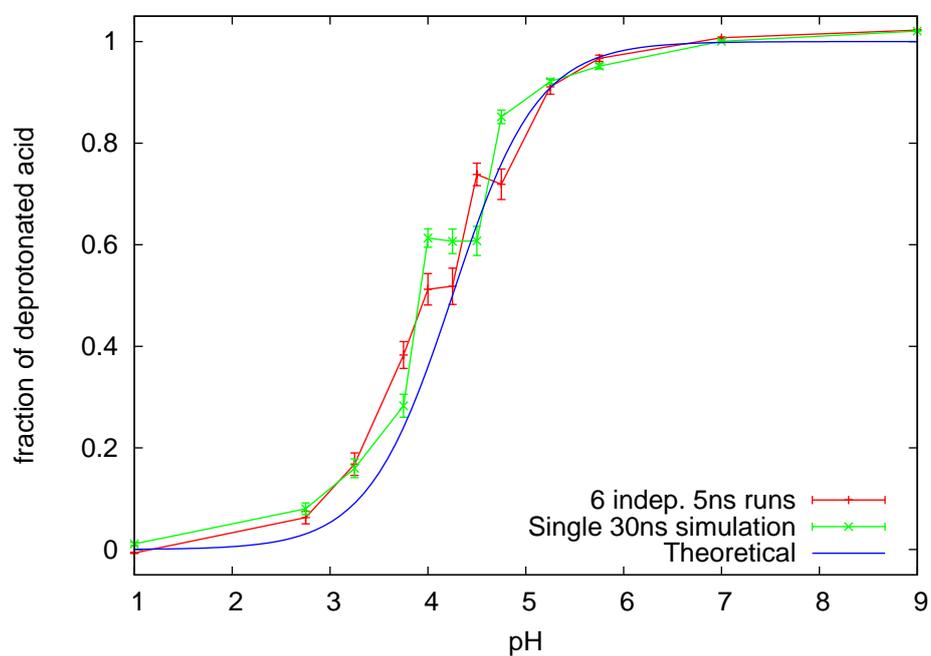


Figure 4.6: Titration curve for glutamic acid in a water box for the average of 6 independent $5ns$ runs compared to the titration curve based on one simulation of $30ns$

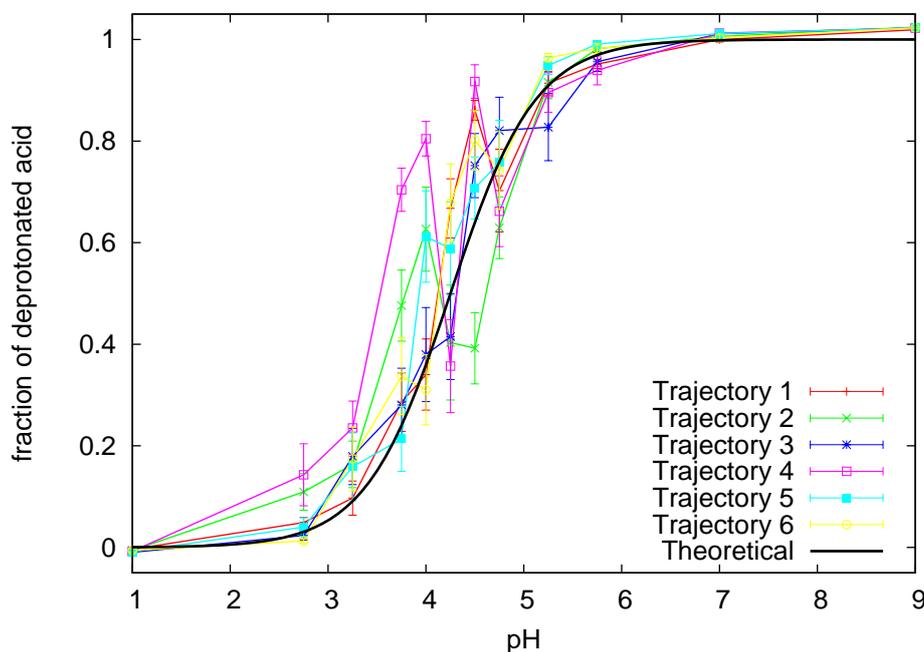


Figure 4.7: Titration curves of six independent λ -trajectories in the same system with the only difference in the starting velocity generator's seed

As can be seen in Figure 4.7 the quality of the titration curve reproduction is different for identical runs of $5ns$ whereby only the seed for the random number generator used in generating the Maxwell-Boltzmann starting velocity distribution was changed. Trajectory 4 is far off and fluctuates while trajectory 3 qualitatively describes the theoretical titration curve quite well. The most heavy fluctuations are around the pH equal to pK_a region where the total energy landscape is quite flat, which is expected. As the finally averaged curve in Figure 4.6 shows, satisfying agreement with the theoretical curve, it can be concluded that for a good titration curve experiment multiple simulations should be run for each point at constant pH.

Another observation is the high fraction of deprotonated acid for pH lower than the pK_a . Especially around $pH=3.25$ and $pH=3.75$ this effect is found in all six λ -trajectories. A potential reason to this aberrant behavior is the barrier of $2.5kJ/mol$ allowing some fluctuation around the end points. The fluctuations underline the difficulties in choosing an optimal barrier height for optimal sampling: On the one hand the barrier is not low enough to allow reg-

ular transitions (see Figure 4.9) and on the other hand it is not high enough to prevent the λ -particle from fluctuating heavily around the end state. Figure 4.8 shows a histogram giving the percentage of time λ is sampling a given interpolation area and Figure 4.9 depicts the λ -trajectory in time. As Figure 4.9 illustrates, only 4 transitions take place during the 5ns long simulation. Increasing the barrier height therefore requires a substantial increase in simulation length as well to be able to observe some transitions. On the other hand decreasing the barrier height could even lead to a sampling improvement but will increase the overall time the system is in an unphysical interpolated state. A possible solution to improve sampling without decreasing the end state population could be the application of an adaptive barrier. As illustrated in Section 2.6 the barrier lowers while the λ -particle is around $\lambda = 0$ or $\lambda = 1$ and increases when too many transitions occur and/or the system stays too long in between both end states. A detailed evaluation and analysis of the parameters involved, as e.g. the thresholds when to lower or raise the barrier, is still under development.

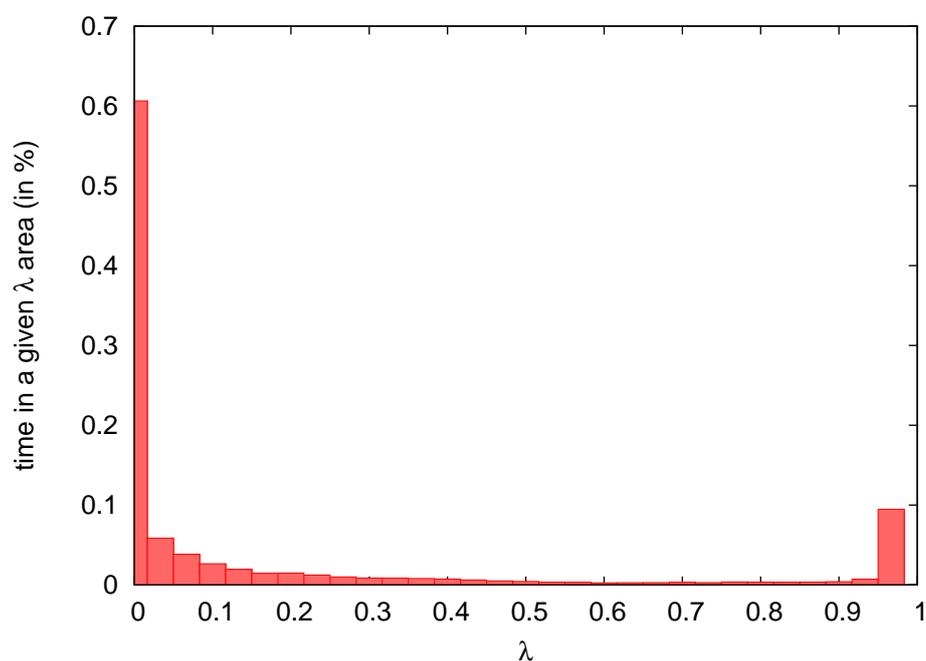


Figure 4.8: Distribution of λ in a simulation at pH=3.25. On the y-axis the time (in percent of total simulation time) the λ -particle stays in the specified region is given. Especially the region from 0.1 to 0.3 is important for the deviation from the theoretical titration curve.

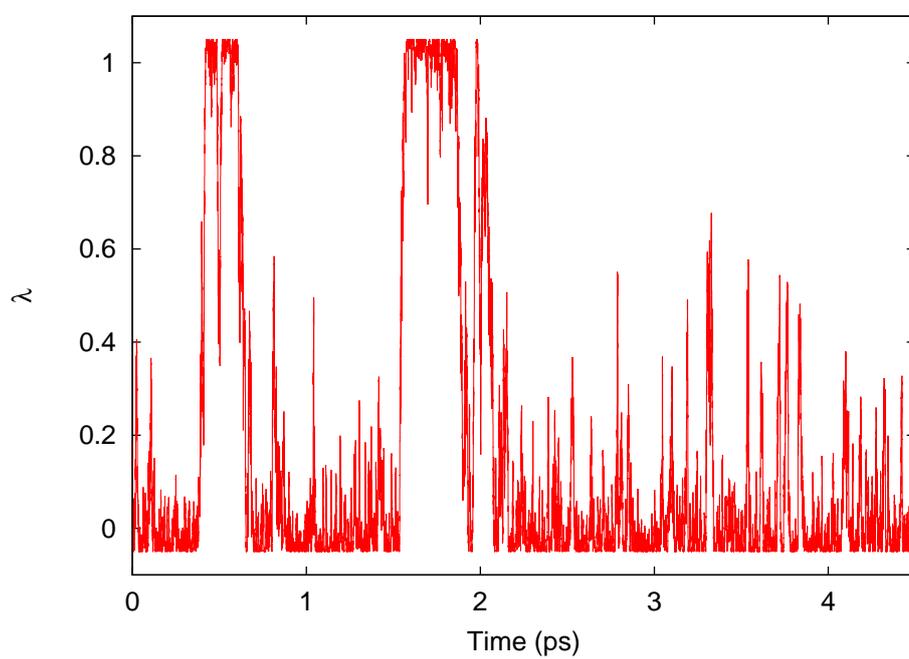


Figure 4.9: λ -trajectory for one glutamic acid in a simulation at constant pH of 3.25. The λ -particle fluctuates heavily around the end state at $\lambda=0$ but only 4 transitions can be observed.

Chapter 5

Results 2 - peptide: oligoglutamic acid

In the following subsections different systems including two or more glutamic acids are presented and the interactions between multiple titratable groups are analyzed.

All simulations were performed after energy minimization and $500ps$ equilibration time. Each titration curve was calculated in a $5ns$ long simulation for 12 distinct pH values. The detailed simulation parameters and methods used in all simulations are tabulated in 5.1. Position restraints are used in all simulations to prevent the system from forming e.g. salt-bridges.

For comparison, a theoretical titration curve was calculated as well. These theoretical curves are constructed by simply multiplying a single amino acid's titration curve with the number of acids and do not represent any experimental data. Nevertheless, influences of the peptide environment are highlighted.

5.1 Duplicated parametrization system

For the first simulation of multiple titratable groups a very simple test system was chosen: The system used in the single glutamic acid analysis as illustrated in Section 4.1 was copied in one direction, effectively doubling the box size and introducing a second glutamic acid. Figure 5.1 depicts the original box and Figure 5.2 shows the new system as a copy of the box in one dimension. At the start of the simulation both glutamic acids are protonated. Each acid is coupled to a single λ -variable and both λ -variables are propagated independent. The

Parameter/Method	Value)
Simulation length	5ns
Timestep	2fs
Long range Coulomb	ReactionField
Solvent representation	Explicit
System temperature	300K
λ temperature	300K
Mass m_λ	5u
λ -dynamics additional barrier height	2.5kJ/mol

Table 5.1: Parameters and methods used in all simulations

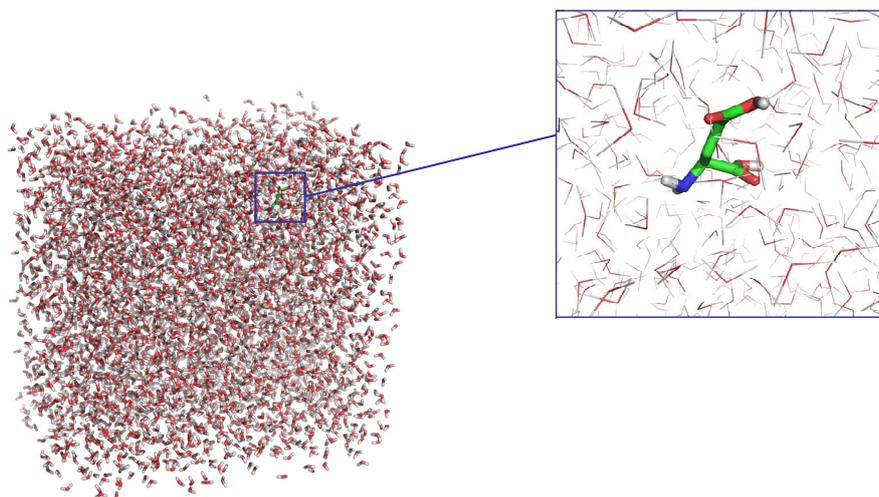


Figure 5.1: Parametrization system from Section 4.1

behavior of the two acids is very similar to the behavior of the single glutamic as described before. This shows that the code allows the computation of multiple titratable states.

Five simulations of 5ns each have been performed for 12 different pH values between 1 and 9. Overall, after averaging at every pH value these simulations yield a titration curve as depicted in Figure 5.3. The theoretical titration curve is qualitatively reproduced although the values at pH 3.25 and pH 3.75 are significantly above the theoretically expected values. As outlined before, a higher barrier height and a significantly longer simulation might allow approaching the theoretical titration curve in the critical pH region as well.

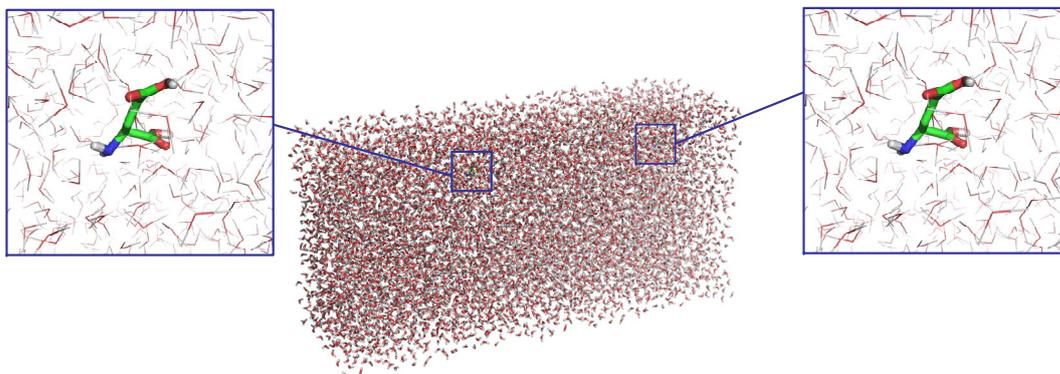


Figure 5.2: Comparison system constructed by copying the parametrization box in x direction.

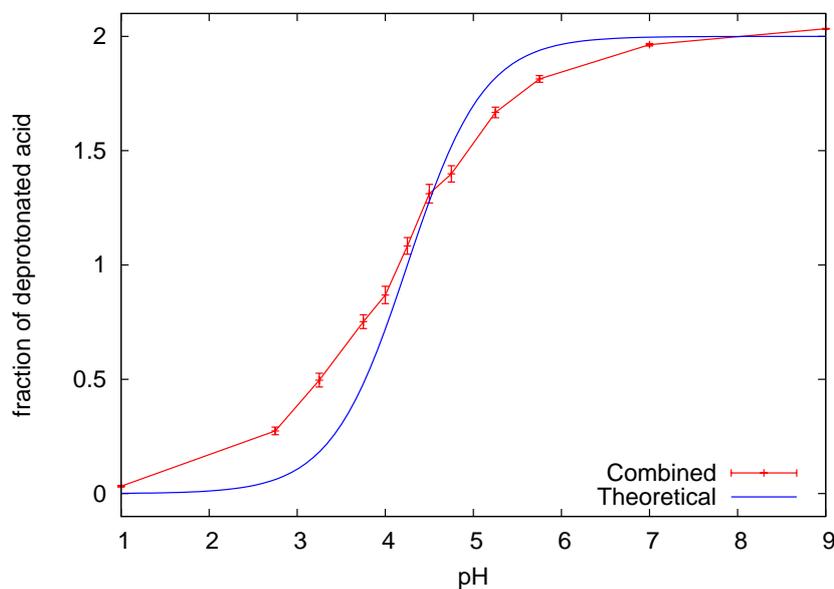


Figure 5.3: Titration curves of the two independent glutamic acids. The acids are position restrained and cannot interact with each other. Therefore each acid should react on the pH without being influenced by other titratable groups and the sum of both fractions of deprotonated acid should approach the theoretical curve for two glutamic acids.

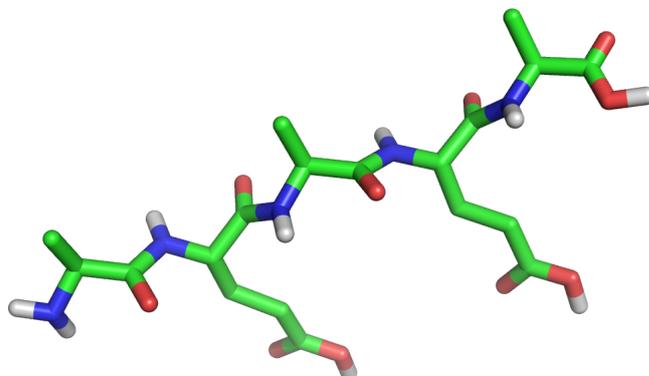


Figure 5.4: Test peptide including two glutamic acids and neutral alanines (AEAEA)

5.2 Test peptide (2 glutamic acids)

After applying the λ -dynamics approach to two independent, distantly located amino acids the interaction of two neighboring acids is analyzed. The second test system (AEAEA) is a peptide with two glutamic acids and three alanine as depicted in Figure 5.4. The terminal groups are neutral and the position of all heavy atoms are restrained.

The titration curves shown in Figure 5.5 were calculated based on three independent runs at 12 distinct pH values. It can be seen that all titration curves are more linear and do not show a strong sigmoidal behavior. Figure 5.6 depicts the combined titration curve calculated by averaging over the independent runs.

Compared to the titration curve of the two separated acids as shown in Figure 5.3 a distinct variation is observed around $\text{pH}=4.25$. As Figure 5.6 supports, a plateau is forming around a $\text{pH} = \text{p}K_a$. A hypothesis for the stabilization of protonation is the influence of the second glutamic acid: While one glutamic acid is deprotonating, the relative $\text{p}K_a$ of the other acid shifts due to the charge of up to $-1e$ imposed by the deprotonated acid. Overall, a higher pH is required to deprotonate the acids due to the effects of the local environment. Figures 5.7 and Figures 5.8 depict the distribution of λ for the first and the

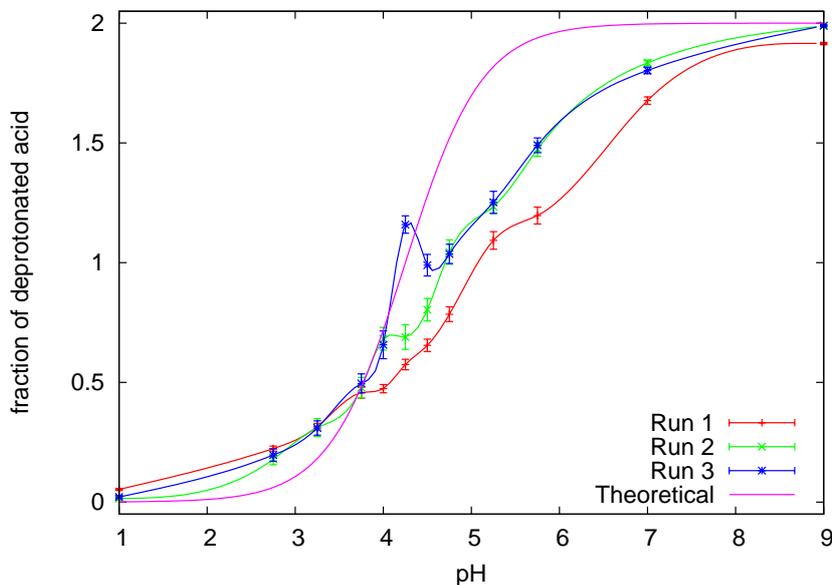


Figure 5.5: Titration curves for 3 independent simulations of two glutamic acids in the AEAEA peptide.

second glutamic acid in time at $pH = pK_a$. The first acid, illustrated in the red histogram and the red trajectory in Figure 5.9, is as often found in the protonated as in the deprotonated state, which is expected at $pH = pK_a$. The second glutamic acid, however, is heavily influenced by the first one. Instead of an equal population of the protonated and the deprotonated state, it is remaining in the protonated configuration (see histogram in Figure 5.8). The trajectory highlighted in blue in Figure 5.9 illustrates the effect of the first (red) glutamic acid. The degree of deprotonation is the higher the more protonated the first acid is (see peaks around 0.3ns, 1.2ns, 1.7ns and 3.4ns). To rule out a potential artifact in the implementation places priority on the first glutamic acid, the order of the residues in the peptide was reversed which did not alter the results.

5.3 Test peptide (3 glutamic acids)

The third test system extends the second system by an additional glutamic acid. Three glutamic acids are embedded between two alanine forming a simple

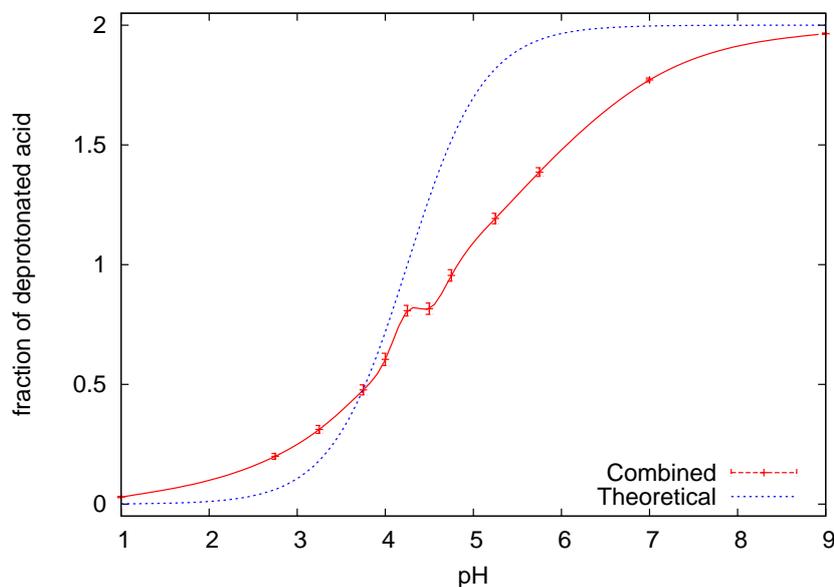


Figure 5.6: The total fraction of deprotonated glutamic acids in the AEAEA peptide, averaged over 3 independent runs of $5ns$ each.

peptide as illustrated in Figure 5.10. Again a series of titration curve simulations is performed to analyze the influence of pH on the overall fraction of deprotonated acid.

Figure 5.11 depicts the titration curves of the test peptide for each of the simulations performed. The curves are again differently shaped compared to the runs simulating one or two glutamic acids. All three curves show a pK_a shift at around $pH = 4.25$ and around $pH = 4.75$. The plateaus forming can be seen best in the averaged titration curve shown in Figure 5.12.

It can be assumed that the plateaus are again results of a pK_a shift originated by the deprotonation of the surrounding amino acids as seen before in the peptide with two glutamic acids. If a single amino acids would be simulated around $pH = 4.25$ the probability to be found in the deprotonated state is 0.5. In the three glutamic acid peptide, each glutamic acid is under the influence of neighboring charges and if the first glutamic acid deprotonates around $pH = 4.0$ to $pH = 4.25$ the local environment stabilizes the protonated state for the remaining glutamic acids. Therefore the total fraction of deprotonated glutamic acid does not change with increasing pH until the pH is high enough

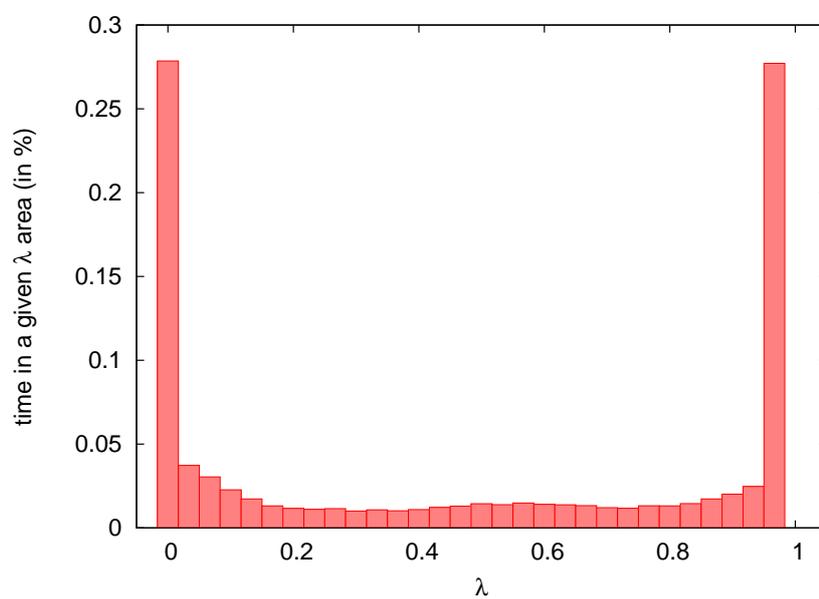


Figure 5.7: Distribution of λ for the first glutamic acid in a constant pH simulation at pH 4.25. The end states around $\lambda = 0$ and $\lambda = 1$ are equally populated.

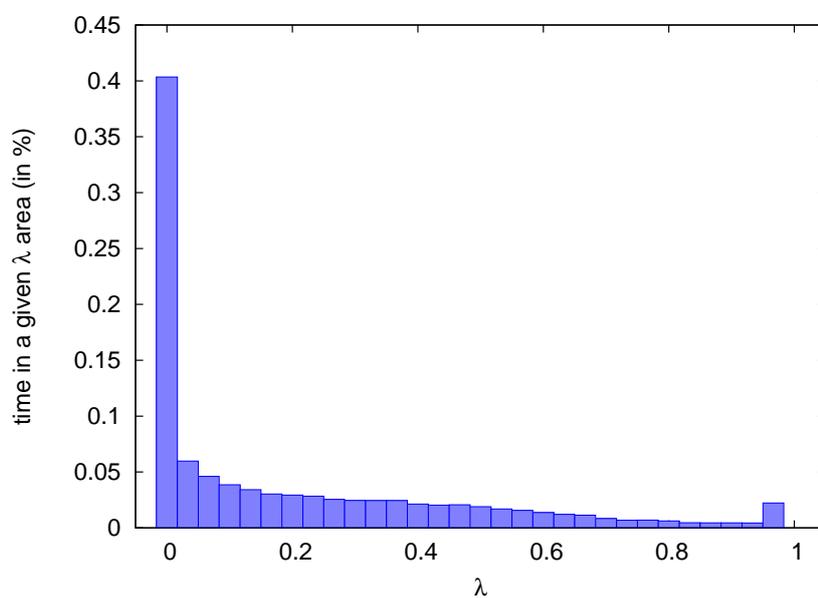


Figure 5.8: Distribution of λ for the second glutamic acid at constant pH of 4.25. The ratio between protonated ($\lambda = 0$) and deprotonated acid ($\lambda = 1$) is suggesting an influence of the first glutamic acid. Instead of equally populating the end states the deprotonated state is favored.

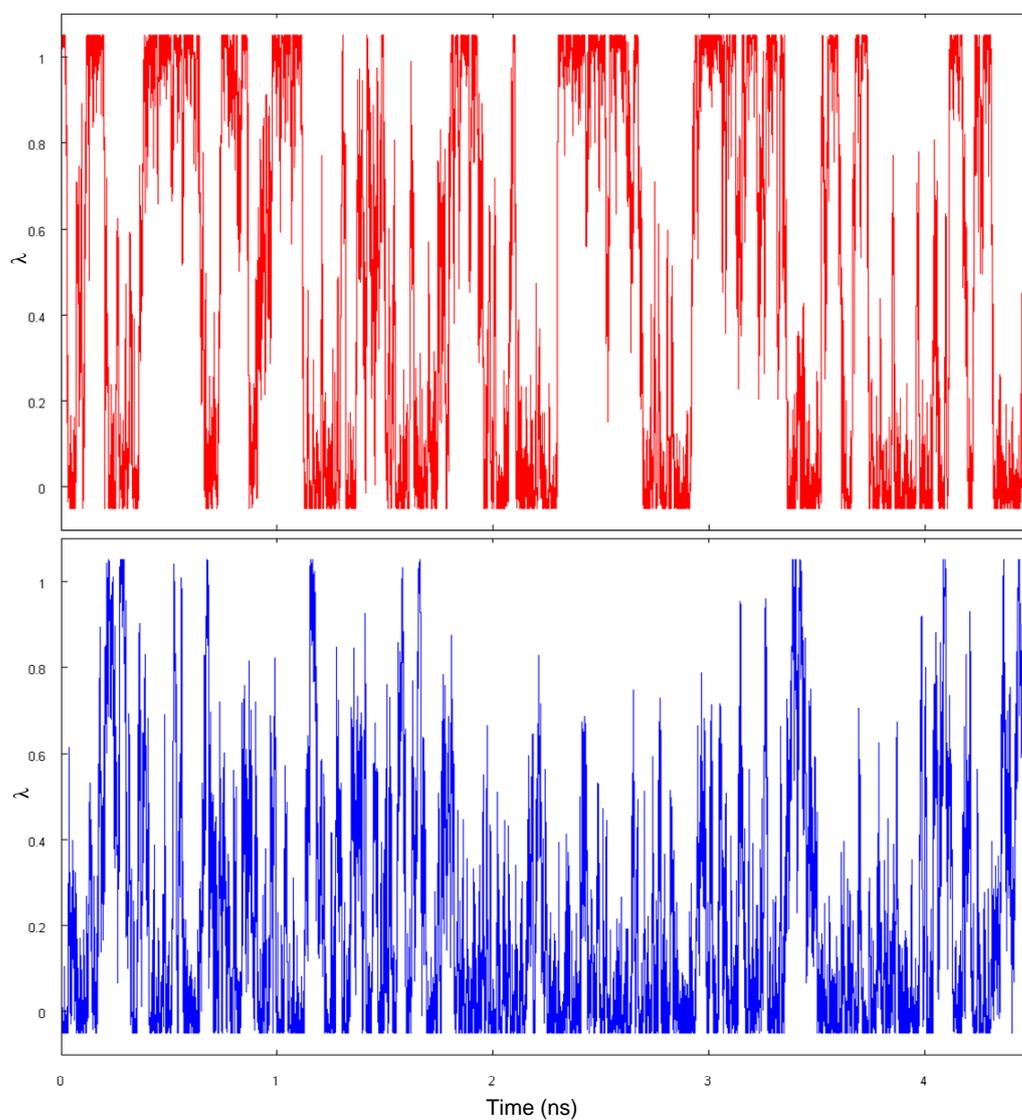


Figure 5.9: Trajectory of the λ -particle for the two glutamic acid. The first glutamic acid (red curve) is interacting with the second (blue curve) as the second can only deprotonates while the first is protonated, e.g. at 0.3ns, 1.2ns, 1.7ns, 3.5ns and 4.4ns.

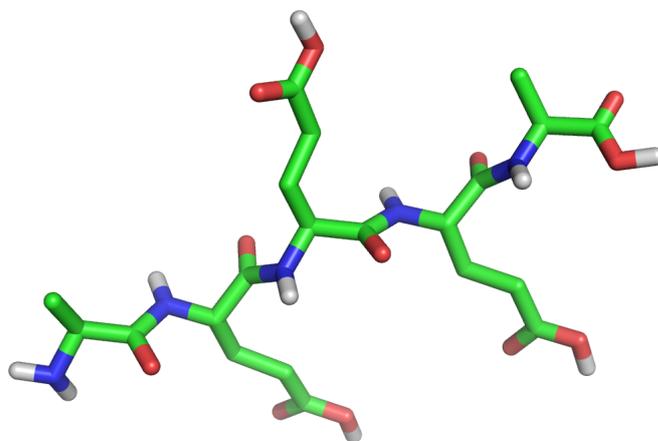


Figure 5.10: Test peptide with three directly attached glutamic acids surrounded by alanines

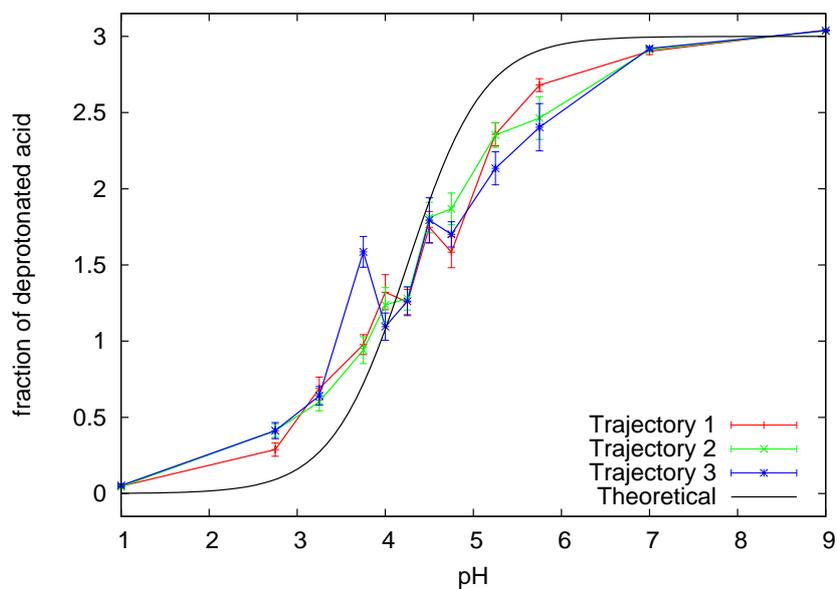


Figure 5.11: Titration curves of three runs of the three glutamic acid peptide.

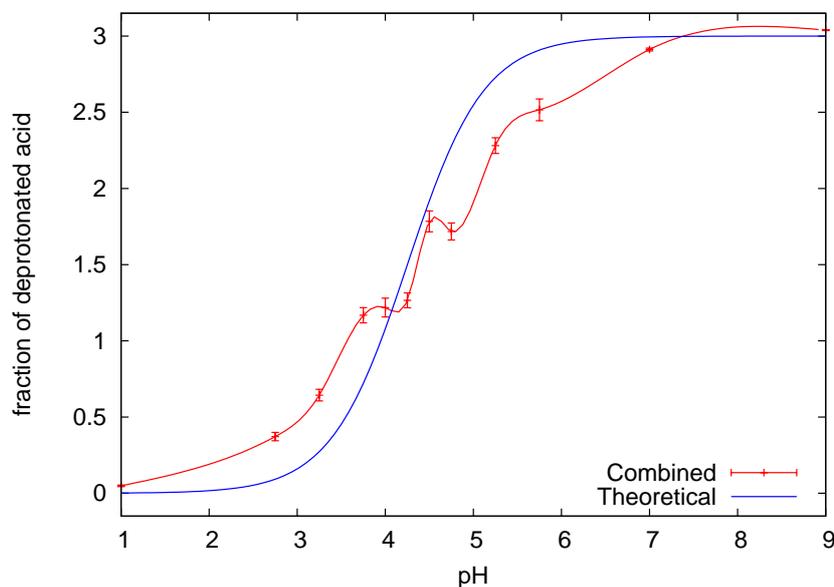


Figure 5.12: Combined titration curve over 3 runs of a three glutamic acid peptide.

that the protonation of a second glutamic acid becomes favorable. This condition is found to be around $pH = 4.75$.

Figure 5.13 supports this argument by three λ -trajectories showing the λ -development over time for each glutamic acid of one run (number 2) at $pH = 4.25$. As can be seen the protonation states of *Glu2* (blue curve) and *Glu3* (brown curve) are heavily correlated. If *Glu2* deprotonates *Glu3* is protonated again (e.g. at $0.9ns$ and $1.7ns$). However, *Glu1* is most of the simulation time in the protonated state which indicates the stabilization of this state at $pH = 4.25$ by the remaining two glutamic acids and their averaged charge of $\approx -1e$. Figure 5.14 depicts the titration curve for this *Glu1* and underlines an untypical ratio of ≈ 0.2 protonated and deprotonated acid around $pH = 4.25$ where a ratio of ≈ 0.5 is expected on the basis of a single glutamic acid. Additionally the curve for higher pH seems to be slightly shifted by half a pK_a unit.

The combined titration curve (Figure 5.12) shows a third stabilized section around a $pH = 5.75$. This could be related to a stabilization of glutamic acid as before. However, this interesting observation was not further investigated yet. Alternatively the effect could be due to statistical error, which are significantly

higher in the peptide simulations as compared to isolated glutamic acids as shown in Figures 5.5 and 4.7.

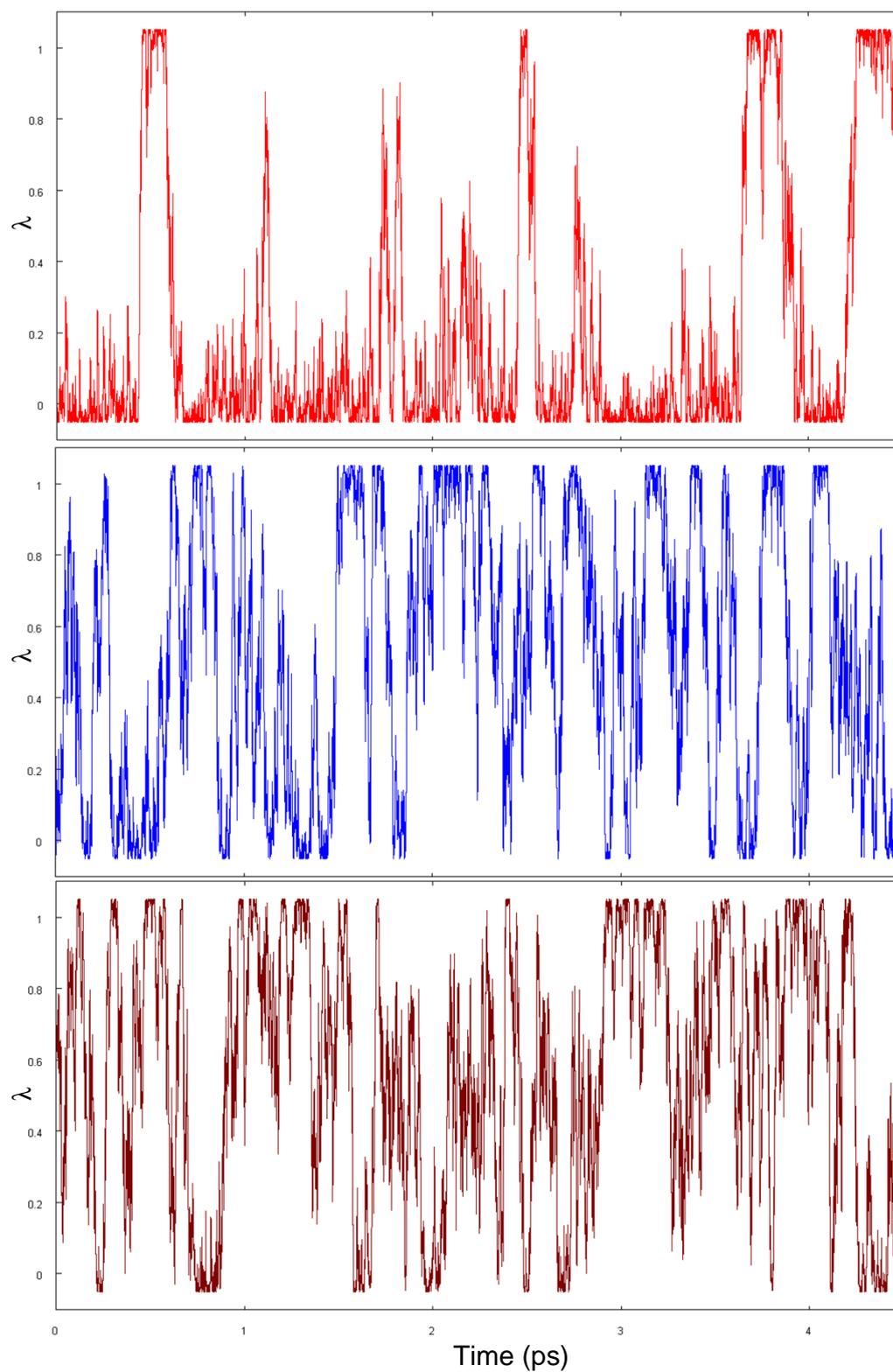


Figure 5.13: The λ -trajectories for the 3 glutamic acids embedded in the peptide in run 2 at a constant $pH = pK_a$.

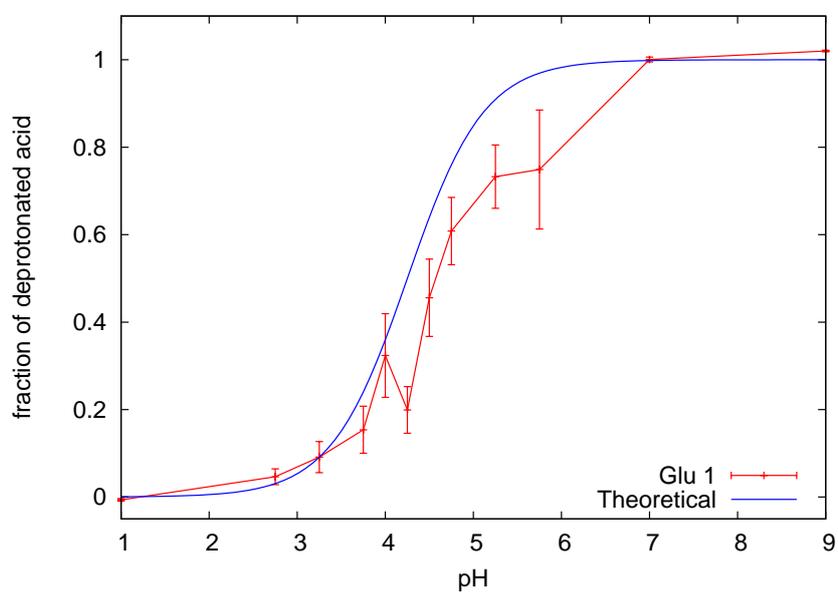


Figure 5.14: Titration curve for *Glu1* in run 2. The low probability for deprotonation around $pH = 4.25$ indicates an influence by the remaining glutamic acids.

Chapter 6

Summary

In this thesis an extended λ -dynamics approach allowing constant pH simulations was presented. While in classical molecular dynamics simulations the protonation states of all titratable groups have to be defined in advance, the protonation states of such groups are included as an additional coordinate that is dynamically updated. This dynamical updating is achieved by coupling two Hamiltonians describing the protonated and deprotonated state of the whole system via a coupling parameter λ . By assigning a mass and velocity to λ a virtual particle is introduced on which MD can be performed. The forces governing the dynamics of the λ -particle are the gradient of the free energy landscape with respect to λ . Constant pH simulations are then achieved by additionally accounting for the pH dependent hydration free energy of the vanishing proton to correct the free energy landscape.

Correction to the free energy landscape were required due to the force field offset error: The force field representation of bonds is not correctly modeling the quantum mechanical effects of bond formation as minimal energies stored in the bonds, but simply uses a harmonic approximation with zero minimal energy. Usually this does not impose any problems because free energy calculations are performed using a thermodynamical cycle, so only relative free energy differences are computed between branches that have the same force field error. Our approach allows the calculation of a single branch in the thermodynamical cycle by λ -dynamics. The force field offset errors are corrected using precalculated reference data provided as a parameter to the simulation.

After compensating for the force field effects the pH dependence was included by adding to the free energy landscape a function κ which includes the influences of the hydration free energy of the proton.

The method was implemented into the GROMACS MD simulation software and supports multiple titratable groups. Currently only glutamic acid was parameterized but other amino acids such as lysin or aspartic acid can be simulated as well by providing the corresponding parameters. Yet, the simulation of histidine is still under development due to the complexity of multiple titratable sites sharing the same atoms. A solution to this problem could be the definition of multidimensional λ particles.

A first test scenario for the constant pH simulations was the simulation of an isolated glutamic acid in a water box at different pH. By performing 12 runs at different pH values of $5ns$ each, a titration curve was derived by computing the average extent of protonation for each simulation. The sigmoidal behavior of the theoretical titration curve was qualitatively reproduced. Additionally it was shown, that best results were achieved by performing a series of simulations for each pH value with different starting velocity distributions.

As a next step, titration curves of systems including multiple titratable sites were calculated. The titration curve of a peptide with two glutamic acid showed the formation of a plateau where the relative pK_a of the second glutamic acid increased due to the charge imposed by the deprotonated first glutamic acid. The same effect could be observed in a peptide containing three glutamic acids directly attached to each other. Two plateaus were formed and a detailed analysis of the individual protonation curves showed the influence of the surrounding glutamic acids. Typically, two glutamic acids interact with each other by having in average one deprotonated and the other protonated while the third acid stays most of the time in the protonated state due to the charge imposed by the other two.

Current limitations are twofold. On the one hand residues with more than one titratable site are more difficult to model, as the sites can not be treated independently. The case of Histidine is therefore not yet solved but the use of multidimensional λ particles seems to be a promising approach. On the other hand the technical implementation in the GROMACS software could be greatly improved. The implementation was done introducing as little code changes as possible but the performance is not optimal. Especially the neighbor searching which has to be performed at each step for each group can be optimized to be calculated only every tenth step which in itself's fastens the simulation by one order of magnitude. Additionally all forces are calculated anew for each λ depending group, but it would suffice to only calculate the $\delta G/\delta\lambda$ at each step which would also lead to a speed gain of around an order of magnitude,

if multiple titratable groups are simulated.

In comparison to other methods allowing constant pH simulations, the presented approach has a series of advantages. After parameterizing the different titratable groups once, no additional effort is required to perform a constant pH simulation and a “black box” usage is possible. The results achievable with the method are encouraging and the additional computational costs are theoretically minimal. After an optimization the computational costs are below the operations required to simulate a few additional water molecules. While hybrid MC/MD simulations require additional free energy calculations or other complex analysis to determine the most probable protonation state, the method presented here is only introducing one more degree of freedom to the system for each titratable group. Compared to the constant pH method presented by Lee et al. [LJI04] the capability of treating explicit solvent has to be highlighted.

The future development will be the analysis and implementation of multidimensional λ to allow the treatment of e.g. histidine. Additionally the parametrization of all amino-acids in different force fields is required to distribute the software and to allow a wide application. With further technical optimizations the additional computational costs can be minimized allowing constant pH simulation to become a standard procedure when performing MD simulations.

Bibliography

- [And80] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384–2393, 1980.
- [Ant08] J. M. Antosiewicz. Protonation free energy levels in complex molecular systems. *Biopolymers*, 89(4)(9999):262–269, 2008.
- [BH01] U. Börjesson and P. H. Hünenberger. Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. *The Journal of Chemical Physics*, 114(22):9706–9719, 2001.
- [BKvG02] R. Bürghi, P. A. Kollman, and W. F. van Gunsteren. Simulating proteins at constant pH: An approach combining molecular dynamics and Monte Carlo simulation. *Proteins: Structure, Function, and Genetics*, 47(4):469–480, 2002.
- [BMP97] A. M. Baptista, P. J. Martel, and S. B. Petersen. Simulation of protein conformational freedom as a function of pH: constant-pH molecular dynamics using implicit titration. *Proteins: Structure, Function, and Genetics*, 27(4):523–544, 1997.
- [BPvG⁺84] 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(8):3684–3690, 1984.
- [BTS02] A. M. Baptista, V. H. Teixeira, and C. M. Soares. Constant-pH molecular dynamics using stochastic titration. *The Journal of Chemical Physics*, 117(9):4184–4200, 2002.

- [BvdSvD95] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, 1995.
- [CP88] D. Chandler and J. K. Percus. Introduction to Modern Statistical Mechanics. *Physics Today*, 41:114, 1988.
- [DA04] M. Dlugosz and M. Antosiewicz. Constant-pH molecular dynamics simulations: a test case of succinic acid. *Chemical Physics*, 302(1-3):161–170, July 2004.
- [DCK04] T.E. Cheatham III C.L. Simmerling J. Wang R.E. Duke R. Luo K.M. Merz B. Wang D.A. Pearlman M. Crowley S. Brozell V. Tsui H. Gohlke J. Mongan V. Hornak G. Cui P. Beroza C. Schafmeister J.W. Caldwell W.S. Ross D.A. Case, T.A. Darden and P.A. Kollman. *AMBER 8*. University of California, San Francisco, 2004.
- [FS96] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press Limited, 1996.
- [FTS⁺] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

- [HGE74] R.W. Hockney, S.P. Goel, and J.W. Eastwood. Quiet High-Resolution Computer Models of a Plasma. *Journal of Computational Physics*, 14:148–158, 1974.
- [HKvdSL08] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 2008.
- [HMV78] J. P. Hansen, I. R. McDonald, and Pieter B. Visscher. Theory of Simple Liquids. *American Journal of Physics*, 46(8):871–872, 1978.
- [KB05] J. Khandogin and III Brooks, C. L. Constant pH Molecular Dynamics with Proton Tautomerism. *Biophys. J.*, 89(1):141–157, 2005.
- [KI96] X. Kong and C. L. Brooks III. Lambda-dynamics: A new approach to free energy calculations. *The Journal of Chemical Physics*, 105(6):2414–2423, 1996.
- [LC03] G. Li and Q. Cui. PKa Calculations with QM/MM Free Energy Perturbations. *Journal of Physical Chemistry B*, 107(51):14521–14528, 2003.
- [LHvdS01] 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(8):306–317, 2001.
- [LJI04] M. S. Lee, F. R. Salsbury Jr., and C. L. Brooks III. Constant-pH molecular dynamics using continuous titration coordinates. *Proteins: Structure, Function, and Bioinformatics*, 56(4):738–752, 2004.
- [MC05] J. Mongan and D. A. Case. Biomolecular simulations at constant pH. *Current Opinion in Structural Biology*, 15(2):157–163, 2005.
- [MCM04] J. Mongan, D. A. Case, and J. A. McCammon. Constant pH molecular dynamics in generalized Born implicit solvent. *Journal of Computational Chemistry*, 25(16):2038–2048, 2004.
- [MGGM⁺85] J. B. Matthew, F. R. N. Gurd, B. E. Garcia-Moreno, M. A. Flanagan, Keith L. March, and S. J. Shire. pH-Dependent Pro-

- cesses in Proteins. *Critical Reviews in Biochemistry and Molecular Biology*, 18(2):91–197, 1985.
- [MP94] J. E. Mertz and B. M. Pettitt. Molecular dynamics at a constant pH. *Int. J. Supercomput. Appl. High Perform. Eng.*, 8(1):47–53, 1994.
- [MRR⁺53] 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(6):1087–1092, 1953.
- [Per78] M. F. Perutz. Electrostatic effects in proteins. *Science*, 201(4362):1187–1191, 1978.
- [RPMP03] D. Reith, M. Pütz, and F. Müller-Plathe. Deriving Effective Mesoscale Potentials from Atomistic Simulations. *Journal of Computational Chemistry*, 24:1624–1636, 2003.
- [RSC05] D. Riccardi, P. Schaefer, and Q. Cui. PKa Calculations in Solution and Proteins with QM/MM Free Energy Perturbation Simulations: A Quantitative Test of QM/MM Protocols. *Journal of Physical Chemistry B*, 109(37):17715–17733, 2005.
- [SCW97] Y.Y. Sham, Z.T. Chu, and A. Warshel. Consistent Calculations of pKa's of Ionizable Residues in Proteins: Semi-microscopic and Microscopic Approaches. *Journal of Physical Chemistry B*, 101(22):4458–4472, 1997.
- [Sim02] T. Simonson. Gaussian fluctuations and linear response in an electron transfer protein. *Proceedings of the National Academy of Sciences*, 99(10):6544–6549, 2002.
- [SLH01] D. v. d. Spoel, E. Lindahl, and B. Hess. *GROMACS User Manual*. University of Groningen, 3.3 edition, 2001.
- [SLH⁺05] D. v. d. 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(16):1701–1718, 2005.
- [Sop96] A. K. Soper. Empirical potential Monte Carlo simulation of fluid structure. *Journal of Chemical Physics*, 202:295–306, 1996.

- [TTB⁺98] G. J. Tawa, I. A. Topol, S. K. Burt, R. A. Caldwell, and A. A. Rashin. Calculation of the aqueous solvation free energy of the proton. *The Journal of Chemical Physics*, 109(12):4852–4863, 1998.
- [TV77] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.
- [War79] A. Warshel. Calculations of chemical processes in solutions. *Journal of Physical Chemistry*, 83(12):1640–1652, 1979.
- [WD08] E. Weinan and L. Dong. The Andersen thermostat in molecular dynamics. *Communications on Pure and Applied Mathematics*, 61(1):96–136, 2008.
- [WSK85] A. Warshel, F Sussman, and G. King. Free energy changes in solvated proteins: microscopic calculations using a reversible charging process. *Biophysics*, 25:8368–8372, 1985.
- [Zwa54] R. W. J. Zwanzig. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics*, 22(8):1420–1426, 1954.

Chapter 7

Appendix

Periodic boundary conditions

All atoms simulated in a MD simulation are contained in a box which is a cubic box in the simplest case. To overcome the boundary problem of treating atoms leaving the box or forces acting behind the box, periodic boundary conditions are introduced. Hereby the box is in all dimensions surrounded copies of itself as illustrated in Figure 7.1 and atoms leaving the box on one side will appear immediately on the other side.

In some simulations this is very helpful, e.g. in simulating a membrane which has then no arbitrary borders, but in others, for example protein or small system simulations, this can lead to unwanted artifacts by the molecule seeing itself although this was not intended. If an isolated system should be analyzed, choosing a sufficient box size is important to overcome these often not easy to find errors. Additionally the geometrical shape can be optimized to achieve high distances to the image of the protein itself while not being required to

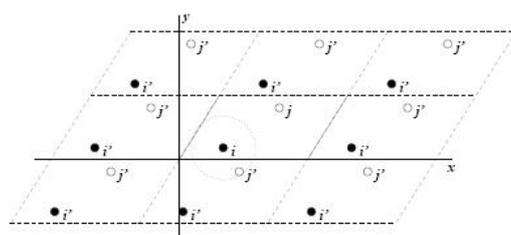


Figure 7.1: Periodic boundary conditions in two dimensions for a triclinic box (Figure from [SLH01])

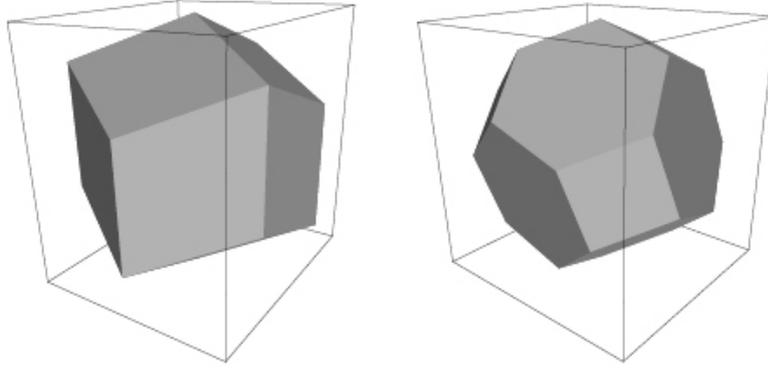


Figure 7.2: The rhombic dodecahedron and a truncated octahedron (Figure from [SLH01])

simulate too many water molecules.

The simple cubic box has for a spherical molecule with a maximal interaction distance d_1 superfluously water-filled corners. For such simulation setups the rhombic dodecahedron as illustrated in Figure 7.2 is optimal: It uses with the same d_1 only 71% of the size and saves therefore about 29% of CPU-time. For the final simulation the image distance d should be $d > d_1$.

Thermo- and barostats

To be able to simulate at constant temperature or pressure (e.g. to simulate an NPT ensemble) these thermodynamical properties need to be maintained by a thermo- or barostat.

Berendsen thermostat

The *Berendsen-Thermostat* [BPvG⁺84] couples the system to an infinite external heat bath with temperature T_{ref} . The heat flow into or out of the system is implemented by scaling the velocities with a factor λ given by:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T} \left\{ \frac{T_{ref}}{T(t - \frac{\Delta t}{2})} - 1 \right\} \right]^{1/2} \quad (7.1)$$

The parameter τ_T is a temperature coupling parameter that translates to $\tau = 2c_v\tau_T/N_{df}k$ for the exponential temperature deviation with constant τ :

$$\frac{dT}{dt} = \frac{T_{ref} - T}{\tau} \quad (7.2)$$

c_v is the total heat capacity of the system, k is the Boltzmann's constant and N_{df} the number of degrees of freedom.

Andersen thermostat

The *Andersen Thermostat* [And80] (for a detailed analysis on extended properties see [WD08]) couples a system to a reference heat bath by selecting new velocities from a Maxwellian velocity distribution. The updating is governed by a coupling parameter τ determining the collision frequency. In every coupling step the velocity is updated if a random number ran from an uniform distribution holds $ran \leq \tau \cdot dt$.

This behavior mimics collisions of the system's atoms with atoms from the heat bath and reproduces an exact Maxwell-Boltzmann distribution independent of the value of the collision frequency τ (see [FS96]). By doing so, the momentum is not conserved and simulating dynamics is difficult due to the fact, that consequently the mean square displacement as a function of time changes for different collision frequencies.

Berendsen barostat

Analogously to the thermostat the pressure is scaled in the *Berendsen Barostat* [BPvG⁺84] accordingly to

$$\frac{d\mathbf{P}}{dt} = \frac{\mathbf{P}_{ref} - \mathbf{P}}{\tau_P} \quad (7.3)$$

which reflects the coupling of system pressure to a 'pressure bath' realized by scaling the coordinates and box vectors every step.

Free energy calculations

Slow-growth

The standard way in GROMACS to perform free energy calculation is to use slow-growth methods. The system is hereby perturbed slowly from state A to state B . Slowness is important to ensure that every state is thermodynamically equilibrated. If that requirement is fulfilled, the process is reversible and a simulation from B to A yields the same results except for the sign. Therefore analysis of the hysteresis is the building block for most slow-growth error

estimates. If the simulation fulfills these requirements integrating over $\delta G/\delta\lambda$ over $\delta\lambda$ yields ΔG .

$$\Delta G = G_{(p,T)}^B - G_{(p,T)}^A = \int_0^1 \frac{\delta H}{\delta\lambda}_{N,p,T;\lambda} \delta\lambda \quad (7.4)$$

Multiple conformation TI

While in Slow-Growth one simulation is performed changing the λ in such small steps that equilibrium is assumed, the thermodynamic integration of multiple conformation consists of a set of single simulations. Hereby the interval $[0 : 1]$ is divided into N parts (e.g. 20) and for each λ a separate simulation of for example shorter length is performed. A linear regression over the averaged $dG/d\lambda$ values allows the final integration to derive ΔG . Usually the first part of the simulation is dropped for equilibration purposes whereby the equilibration time depends heavily on the system's complexity. An error estimation can be done using block averaging algorithms which separate the coupled $\delta G/\delta\lambda$ in uncoupled blocks of minimal size so that standard error estimators as standard deviation can be applied over the block's averaged values.

Gibb's free energy instead of Helmholtz free energy

Often the for chemists generally more useful Gibb's free enthalpy G instead of the Helmholtz free energy F is computed in free energy calculations. Whereas F is related to an equilibrium MD simulation generating an N, V, T ensemble with constant atom numbers, constant volume and constant temperature, the Gibb's free enthalpy G is related to the N, p, T ensemble under constant pressure instead of volume. For the purpose of MD simulations F and G will be used interchangeable for the following reasoning:

$$G = U + p \cdot V - T \cdot S = F + p \cdot V \quad (7.5)$$

$$dG = dF + dp \cdot V + p \cdot dV = dF + 0 + p \cdot dV \approx dF \quad (7.6)$$

The term $p \cdot dV$ is usually negligible due to the very small influence of pressure scaling resulting in volume changes (see [SLH01]). The growth of a water molecule from nothing in a bath of 1000 water molecules would produce an additional pressure of 22 bar and would result to an energy correction of -20 J/mol. Keeping in mind that errors in such free energy calculations are usually between 5 kJ/mol and 10 kJ/mol an influence of -20 J/mol is orders of magnitudes below the simulation's accuracy.