GMCT : A Monte Carlo Simulation Package for Macromolecular Receptors

R. Thomas Ullmann[a]* and G. Matthias Ullmann[a]*

Generalized Monte Carlo titration (GMCT) is a versatile suite of computer programs for the efficient simulation of complex macromolecular receptor systems as for example proteins. The computational model of the system is based on a microstate description of the receptor and an average description of its surroundings in terms of chemical potentials. The receptor can be modeled in great detail including conformational flexibility and many binding sites with multiple different forms that can bind different ligand types. Membrane embedded systems can be modeled including electrochemical potential gradients. Overall properties of the receptor as well as properties of individual sites can be studied with a variety of different Monte Carlo (MC) simulation methods. Metropolis MC, Wang-Landau MC and efficient free energy calculation methods are included. GMCT is distributed as free open source software at www.bisb.uni-bayreuth.de under the terms of the GNU Affero General Public License. © 2012 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.22919

Introduction

Generalized Monte Carlo titration (GMCT) is a versatile package of computer programs for the simulation of complex receptor systems with Monte Carlo methods. The primary targets of the software are biomolecular receptors like proteins that bind or transfer protons, electrons or other small-molecule ligands. The software is, however, also useful for studying polyelectrolytes in a broader sense or other systems like Ising or Potts models. The properties of large systems can be studied using a variety of simulation methods. This makes GMCT ideally suited to study the thermodynamics of ligand binding and charge transfer processes in bioenergetic complexes and other complex biomolecular systems.

The theoretical basis of the software is a general, chemical potential based formulation of ligand binding theory that extends earlier work.[1–16] This formulation results in an energy function which facilitates understanding and interpretation of simulation results. A schematic view of the receptor model used in GMCT with all possible features is shown in Figure 1. All these features enable the software to simulate a realistic receptor model that accounts for intrinsic properties of the receptor and for realistic conditions that can be matched to the conditions in a natural or experimental environment.

There is a variety of Monte Carlo simulation software for molecular mechanics models in continuous coordinates with capabilities that are comparable to those of molecular dynamics codes.[17, 18] Equivalent programs for simulations of models that are defined on discrete configuration spaces like the receptor model of GMCT have not been available. Such models are widely used in the study of proton binding and reduction processes that occur in biomolecules.[19–23] There are several programs for Metropolis Monte Carlo Simulations based on similar physical models: WhatIf,[6, 24] DOPS,[8, 22, 25] MCCE,[26, 27] Karlsberg,[28] MCTI,[4] Petite,[29] and Monte.[30] GMCT offers additional methods and an extended range of applications. Unique features of GMCT are efficient free energy calculation methods that can be used to calculate free energy differences for freely definable transformations and for the calculation of free energy measures of cooperativity. Namely, free energy perturbation,[31, 32] thermodynamic integration,[33] the non-equilibrium work method[34] and the Bennett acceptance ratio method[35] have been implemented. Further features that have not been previously available are the possibility to account for electrochemical potential gradients[36, 37] and the possibility to study global receptor properties with Wang-Landau Monte Carlo. All simulation methods have an analytical counterpart that can be used for validation purposes on sufficiently small systems.

GMCT is open source software written in C/C++. GMCT was originally developed for Linux systems but should be compilable on any system with a standard C/C++ compiler. The programs are frugal in terms of hardware requirements but can of course profit from modern hardware. GMCT was optimized in many parts for simulation performance and minimum memory requirements. Especially high priority was attributed to the aim to make GMCT flexible and user friendly.

The next section describes the energy function utilized by GMCT and gives a brief introduction to its theoretical basis. The theory section is followed by an overview of the design and usage of the programs. Afterwards, a description of the simulation methods implemented in GMCT is given. These simulation methods can be used for the calculation of various properties

[a] *R. Thomas Ullmann, G. Matthias Ullmann*
Structural Biology/Bioinformatics, University of Bayreuth, Universitätsstr. 30, BGI, Bayreuth 95447, Germany
E-mail: thomas.ullmann@uni-bayreuth.de; matthias.ullmann@uni-bayreuth.de

Contract/grant sponsor: Deutsche Forschungsgemeinschaft; Contract/grant number: SFB 840 research project B2

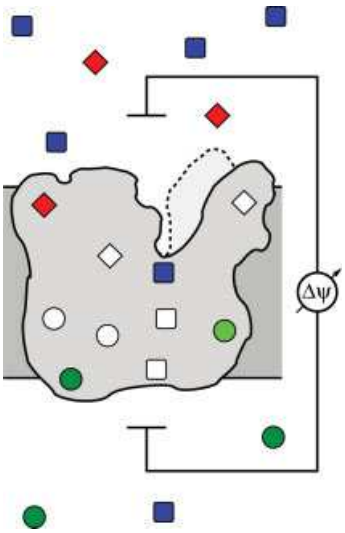

Figure 1. The conceptual model of a macromolecular receptor in GMCT with all possible features: The receptor (central region) can contain multiple sites of different types as depicted by the differently shaped symbols. The sites may bind one or more ligands of differing types from the surrounding solution and may have multiple conformations or charge forms indicated by varying color shades. The receptor is surrounded by a solution containing the different ligand types at known electrochemical potentials. The ligand types are depicted by the differently shaped symbols in the solvent regions (upper and lower regions). Optionally, the receptor may be embedded in a membrane which divides the solution in two compartments. Each site is connected to only one of the compartments. The chemical potential of the ligands and the electric potential may differ between the compartments. That is, there may be electrochemical gradients across the receptor-membrane system as symbolized by the differing numbers of ligands in the solvent regions and the voltmeter. The receptor may also have several global conformations, e.g., due to a movement of protein domains relative to each other as depicted by the lobe in broken lines and lighter shade. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

that can aid in the analysis and understanding of the simulated systems. Possible applications of the simulation methods are discussed and illustrated with examples from our own work. Special emphasize is thereby put on issues arising in the practical use of the methods with GMCT. We close with some conclusions regarding possible applications of GMCT.

Theory

This section contains a brief description of the theoretical basis for the treatment of binding equilibria within the program suite GMCT. The theoretical framework behind GMCT extends earlier work considering protonation and redox equilibria in our[12, 13, 15, 16] and other groups.[1–11, 14] We briefly introduce the most important quantities of the formalism and the conceptual model behind the energy function used in GMCT.

The microstate description of the receptor-ligand system

The conceptual model of GMCT is based on a microstate description of the receptor-ligand system,[14, 15, 23, 38] while the surrounding ligand solution is described by electrochemical potentials in a mean field approach.[33, 39] Each microstate is defined by a set of definite values for all explicitly considered variables. For classical systems, momentum dependent terms cancel from free energy differences and thus do not need to be considered explicitly.[40] In this case, the microstate of the system is fully defined by its configuration.

A schematic view of the receptor model is depicted in Figure 1. The receptor can exist in several global conformations whose number is given by N^{confs}. A subdivision of the receptor into a largely invariable background (e.g., the protein backbone in a defined global conformation) and multiple variable sites (e.g., protonatable amino acid sidechains) completes the description of the system configuration

$$\mathbf{q} = \{c, \vec{s}\} \tag{1}$$

Here, c is the global conformation currently adopted by the background and the N^{sites} elements of the state vector $\vec{s}$ specify the specific form adopted by each site. The different forms of a site are termed instance in GMCT. An instance is characterized by the number of bound ligands of each type, the spatial coordinates of the atoms and the charge distribution. With this description of the sites one can account for example for multiple redox and/or protonation forms, multiple tautomeric forms, multiple sidechain rotamers and other sources of variability the user might want to represent.

The partitioning of the protein into background and individual sites and the representation of the sites by a set of instances must be made such as to comply with the assumption of pairwise interactions. That is, the partitioning of the receptor must be chosen such that any change in a receptor constituent has only a negligible effect on the electronic configuration of the other receptor constituents. Thus, changes in the electronic configuration must be largely restricted to a single receptor constituent. If, for example, the protonation behavior of a histidine sidechain is to be considered, all of its protonation forms have to be included in one site instead of partitioning the histidine sidechain into two sites each including one of the protonatable nitrogen atoms (see Fig. 2).

A subset of the microstates, called substate in GMCT, can be selected on the basis of an observable. This observable might, for example, be the total number of bound ligands of a certain type, a certain number of ligands bound to a particular site of the receptor, a certain conformation of the receptor or any other criterion.

Different models for the practical implementation of the microstate description have been described in the literature.[2, 6–9, 12, 23, 24, 27–29, 41] Most of these models are based on continuum electrostatics. Some of them use a combination of continuum electrostatics and molecular mechanics. The models differ in the degree of detail which is used in the description of the sites and various other aspects. GMCT leaves the choice of a specific model to the user.

The microstate energy function

The concept behind our microstate energy function can be interpreted as a thought, stepwise assembly process of the receptor-

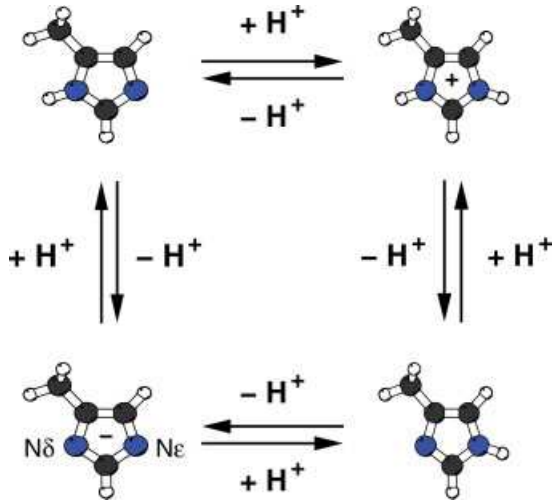

Figure 2. The microscopic protonation forms of a histidine sidechain. One proton can be bound at the δ nitrogen atom and one proton can be bound at the ϵ nitrogen atom indicated for the fully deprotonated form at the bottom left. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

ligand system from its constituents in a certain microstate. The receptor is constituted by the background and a number of sites. Each receptor constituent *i* is taken in a particular form from a reference environment and transferred into the receptor environment.

The electrochemical potential of the receptor constituent *i* in this particular form relative to that of some reference form of the constituent is termed intrinsic energy[5]

$$E_i^{\text{int}} = \bar{\mu}_i^{\circ} - \bar{\mu}_i^{\circ,\text{ref}} \tag{2}$$

The intrinsic energy of the constituent in the receptor environment, $E_i^{\text{int,r}}$, is computed relative to the intrinsic energy of the model compound in a reference environment.[2,42] The shift of the intrinsic energy upon transfer of the model compound from the reference environment to the receptor environment is given by

$$\Delta E_i^{\text{int}} = E_i^{\text{int,r}} - E_i^{\text{int,m}} \tag{3}$$

The intrinsic energy of this form in the reference environment, $E_i^{\text{int,m}}$, is called model energy, because it is known from experiments or calculations on model compounds in a reference environment. The calculation of the model energy from experimental data is addressed in the user manual of GMCT. The calculation of the model energy with quantum chemical methods is addressed in Ref. [16]. The model energy contains all energy contributions due to the formation or breaking of covalent bonds within a site. Energy contributions due to the covalent bonding of the receptor constituent to other receptor constituents, e.g., bonding of a protonatable amino acid to the neighboring residues in a protein) are assumed to be equal for all instances of a site. Thus, these energy contributions cancel out from relative intrinsic energies between two forms of the constituent and have no influence on the receptor behavior. The intrinsic energy in the receptor environment accounts for the interaction with all invariable parts of the system, i.e., for the interaction with the background and the receptor environment. The interactions among sites are treated separately. The receptor behavior depends on all variables of the system, that is on the global conformation and on the instance adopted by each site.

The energy of the microstate $\boldsymbol{n}$ is a function of its system configuration $\mathbf{q}_n$[5,11,16]

$$E_{\boldsymbol{n}}^{\text{micro}} = E_c^{\text{conf}} + \sum_{i=1}^{N^{\text{sites}}} \left(E^{\text{int,r}}(c, \boldsymbol{s}_i) - \sum_{m}^{\mathfrak{L}} \nu_m(c, \boldsymbol{s}_i)\bar{\mu}_m \right) + \sum_{i=1}^{N^{\text{sites}}} \sum_{j=1}^{j<i} W(c, \boldsymbol{s}_i, \boldsymbol{s}_j) \tag{4}$$

where the state vector elements $\boldsymbol{s}_i, \boldsymbol{s}_j$ are the instances occupied by site *i* and site *j*, respectively. The first sum runs over all N^{sites} sites of the system. The second sums run over all pairs of sites in their currently occupied instances. $W(c, \boldsymbol{s}_i, \boldsymbol{s}_j)$ is the interaction energy of instance $\boldsymbol{s}_i$ of site *i* with instance $\boldsymbol{s}_j$ of site *j* for the global conformation *c* adopted by the system. Equation (4) describes the partitioning of the receptor into background and sites and reflects its stepwise assembly from background and sites including the bound ligands.

The assembly process starts with the insertion of the background adopting the global conformation *c* into the receptor environment. The energy cost E_c^{conf} of this process is given relative to one of the global conformations chosen as reference. That is, the conformational energy is equivalent to the intrinsic energy of a site, except that no other receptor constituents are present yet when adding the background to the system. Thus, the global conformational energy does not need to be specified if the receptor is modeled with one global conformation only.

The assembly process continues with the successive addition of the sites. Each site adopts a particular instance $\boldsymbol{s}_i$. The intrinsic energy of the added site $E^{\text{int,r}}(c, \boldsymbol{s}_i)$ is shifted relative to the intrinsic energy of the model compound by the exchange of the reference environment against the receptor environment and by the additionally gained interaction with the background.

The microstate energy is also influenced by the energy cost of removing the ligands bound to each site from the surrounding solution. The magnitude of this energy cost is determined by the number of ligands bound from the solution and by the electrochemical potential of the ligands in solution. The number of ligands of type *m* bound to site *i* is given by the stoichiometric coefficient $\nu_m(c, \boldsymbol{s}_i)$, which is specific to the global conformation of the receptor *c*, and to the instance $\boldsymbol{s}_i$ adopted by the site. The electrochemical potential of ligand type *m* is denoted by $\bar{\mu}_m$. The ligands might be bound from two separate compartments with different electrochemical potential. In this way, electrochemical gradients for the ligands between two compartments separated by a membrane can be accounted for.[36,37] The intrinsic energies, the global conformational energies and the interaction energies together determine how favorable the binding of the ligand to the site is. Thus, each site can be perceived as a small receptor

on its own whose energetics is influenced by the other receptor constituents and the receptor environment.

The last term on the right hand side of eq. (4) sums up the interactions with the previously added sites $j < i$, each of which adopts a particular instance s_j. Besides the global conformational energy, also the intrinsic energies and the interaction energies can be altered if the global conformation of the receptor is changed, because the spatial arrangement of explicitly and implicitly treated components of the system can change.

GMCT uses a uniform system of units to provide clarity and to facilitate understanding of the system response to parameter changes. In GMCT, all chemical and electrochemical potentials are given in kcal/mol, which is the commonly used unit in chemistry. Consistently, also energies are specified as molar energies in kcal/mol. All other quantities are expressed in SI units.

Program Usage and Design

The GMCT suite consists of multiple programs for specific calculation tasks. All programs share the same functions for the reading of input files and the same data structures for storing the input data. The user manual contains detailed descriptions of the separate programs, their purpose and usage including all program parameters. In the following, we will give an overview over the necessary input files and the generated output of a calculation with GMCT. Detailed descriptions of the file contents and formats can be found in the user manual. In addition, the GMCT distribution contains multiple examples for calculation setups including all the necessary input files and scripts with the commands necessary to run the calculations. An overview over the function of the programs with the necessary input files and the generated output is shown in Figure 3.

A user-defined "*name*" for the receptor is used as common prefix for files that contain information about the whole system. All input files and most output files are free-format text files which facilitates editing, verification and further processing the input files can contain comments for documentation. In some cases, there are multiple options for the user to state the properties to be calculated, where each option is especially easy to use for a particular purpose. Optionally selectable program features enable the automated setup of typical calculations.

Program input

The program parameters are defined in *name*.setup. The program parameters comprise the physical conditions and necessary simulation parameters. The physical conditions that can be specified are the temperature, the ranges of the chemical potentials of each ligand type to be considered and the electrostatic membrane potential. Examples for the large number of simulation parameters that can be specified are the numbers of MC scans to be used in a simulation for equilibration and production.

The microscopic receptor properties are specified in multiple files as follows: The file *name*.conf contains a list of all global conformations of the receptor. For each global conformation, the file specifies a user defined name and the properties of the receptor background in the corresponding global conformation.

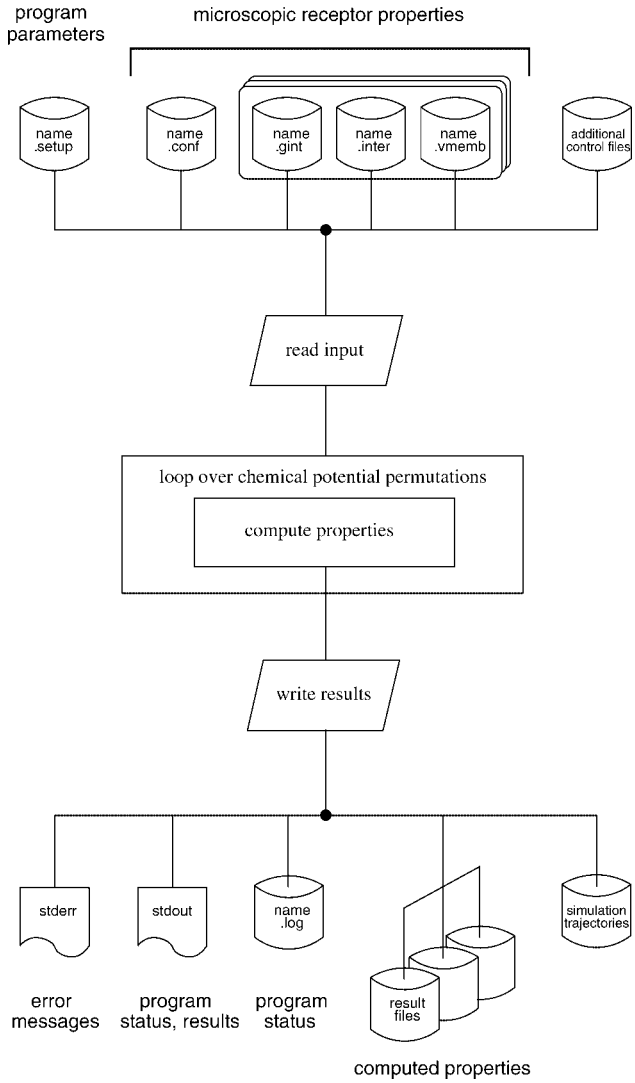

Figure 3. Overview flowchart of the programs of the GMCT suite.

The properties considered for each global conformation are the conformational energy not including the transmembrane potential dependent contributions[39] and the transmembrane potential dependent contributions. These contributions are the interaction energy of the background with the electrostatic transmembrane potential normalized to a membrane potential of +1 V and the capacitance of the receptor-membrane system. For each global conformation, a directory with the name of the global conformation is expected to exist. This directory contains three files with the site specific receptor properties. The file *name*.gint contains the intrinsic energies of each instance of the receptor sites except for the contribution due to the interaction of the site with the electrostatic transmembrane potential. The number of ligands of each ligand type bound to the instances of the sites are also stated in the file *name*.gint. The interaction energies of the instances of each site with the electrostatic transmembrane potential,[39] normalized to a membrane potential of +1 V, are defined in *name*.vmemb. The site-site interaction energies for each pair of instances of the sites are defined in *name*.inter. All properties related to the

electrostatic transmembrane potential are only needed if there is a non-zero electrostatic transmembrane potential.

A program for the automated calculation of the microscopic receptor properties and the preparation of the corresponding input files for GMCT is provided at www.bisb.uni-bayreuth.de. This program uses an extended version of the MEAD library[23,43] to compute the microscopic receptor properties from a continuum electrostatics/molecular mechanics model and will be subject of a separate publication.[44] Currently, most of the theory underlying this program is described in the user manual of GMCT. Examples for the practical use of the program with the necessary input files and commands to run the program are included with the program distribution.

Additional control files can be used for multiple purposes. An example of such a purpose is the specification of substates whose occupation is to be monitored during the simulations or that constitute the end states of transformations.

Program output

During the calculations, the programs give out status information and some results to stdout. More status information, the parameters used in the calculation and and information about program internals are written to a logfile named *name*.log. The amount of information given out to stdout and the logfile about running calculations can be adjusted by the user between almost none (for routine calculations) and very verbose (for analysis and trouble shooting). Error messages are printed to stderr.

The requested properties are computed for each permutation of chemical potential values. The chemical potential values, the electrostatic transmembrane potential and the properties computed under these conditions are written to result files that can be used for plotting or further analysis. The format and the contents of the output files are described in the user manual.

Optionally, simulation trajectories can be written to compressed binary files for further analysis. GMCT makes use of the free open source library zlib (http://zlib.net) to read and write the compressed trajectory files.

Monte Carlo Simulation Methods in GMCT

Monte Carlo (MC) simulations make it possible to investigate complex systems that cannot be treated by analytical methods within acceptable computation time.[45–48] Due to their stochastic nature, MC methods require a reliable source of random numbers.[49] The default random number generator used in GMCT is the Mersenne Twister,[50] because it offers very good performance in terms of speed and quality of the generated random numbers.

The simulation methods implemented in GMCT are briefly introduced in the following parts of this section. For more details, the interested reader is referred to the cited references and to the user manual. The user manual contains a detailed description of the simulation methods including implementation specific details and some theoretical background information.

The next section introduces two example systems that will be used in the following sections to demonstrate possible applications for these methods and practical issues in the use of GMCT. Currently, GMCT features two basic MC simulation techniques that have utility in different applications. The sections "Metropolis Monte Carlo" and "Wang-Landau Monte Carlo" introduces their applications within GMCT. Free energy calculation methods that build on Metropolis MC and their applications are addressed in the "Free Energy Calculations with the Program Suite GMCT" section. The "Cooperativity Measures" section introduces cooperativity measures that can be calculated with GMCT to detect and quantify thermodynamic coupling between events in a molecular system.

Example systems

The simulation methods available in GMCT will be introduced at two example systems of low and moderate complexity, respectively. These example systems are chosen for ease of understanding but do not require all capabilities of GMCT. Despite their limited complexity, these example systems exhibit interesting behavior that is well suited to demonstrate the usefulness of our programs for studying and understanding complex ligand binding systems.

DTPA. The first example system is diethylene-triaminepentaacetate (DTPA). DTPA is a protonatable molecule with a very interesting, strongly irregular titration behavior. The chemical structure of DTPA is shown on the top of Figure 4a. DTPA contains three nitrogen atoms that can bind or release a proton in a range of pH values from 3 to 13 and five carboxyl groups that are always deprotonated in this range.[51] Experimental protonation probabilities in dependence on the pH value are available from nuclear magnetic resonance measurements in this range of pH values.[51,52] A model of this receptor molecule obeying the microstate energyfunction eq. (4) was fitted to closely reproduce the titration behavior in the given range of pH values.[53] The model system possesses a single global conformation and consists of three sites where each site denotes one of the titratable nitrogen atoms in DTPA. The intrinsic energies of the deprotonated instances of each site is set to zero by convention. The intrinsic energies of the protonated instances are expressed relative to those of the deprotonated instances. The intrinsic energies are calculated from the intrinsic pK_a values given in Ref. [53] via $E_i^{\mathrm{int,m}} = -\beta^{-1} \ln 10\, \mathrm{p}K_{a,i}^{\mathrm{int}}$. The protonated instances of the terminal nitrogen atoms have intrinsic energies of -14.6 kcal/mol while that of the central nitrogen atom has an intrinsic energy of -15.3 kcal/mol. The interaction energies between the neutral deprotonated instances and between protonated and deprotonated instances of a pair of sites are set to zero. The interaction energies between the positively charged, protonated instances of the sites are taken from Ref. [53]. The interaction energy between the two terminal sites is equal to 2.2 kcal/mol while the interaction energy of the central site with each of the terminal sites is 4.4 kcal/mol.

Plots of the protonation probabilities of the central and terminal nitrogen atoms in dependence on the pH value of the

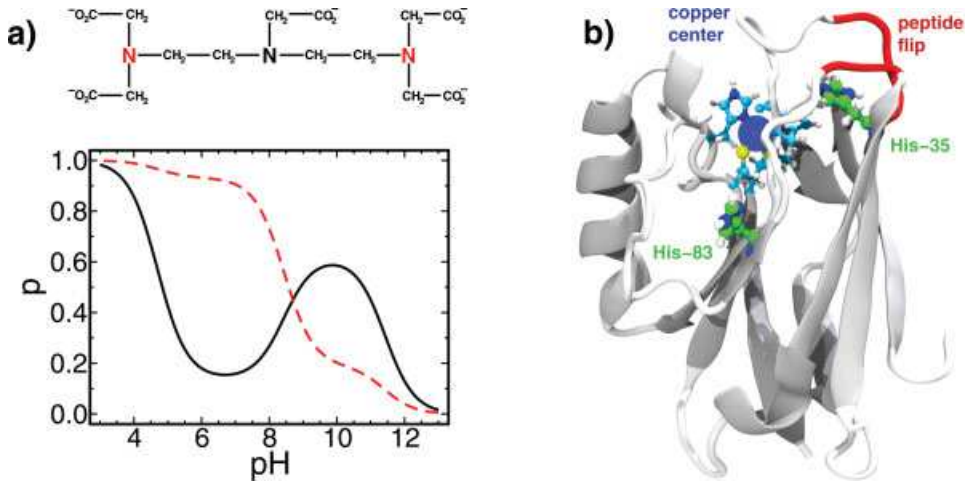

Figure 4. a) The triprotic acid DTPA as example system. DTPA has three protonatable nitrogen atoms. The five carboxylic groups are deprotonated in the range of pH values between 3 and 13. Top: chemical structure of DTPA Bottom: protonation probability of the central nitrogen atom (black solid curve) and of the two equivalent terminal nitrogen atoms (red broken curve) as a function of the pH value. b) The small electron carrier azurin from *Pseudomonas aeruginosa* (PaAz) as example system taken from Ref. [54]. The type-1 copper center is formed by the copper ion shown as blue sphere and coordinating protein sidechains. The reduction of the copper center is coupled to the protonation of amino acid residues through electrostatic interactions, leading to a pH dependent reduction potential. The reduction is most strongly coupled to the protonation of His-35 labeled in the figure. The protonation of this residues is in turn coupled to a conformational rearrangement in the region highlighted in red that involves the flip of a peptide bond. The figure was prepared with VMD[55] and Tachyon.[56]

solution are shown in Figure 4a. The central nitrogen atom shows a highly irregular, nonmonotonic titration curve due to the strong interaction with the other sites and due to the similar intrinsic energies of all three sites.[38] The small size of the model system allows an analytical calculation of all quantities for comparison with the results of simulations. The comparison can give a feeling for the accuracy that can be expected of the simulation results for a typical, affordable amount of computational effort.

Azurin from *Pseudomonas aeruginosa*. While DTPA will be used for most demonstration purposes, a second, larger system will be used to give motivating examples of applications to complex ligand binding systems. This example system is the electron transport protein azurin from *Pseudomonas aeruginosa* (PaAz) depicted in Figure 4b. PaAz is an example system for studying the coupling of protonation and reduction. In the context of GMCT, PaAz provides an example for the coupling of the chemical potentials of multiple ligand types. That is, binding free energies for one ligand type, like the reduction potential of PaAz, generally depend on the chemical potential of all ligand types. The blue copper center of PaAz can be reduced by one electron while the titratable protein residues can bind protons. The reduction potentials of PaAz and other redox-active proteins consequently depend on both, pH value and reduction potential (protonic and electronic potentials) of the solution.[13] Details of the parameters and simulation methods used in the calculations on PaAz can be found in Ref. [54]. This work contains a detailed analysis of the coupling of protonation, reduction and conformational change in PaAz.

Metropolis Monte Carlo

In this section, we discuss specific details of Metropolis MC[57] simulations with GMCT and give some application examples. The application of Metropolis MC to titration problems similar to those considered here was pioneered by Beroza et al.[4] The method allows the treatment of such large systems as the bacterial photosynthetic reaction center, cytochrome *c* oxidase and cytochrome bc_1.[4,58–64]

General principle and possible applications. A Metropolis MC[57] simulation generates a sequence of microstates starting from a randomly selected microstate. The respective next microstate of the sequence is obtained by attempting a randomly selected MC move from a set of possible moves. The MC move can involve a change in one or more variables of the configuration **q**. In the simplest case, the MC move changes the instance occupied by one site, where the site and the new instance are chosen randomly. The move is accepted or rejected according to the Metropolis criterion. The Metropolis criterion ensures the importance weighting of the microstates, i.e., by their Boltzmann factor $\exp[-\beta E^{\text{micro}}]$. The length of an MC simulation is measured in MC scans. One MC scan consists of a number of MC steps that ensures that there is on average at least one attempt per MC scan to change the value of each variable. The sequence of visited microstates recorded after each scan is called trajectory.

A trajectory from Metropolis MC can be used directly to compute thermodynamic averages of observables as

$$\langle O \rangle = \frac{1}{N^{sa}} \sum_{n=1}^{N^{sa}} O_n \qquad (5)$$

where N^{sa} is the number of steps in the trajectory. The typical application for Eq. (5) in GMCT is the calculation of titration curves (see Fig. 5).

In principle, free energy changes between two substates can be calculated from a Metropolis MC trajectory. In practice, however, it is very difficult to obtain reasonable statistical accuracy

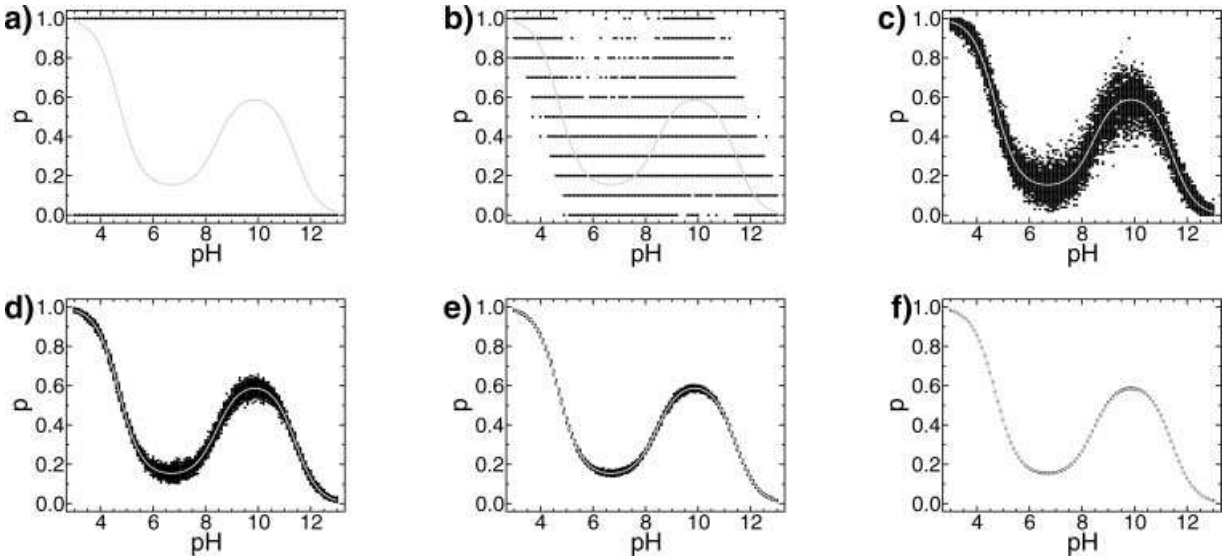

Figure 5. Protonation probability of the central nitrogen atom of DTPA as a function of the pH value computed from Metropolis MC simulations of different length. The data spread due to the statistical uncertainty is indicated by plotting 100 data points from independent simulations for each pH value. The analytically calculated protonation curve is shown as solid gray curve for comparison. a) 1 MC scan, b) 10 MC scans, c) 100 MC scans, d) 1000 MC scans, e) 10,000 MC scans, f) 100,000 MC scans.

with this approach if one or both occupation probabilities are very small, which will always be the case if $|\Delta G|$ is significantly larger than the thermal energy β^{-1}. More sophisticated free energy simulation methods that build on the Metropolis MC algorithm are available in GMCT (see "Free Energy Calculations with the Program Suite GMCT" section). Another possibility for the calculation of free energy differences with GMCT is the Wang-Landau MC method presented in "Wang-Landau Monte Carlo" section.

Example applications. Equation (5) can for example be used to calculate protonation probabilities of protonatable sites in a receptor from a Metropolis MC simulation. Figure 5 shows the protonation probability of the central nitrogen atom of DTPA computed with Metropolis MC simulations of different length. For each pH value, the protonation probabilities from 100 independent simulations are overlaid on the analytical titration curve. The results of an MC simulation will be inevitably afflicted by the statistical uncertainty that manifests itself in the spread of the data points in Figure 5. The statistical uncertainty must be properly assessed in order to be able to judge the reliability of the obtained results.

Besides the inherent width of the variable's distribution, there is the problem of autocorrelation of the samples within one MC simulation,[4,47] i.e., the samples might not be totally independent. The data points in Figure 5, are taken from independent and thus uncorrelated simulations (started from different random number seeds). The statistical uncertainty of one sample can be estimated from the standard deviation of all samples in this case.

Often, only a single trajectory is used to obtain average observables and to assess their statistical error. In this case, one has to account for possible autocorrelations between the individual samples of the trajectory.[4] This error estimation is optional and implemented separately from the actual simulation and requires the recording of the trajectories visited during the Metropolis MC simulations. The autocorrelation of samples is seldom a serious problem because GMCT automatically adjusts the number of MC steps per MC scan to the system size. The automatic adjustment ensures that 10,000–30,000 MC scans suffice in most cases for computing titration curves of high quality (see, e.g., Fig. 5).

Titration calculations with Metropolis MC are widely used to study protonation, reduction and other binding equilibria in biological macromolecules.[19,21,65,66] GMCT was used to study the effect of serine and histidine phosphorylation on the protonation and interaction properties of the protein HPr.[67,68] In two further publications, protonation probabilty calculations with GMCT were employed to study the role of protonatable amino acid sidechains in the enzymatic mechanism of ferredoxin reductases.[69,70] Titration curves computed with GMCT were also used to investigate the protonation behavior of titratable amino acid sidechains in recently determined crystal structures of xenobiotic reductase A[71] and 4-hydroxyphenylacetate decarboxylase.[72] Further analysis of computed titration curves and their correlation with experimentally measured titration curves can be used to identify functionally important titratable groups[38,53,74–76] or to analyze the effect of mutations on the titration behavior of a protein.[20,73,74] The analysis of computed titration curves can also be used to understand cooperative binding like the allosteric binding of oxygen and protons to hemoglobin,[77] the coupling of protonation and reduction in electron transfer proteins and oxidoreductases.[13,54,64,78,79] Titration calculations with GMCT can also help to understand the effect of electrochemical transmembrane gradients on the conformation and binding behavior of integral membrane proteins.[36,37,39] GMCT can be used to compute cooperativity measures that facilitate the analysis of cooperative binding events as discussed in the "Cooperativity Measures" section.

Wang-Landau Monte Carlo

Wang-Landau MC[80,81] is a relatively new simulation method that has not been applied before in the context of receptor models like those considered by GMCT. In this section, we discuss specific details of Wang-Landau MC simulations with GMCT and give some application examples. Applications and limitations of the method are illustrated with the aid of example systems.

General principle. The Wang-Landau MC method[80,81] differs from Metropolis MC in that the simulation trajectory is not build up with the aim to restrict the system to the low-energy regions of the state energy surface, but with the aim to visit all possible energy levels with equal probability. The range of possible microstate energies is divided in bins i extending from a given microstate energy E_i^{micro} and extending over a finite bin width $\Delta E_i^{\text{micro}} = E_{i+1}^{\text{micro}} - E_i^{\text{micro}}$. In the case of discrete microstates and a number of discrete microstate energy values i, there is a integral number g_i of microstates available to the system within each energy bin i.

The distribution of the microstates on the microstate energy surface for the small triprotic acid DTPA is depicted in Figure 6a. With increasing system size, the number of discrete energy values can become very high. In effect, g can appear as a continuous function of the microstate energy already for systems of moderate size as shown for reduced PaAz in Figure 6b.

The program stores two arrays containing the numbers of microstates g_i for each energy bin and a histogram of the number of visits to each energy bin h_i, respectively. An additional histogram array h_a^{subst} with the same number of entries as the overall histogram H is stored for each substate a that is to be monitored. A Wang-Landau MC simulation is similar to a Metropolis MC simulation, in that the MC trial moves are proposed and afterwards accepted or rejected with a certain probability. During the simulation, the distribution of the number of the microstates is refined with a multiplicative factor f each time an energy bin is visited.

Every time a certain number of MC scans is reached, the modification factor is decreased. The modification factor is not decreased if the histogram flatness criterion (see later) is larger than the minimum flatness criterion reached in the simulation. The modification factor is also not reduced if an occupied energy bin was visited for the first time in the current iteration cycle. In GMCT, the new modification factor is computed from the current modification factor via $f_{i+1} = f_i^{1/\alpha}$, where $\alpha > 1$. For typical systems in GMCT, it was found that it is often advantageous to choose a value of α significantly smaller than 2 (default 1.2). Reasonable values for the initial modification factor are $1.1 \leq f_0 \leq 2$. The simulation stops if the modification factor reaches a prescribed value that is typically close to 1 (default $\exp[5 \times 10^{-7}]$), so that there is effectively no further modification of g.

In the course of the simulation the distribution of the number of microstates g converges to the true distribution up to a multiplicative constant resulting in equal probabilities of visiting each energy bin. The convergence is judged by the flatness of a histogram of visits to each energy level maintained during the simulation. The simulation is considered to be converged if no histogram value has a relative deviation $|1 - h_i/h^{\text{mean}}|$ of more than τ^{flat} from the mean histogram value. Here, the mean is taken over all energy bins with nonzero histogram value. Reasonable values for the flatness criterion were found to be $10^{-1} \geq \tau^{\text{flat}} \geq 10^{-3}$ (default 10^{-3}). If τ^{flat} is set to a value very close to zero, the simulation proceeds until the modification factor falls below the prescribed minimum value. It is advisable to check also the convergence of the ratio of the maximum to minimum value of g stated in the program output.

Applications. The Wang-Landau MC method is used in GMCT to compute properties related to macroscopic binding states. Here, macroscopic binding states means that only the total number of ligands of each type is specified while the binding sites are not. A binding macrostate can comprise many microstates. The study of macroscopic binding states is useful to predict and

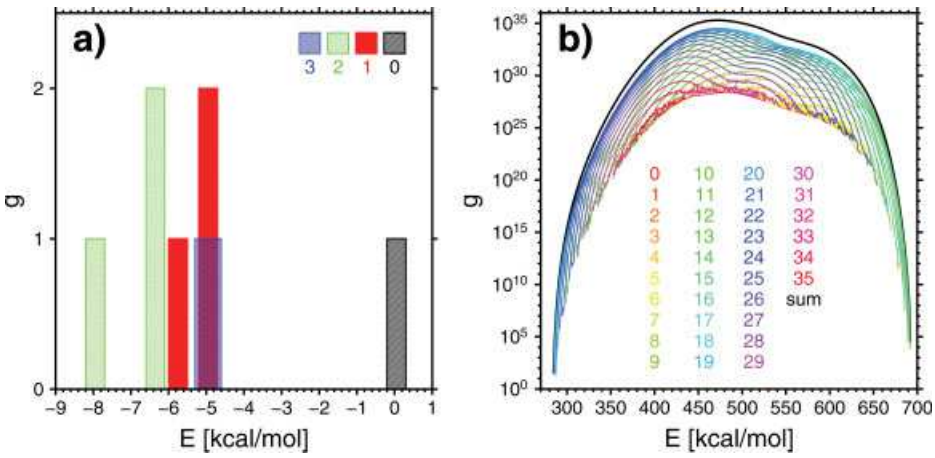

Figure 6. Number of microstates as function of the microstate energyfor macroscopic protonation states at pH 7.0. The numbers in the labels indicate the total number of protons bound to the receptor at the respective macroscopic protonation state. The label sum indicates the total number of microstates summed over all macroscopic protonation states. a) DTPA b) reduced azurin from *Pseudomonas aeruginosa*. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

understand the overall behavior of the receptor in response to the properties of the environment of the receptor.

The distribution g of the microstates on the microstate energy surface can be used to compute global thermodynamic properties of the receptor. The partition function can be computed very accurately even for large systems via

$$\mathcal{Z} = \sum_{i}^{N^{bins}} g_i \exp[-\beta E_i] \quad (6)$$

where g_i is the number of microstates in a given energy bin i of minimum energy E_i. Bin widths of 0.1 kcal/mol are sufficient for very accurate results. The partition function of the system can be used to compute the free energy, the enthalpy and the entropy of the system. Entropic contributions to the free energy due to implicitly modeled degrees of freedom are not resolved but included in the computed enthalpy. Such implicitly modeled degrees of freedom in a continuum electrostatic model could, for example, be solvent degrees of freedom or implicitly represented conformational flexibility of the receptor.

The partition function of a substate a can be computed analogously to eq. (6) using the corresponding distribution of microstates that are part of the substate. The computed partition functions can be used to compute the occupation probability of a substate or the free energy difference between two substates. Macroscopic binding constants and macroscopic pK_a values can be calculated from macroscopic binding free energies as shown in Ref. [38].

Figure 7 provides a motivating example for the application of Wang-Landau MC to complex ligand binding systems with coupled binding phenomena. The macroscopic protonation probabilities of PaAz are markedly altered upon reduction or oxidation of the copper center. The quantitative analysis of this alteration helped to understand the coupling of protonation and reduction in PaAz as shown in Ref. [54]. Furthermore, it can be seen that for every pH value there are multiple macroscopic protonation states that occur with significant probability. This information could for example be important in the planning and setup of molecular dynamics simulations to decide whether it is sufficient to simulate a single protonation state of the system or if a constant-pH molecular dynamics simulation is required.

Free energy calculations with the program suite GMCT

Free energy is the central quantity of thermodynamics and thus of outstanding importance in physics, chemistry, and biology. Free energy differences provide the driving force for any physical process. Free energy calculations with GMCT can, e.g., be used to study the thermodynamics of binding processes and their coupling under realistic conditions including electrochemical potential gradients.

A variety of free energy simulation techniques exists that makes the calculation of free energies for arbitrary transformations possible.[82–87] These methods ensure that the system visits both end states, even if one or both of the end states are not readily observed in an equilibrium simulation. The implementation of all free energy calculation methods presented below follows the concepts presented in Ref. [32].

The biased Monte Carlo method. Biased MC[58] is conceptional intriguingly simple and despite that, does still give accurate results. A biased Metropolis MC simulations is run in which the system is restricted to the end states of the reaction. A bias energy B between the end states of the transformation is iteratively refined with the aim to obtain equal probabilities of 0.5 for visiting both end states. The transformation free energy is then obtained as the negative average bias potential

$$\Delta G_{0\to 1} = -\frac{1}{N^{sa}} \sum_{s}^{N^{sa}} B_s \quad (7)$$

computed from a number N^{sa} of trial values for the bias. The statistical uncertainty of the bias energy is often high, especially in larger systems or when dealing with complex transformations involving reactive subsystems with many different configurations, e.g., in the calculation of cooperativity free energies. The

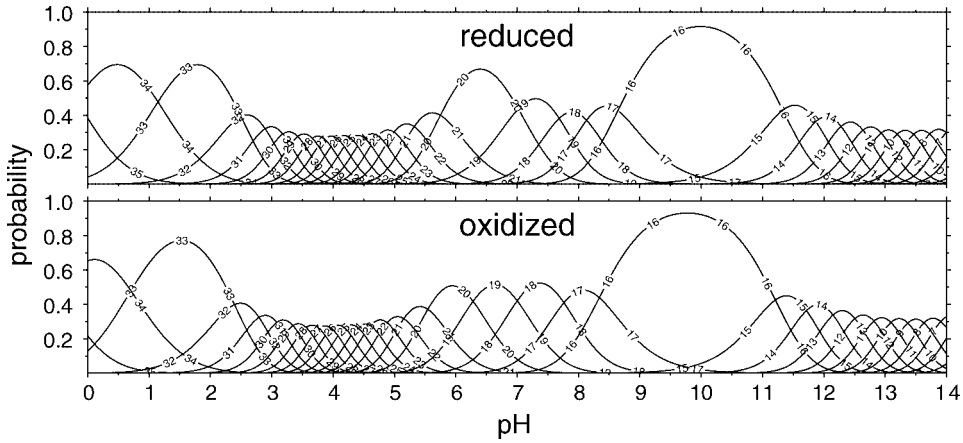

Figure 7. Equilibrium occupation probabilities of macroscopic protonation states in oxidized and reduced azurin from *Pseudomonas aeruginosa*. The curves are labeled with the total number of protons bound to the receptor at the respective macroscopic protonation state. Reprinted from Ref. [54], with permission from the American Chemical Society.

free energy estimate converges very slowly in such cases. More rapidly converging methods are presented below.

Staging with chimeric intermediates. In principle, the free energy difference between two regions of configuration space can be calculated by performing a free energy calculation just between the end states. However, if the configuration space overlap between the end states is very small, the free energy estimate obtained from the simulation will be afflicted with a high statistical error. In such cases, one will need to perform many or a very long simulation to obtain a reliable free energy estimate with acceptable statistical error.

An alternative way of performing the simulation is to break down the problem into multiple steps (also called stages) by introducing intermediate states along a transformation coordinate $0 \leq \lambda \leq 1$.[87,88] These intermediate states are chimeras of the two end states, where the relative weight of the end states is determined by λ (see user manual for details). The introduction of stages increases the overlap between adjacent states along the transformation coordinate and can thus accelerate the convergence of a free energy calculation.

We performed test calculations on a variety of model systems with different numbers of intermediates employing the free energy perturbation (FEP),[31,32] the non-equilibrium work (NEW)[34] method or thermodynamic integration (TI)[33,82,84] (see below). Our test calculations did not lead to a clear suggestion for an optimum parameter choice. The optimum setting in terms of simulation time varied widely between the test systems. For the calculation of free energies for simple transformations with the FEP method, one intermediate state was sufficient in most cases, while especially for higher order free energies of cooperativity two or three intermediates were sometimes better. The use of a larger number of intermediates with the FEP method was seldom favorable. For free energy calculations with the NEW method, 21 intermediate states were found to be a good choice for a first guess. In some cases, the optimum number of intermediates for the NEW method was significantly larger. For free energy calculations with the TI method, it is advisable to use a higher number of intermediates to ensure that the numeric integration works well. Using 101 intermediate states was found to be a choice which is quite sure to work also in case of complicated shapes of the integrand.

The thermodynamic integration method. The thermodynamic integration (TI) method[33] is a very common and popular free energy calculation method.[89–91] GMCT uses the discrete version of TI instead of slow growth TI to ensure that the systems is always in equilibrium.[84] That is, the integral over the average derivative of the state energy with respect to λ is computed by evaluating the integrand at a number of discrete values of the coupling parameter λ and applying a numeric integration technique.[89,90] The integrand is evaluated at the two end states and a number of N^{int} chimeric intermediates. By default, we use Simpson's rule for the numeric integration. Simpson's rule is only applicable if an odd number of chimeric intermediates N^{int} is used. In case of an even number of chimeric intermediates the program falls back to the less advantageous use of the trapezoidal rule. The numeric integration techniques and their performance are described in detail by Bruckner and Boresch.[89,90]

In comparison to the free energy perturbation and nonequilibrium work methods introduced below, TI involves an additional source of error introduced by the numeric integration. The requirement of running full Metropolis MC simulations for a relatively large number of intermediates to obtain reliable results makes TI often less efficient than the FEP and NEW methods described in the next sections. However, the TI method can have superior performance if many degrees of freedom are involved in a transformation. Such a case can for example occur if a large number of binding sites change their binding form in the transformation.

The free energy perturbation method. Free energy perturbation (FEP) is a very precise[89,92,93] and efficient free energy simulation method with a long tradition.[31,82,85–87,94] The implementation of the FEP method in GMCT is based on the initial work of Zwanzig[31] and our recent generalization of the FEP theory.[32] Our generalized FEP theory was inspired by MC methods and is especially well suited for application to systems that are defined on discrete configuration spaces like the receptor models in GMCT.

There are two efficient FEP simulation schemes in GMCT. Both simulation schemes are based on an equilibrium simulation of the system in the initial state with Metropolis MC (see "Metropolis Monte Carlo" section). Periodically, in distances of one or more MC scans, the system is forced to visit the adjacent end state with one or more so-called FEP move(s).[32] The multimove simulation scheme collects microstate energy difference samples for all possible FEP moves. The random single-move simulation scheme collects a single microstate energy difference sample for a single, randomly selected FEP move.

The performance of the multimove and the random-single-move simulation schemes is often comparable. Using the multimove simulation scheme can be advantageous if the number of configurations of the reactive subsystem in the final state is not too large (e.g., in the calculation for Fig. 9). With a growing number of these configurations, the advantage of using more information in the calculation is outweighed by the disadvantage of having to compute the microstate energy more often. In addition, the memory demand is higher because multiple microstate energy difference samples instead of a single one have to be stored per sampling step.

The nonequilibrium work method. The nonequilibrium work (NEW) method of Jarzynski[34] shares similarities with the free energy perturbation method described in the previous section. The transformation of the initial to the final state can also be partitioned into a number of steps by introducing chimeric intermediates. The NEW method takes the average over work values done in microscopic trajectories between the initial and the final state including all the discrete, intermediate states. In contrast, the FEP method averages over state energy differences and the averages are taken for each step separately.

In test calculations, the overall performance of NEW method was in most cases similar to that of the FEP method.

The Bennett acceptance ratio method. The Bennett acceptance ratio estimator[35] can be used in free energy calculations with the FEP or NEW methods to compute a maximum likelihood estimate of the free energy difference.[35, 95, 96] This estimator incorporates energy difference or work samples collected in forward and reverse direction. The use of the Bennett acceptance ratio method often dramatically accelerates the convergence of a free energy calculation.[35, 89, 92, 93, 96–98]

Applications of free energy calculations. The range of applications for free energy calculations with GMCT is broad. One can study isolated events like conformational changes, binding reactions and charge transfer reactions, as well as concerted events. Derived properties that can be calculated from the computed free energy differences include binding constants, pK_a values,[38, 54, 76] and reduction potentials.[13, 54] The free energy simulation methods are also the basis for the calculation of free energy measures of cooperativity introduced in the "Cooperativity Measures" section, that can be used to study the thermodynamic coupling of events.

The end states of a transformation can be freely defined in a user-friendly and flexible way. It is possible to use wild cards in the specification of conformation and site and instance names. One can also specify fixed values or value ranges for the number of ligands bound to a site. For convenience, there is the possibility to let the program automatically determine all stepwise binding reactions with the corresponding end states for all sites or a subset of sites.

The desired precision of the computed free energy estimates can be adjusted. The simulation is run until the standard deviation of the mean free energy estimate falls below a user-defined tolerance value. The quality of the free energy differences that can be obtained with different tolerance values is illustrated by the example of DTPA in Figure 8. The quality of a free energy estimate can be impaired if there are sampling problems. Sampling problems are rarely encountered with Metropolis MC simulations with GMCT, because the MC moves ensure efficient sampling in most cases. See, however, refs. [86] and [87] for further information on how to assess the quality of free energy estimates obtained from a simulation and the detection of sampling problems.

Figure 9 shows that very good accuracy can be obtained also in complex systems for a large number of different conditions. The protonation free energy of His-35 of PaAz is afflicted with an especially high intrinsic statistical uncertainty because of a low overlap of the end states. Figure 9 is also an example for

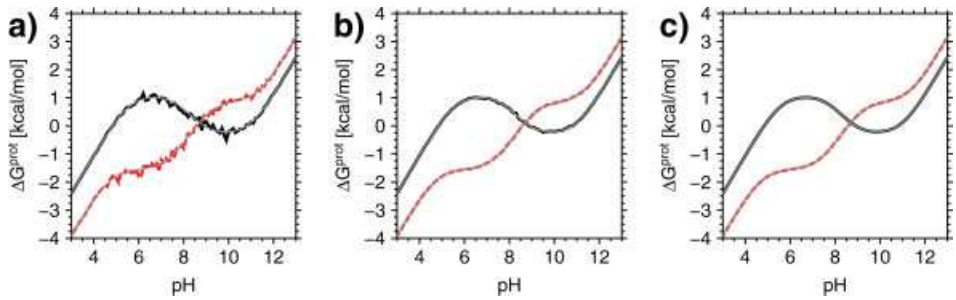

Figure 8. Free energies of binding a proton to the sites of DTPA as a function of the pH value for different values of the statistical error tolerance. The protonation free energy for the central nitrogen atom is indicated by the black solid curve. The protonation free energy of the two equivalent terminal nitrogen atoms is indicated by the red broken curve. The binding free energies are computed with the random single-move FEP simulation scheme combined with the Bennett acceptance ratio method. For comparison, the analytically computed free energy differences are overlaid as gray solid lines. The statistical error tolerance was set to a) $\sigma = 0.1$ kcal/mol b) $\sigma = 0.01$ kcal/mol a) $\sigma = 0.001$ kcal/mol.

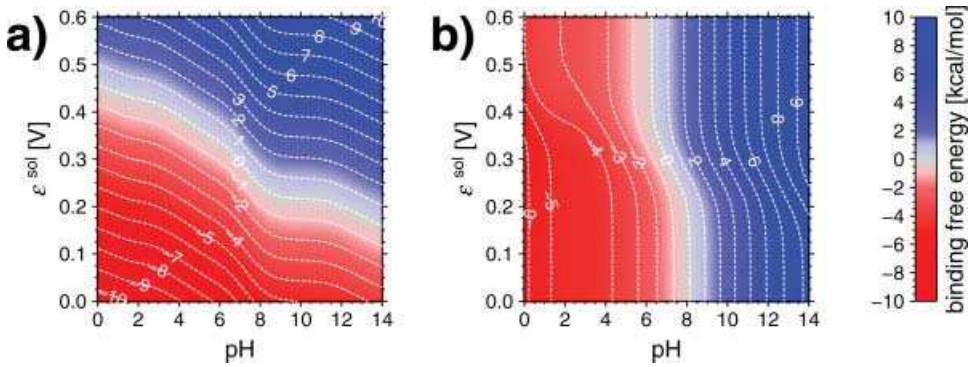

Figure 9. Binding free energies of two important sites involved in the coupling of protonation and reduction in azurin from *Pseudomonas aeruginosa* (see Fig. 4b). The binding free energies are plotted as functions of the pH value and the reduction potential of the solution $\mathcal{E}^{sol}$. Contour values are given in kcal/mol. The protonation free energies are computed with the multi-move FEP simulation scheme using one intermediate state and the Bennett acceptance ratio method. The statistical error tolerance was set to 0.01 kcal/mol. a) free energy of binding an electron to the copper center (reduction free energy) b) free energy of binding a proton to His-35 (protonation free energy).

the application of GMCT to study the coupling of ligand binding processes.[54] It can be seen that the reduction free energy of the copper center does not only depend on the reduction potential but also on the pH value of the solution. The protonation free energy energy of His-35 does also depend on both chemical potentials. The coupling of binding processes can be studied with the cooperativity measures that are introduced in the next section.

Cooperativity measures

Cooperativity often forms the basis of biomolecular function and is thus also a key to its understanding. The term cooperativity means that two or more events influence each other.

The covariance as measure of cooperativity. The covariance or (unnormalized) correlation of two substates, at the same point of an equilibrium trajectory is given by[36,38]

$$\mathcal{C}_{a,b}(\tau = 0) = p(a)p(b) - p(a,b) \tag{8}$$

where $p(a)$ and $p(b)$ are the equilibrium probabilities of finding substates a and b without regard to the respective other substate, while $p(a,b)$ is the equilibrium probability of finding both substates occupied by a sample from the trajectory. Thus, the covariance measures if there is an increased or decreased probability of finding the two substates occupied in the same sample relative to the probability of finding them independently. If one wishes to investigate the coupling of two events one can describe these events as transformations (or reactions in a general sense). The coupling of the events can then be measured with the covariance between the initial or final states of the transformations.

Free energy measures of cooperativity. The cooperativity between two events can be measured with the cooperativity free energy $G^{\mathrm{coop(n)}}$.[99] where n in the superscript denotes the order of the cooperativity or the number of events between which the cooperativity is measured. Each event defines a transformation of the system from an initial state to a final state (see "Free Energy Calculations with the Program Suite GMCT" section). The free energy of cooperativity is given by

$$\begin{aligned} G^{\mathrm{coop(n)}} = {} & G_{(0,0,\ldots)\to(1,1,\ldots)} \\ & - (G_{(0,0,\ldots)\to(1,0,\ldots)} + G_{(0,0,\ldots)\to(0,1,\ldots)} + \ldots) \end{aligned} \tag{9}$$

The cooperativity free energy measures the difference of the free energy cost for performing all events simultaneously and the sum of the free energy costs for each separate event while fixing all the other reactive subsystems in the initial state. Thermodynamic schemes for the calculation of free energies of cooperativity for pairs and triplets of transformations can be found in Ref. [54] or the user manual. The cooperativity free energy is a sensitive quantitative measure of cooperativity that has a clear thermodynamic interpretation as effective interaction energy between the sites undergoing the transformations. Note that the cooperativity free energy for a pair or triplet of events is conceptually identical to the effective interaction energy measured in a double-mutant or triple-mutant cycle experiment, respectively.[100,101]

Applications. The cooperativity measures introduced above indicate and quantify an effective thermodynamic coupling between distinct events. This knowledge can, for example, be used to understand the basis of cooperative receptor behavior or energy transducing biomolecular complexes on statistical thermodynamics grounds.[54,99] The analysis of a system with cooperativity measures can help to understand its behavior or mechanism in a structural context.

An example for the application of the cooperativity measures can be found in Ref. [54]. In this work, we applied cooperativity measures for a detailed study of the coupling of protonation, reduction and conformational change in PaAz. We show, that cooperativity free energies are especially useful for analyzing the thermodynamic coupling of events in complex molecular systems.

Conclusion

GMCT allows for a detailed modeling of complex macromolecular receptor systems like proteins or other polyelectrolytes under realistic conditions. A variety of Monte Carlo simulation methods can be used to study overall properties of the receptor as well as properties of individual sites. The description of the system in terms of discrete microstates of the receptor and chemical potentials of the ligands renders the simulations computationally very inexpensive relative to all-atom simulations. This computational efficiency enables very accurate calculations of receptor properties with low statistical uncertainty.

Properties of binding processes that can be calculated with GMCT are for example binding probabilities (titration curves), binding free energies and binding constants. These properties can be computed from a microscopic viewpoint for studying the behavior of separate sites or groups of sites in the receptor or from a macroscopic viewpoint for studying the overall behavior of the receptor. Midpoint reduction potentials $\mathcal{E}_{1/2}$ and $\mathrm{p}K_{1/2}$ values can be derived from computed titration curves. Binding free energies can be expressed in terms of thermodynamically defined reduction potentials and $\mathrm{p}K_{\mathrm{a}}$ values.[13,38,76]

The free energy calculation methods of GMCT can also be used to study charge transfer reactions, conformational transitions and any other process that can be described within the receptor model of GMCT. A particularly interesting feature of GMCT is the possibility to calculate free energy measures of cooperativity that can be used to study the coupling of different processes in the receptor. An example of special interest in our lab is the coupling between binding and transfer processes of charged ligands in bioenergetic protein complexes.[19,64,102]

GMCT can also be helpful in setting up and complementing molecular dynamics (MD) simulations.[67,68] The preparation of a protein structure for MD simulations does often require the specification of protonation states and tautomeric states occupied by titratable residues. This information can be obtained from

Metropolis MC calculations with GMCT. In addition, protonation state calculations can be used to assess whether the modeling of a system could require a constant-pH MD method.[103–106] This would for example be the case for our example system PaAz, because this protein can occur in more than one macroscopic protonation state over the whole range of physiologically relevant pH values as can be seen from Figure 7.

The GMCT distribution includes a number of example calculations including the examples used in this paper. The examples are provided with all necessary input files, commented setup files and a tutorial on how to run the calculations. The tutorial examples will make it easy for potential user to familiarize themselves with the programs and can serve as primers for setting up own calculations.

The use of free format, text based input files, and its extensibility as free open source software should allow many researchers to use the features of GMCT. The software can be used as is for simulations on any receptor model that is formulated in terms of discrete microstates whose energy function can be expressed in the form of eq. (4).[19,21–23] GMCT can be obtained free of charge under the terms of the GNU Affero General Public License from www.bisb.uni-bayreuth.de.

Acknowledgments

R.T.U. Thank the state of Bavaria for a doctoral fellowship according to the Bayerische Eliteförderungsgesetz.

Keywords: ligand binding • Monte Carlo simulation • electrochemical potential • free energy • cooperativity • pKa • reduction potential • membrane potential • bioenergetics • cooperativity

How to cite this article: R. T. Ullmann, G. M. Ullmann, *J. Comput. Chem.* **2012**, *33*, 887–900. DOI: 10.1002/jcc.22919

[1] M. K. Gilson, B. H. Honig, *Proteins* **1988**, *4*, 7.
[2] D. Bashford, M. Karplus, *Biochemistry* **1990**, *29*, 10219.
[3] D. Bashford, M. Karplus, *J. Phys. Chem.* **1991**, *95*, 9557.
[4] P. Beroza, D. R. Fredkin, M. Y. Okamura, G. Feher, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5804.
[5] M. K. Gilson, *Proteins* **1993**, *15*, 266.
[6] A.-S. Yang, M. R. Gunner, R. Sampogna, B. Honig, *Proteins* **1993**, *15*, 252.
[7] T. You, D. Bashford, *Biophys. J.* **1995**, *69*, 1721.
[8] J. Antosiewicz, J. M. Briggs, A. H. Elcock, G. M. K., J. A. McCammon, *J. Comput. Chem.* **1996**, *17*, 1633.
[9] P. Beroza, D. R. Fredkin, *J. Comput. Chem.* **1996**, *17*, 1229.
[10] E. G. Alexov, M. R. Gunner, *Biophys. J.* **1997**, *74*, 2075.
[11] V. Z. Spassov, D. Bashford, *J. Comput. Chem.* **1999**, *20*, 1091.
[12] G. M. Ullmann, E. W. Knapp, *Eur. Biophys. J.* **1999**, *28*, 533.
[13] G. M. Ullmann, *J. Phys. Chem. B* **2000**, *104*, 6293.
[14] A. M. Ferreira, D. Bashford, *J. Am. Chem. Soc.* **2006**, *128*, 16778.
[15] T. Becker, R. T. Ullmann, G. M. Ullmann, *J. Phys. Chem. B* **2007**, *111*, 2957.
[16] T. Essigke, A continuum electrostatic approach for calculating the binding energetics of multiple ligands, PhD thesis, University of Bayreuth, **2008**.
[17] A. Vitalis, R. V. Pappu, In Annual reports in computational chemistry, Vol. 5; R. A. Wheeler, Ed.; Elsevier B. V., Amsterdam, **2009**; pp. 49–76.
[18] W. Jorgensen, J. Tirado-Rives, *J. Phys. Chem.* **1996**, *100*, 14508.
[19] G. M. Ullmann, E. Kloppmann, T. Essigke, E.-M. Krammer, A. R. Klingen, T. Becker, E. Bombarda, *Photosynth. Res.* **2008**, *97*, 33.
[20] A. R. Klingen, E. Bombarda, G. M. Ullmann, *Photochem. Photobiol. Sci.* **2006**, *5*, 588.
[21] M. R. Gunner, J. Mao, Y. Song, J. Kim, *Biochim. Biophys. Acta.* **2006**, *1757*, 942.
[22] Z. Piłat, J. M. Antosiewicz, *J. Phys. Chem. B* **2010**, *114*, 1393.
[23] D. Bashford, *Frontiers in Bioscience* **2004**, *9*, 1082.
[24] J. E. Nielsen, G. Vriend, *Proteins* **2001**, *43*, 403.
[25] Z. Piłat, J. M. Antosiewicz, *J. Phys. Chem. B* **2008**, *112*, 15074.
[26] R. E. Georgescu, E. G. Alexov, M. R. Gunner, *Biophys. J.* **2002**, *83*, 1731.
[27] Y. Song, J. Mao, M. R. Gunner, *J. Comput. Chem.* **2009**, *30*, 2231.
[28] G. Kieseritzky, E.-W. Knapp, *Proteins* **2008**, *71*, 1335.
[29] A. M. Baptista, C. M. Soares, *J. Phys. Chem. B* **2001**, *105*, 293.
[30] V. Couch, A. Stuchebruckhov, *Proteins* **2011**, *79*, 3410.
[31] R. W. Zwanzig, *J. Chem. Phys.* **1954**, *22*, 1420.
[32] R. T. Ullmann, G. M. Ullmann, *J. Phys. Chem. B* **2011**, *115*, 507.
[33] J. G. Kirkwood, *J. Chem. Phys.* **1935**, *3*, 300.
[34] C. Jarzynski, *Phys. Rev. E* **1997**, *56*, 5018.
[35] C. H. Bennett, *J. Comput. Phys.* **1976**, *22*, 245.
[36] N. Calimet, G. M. Ullmann, *J. Mol. Biol.* **2004**, *339*, 571.
[37] E. Bombarda, T. Becker, G. M. Ullmann, *J. Am. Chem. Soc.* **2006**, *128*, 12129.
[38] G. M. Ullmann, *J. Phys. Chem. B* **2003**, *107*, 1263.
[39] B. Roux, *Biophys. J.* **1997**, *73*, 2980.
[40] H.-X. Zhou, M. K. Gilson, *Chem. Rev.* **2009**, *109*, 4092.
[41] Y. Y. Sham, Z. T. Chu, A. Warshel, *J. Phys. Chem. B* **1997**, *101*, 4458.
[42] A. Warshel, *Biochemistry* **1981**, *20*, 3167.
[43] D. Bashford, In Scientific computing in object-oriented parallel environments, Y. Ishikawa, R. Oldehoeft, J. Reynders, M. Tholburn, Eds.; Lecture notes in computer science, Vol. 1343, Springer: Berlin, **1997**; pp. 233–240.
[44] R. T. Ullmann, G. M. Ullmann, **2011**, to be submitted for publication.
[45] M. Allen, D. Tildesley, Computer simulation of liquids; Clarendon Press: Oxford, **1990**.
[46] D. Frenkel, In Computational soft matter: from synthetic polymers to proteins, lecture notes, Vol. 23 of NIC Series; John von Neumann Institute for Computing, Jülich, **2002**; pp. 29–60.
[47] B. A. Berg, In Markov Chain Monte Carlo, Vol. 7 of Lecture notes series, Institute for Mathematical Sciences, National University of Singapore; World Scientific, **2005**; Chapter 1.
[48] M. Lewerenz, In Quantum simulations of complex many-body systems: from theory to algorithms, lecture notes, Vol. 10 of NIC Series; John von Neumann Institute for Computing, Jülich, **2002**; pp. 1–24.
[49] T. H. Click, A. Liu, G. A. Kaminski, *J. Comput. Chem.* **2011**, *32*, 513.
[50] M. Matsumoto, T. Nishimura, *ACM Trans. Model. Comput. Simul.* **1998**, *8*, 3.
[51] P. Letkeman, *J. Chem. Ed.* **1979**, *56*, 348.
[52] J. L. Sudmeier, C. N. Reilley, *Analyt. Chem.* **1964**, *36*, 1698.
[53] A. Onufriev, D. A. Case, G. M. Ullmann, *Biochemistry* **2001**, *40*, 3413.
[54] R. T. Ullmann, G. M. Ullmann, *J. Phys. Chem. B* **2011**, *115*, 10346.
[55] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33.
[56] Stone, J. An efficient library for parallel ray tracing and animation, Master's thesis, Computer Science Department, University of Missouri-Rolla, April, **1998**.
[57] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, *J. Chem. Phys.* **1953**, *21*, 1087.
[58] P. Beroza, D. R. Fredkin, M. Y. Okamura, G. Feher, *Biophys. J.* **1995**, *68*, 2233.
[59] C. R. Lancaster, H. Michel, B. Honig, M. R. Gunner, *Biophys. J.* **1996**, *70*, 2469.

[60] B. Rabenstein, G. M. Ullmann, E. W. Knapp, *Biochemistry* **1998**, *37*, 2488.
[61] B. Rabenstein, G. M. Ullmann, E. W. Knapp, *Eur. Biophys. J.* **1998**, *27*, 626.
[62] A. Kannt, C. R. D. Lancaster, H. Michel, *Biophys. J.* **1998**, *74*, 708.
[63] J. Köpke, E.-M. Krammer, A. R. Klingen, P. Sebban, G. M. Ullmann, G. Fritzsch, *J. Mol. Biol.* **2007**, *371*, 396.
[64] A. R. Klingen, H. Palsdottir, C. Hunte, G. M. Ullmann, *Biochim Biophys Acta-Bioenergetics* **2007**, *1767*, 204.
[65] B. García-Moreno E, C. A. Fitch, In Methods in enzymology, Vol. 380; J. M. Holt, M. L. Johnson, G. K. Ackers, Eds.; Academic Press, London, **2004**; pp. 20–51.
[66] J. E. Nielsen, J. A. McCammon, *Prot. Sci.* **2003**, *12*, 1894.
[67] N. Homeyer, T. Essigke, G. M. Ullmann, H. Sticht, *Biochemistry* **2007**, *46*, 12314.
[68] N. Homeyer, T. Essigke, H. Meiselbach, G. Ullmann, H. Sticht, *J. Mol. Model* **2007**, *13*, 431.
[69] V. I. Dumit, N. Cortez, G. M. Ullmann, *J. Mol. Biol.* **2010**, *397*, 814.
[70] V. I. Dumit, N. Cortez, G. M. Ullmann, *Proteins* **2011**, *79*, 2076.
[71] O. Spiegelhauer, S. Mende, F. Dickert, S. H. Knauer, G. M. Ullmann, H. Dobbek, *J. Mol. Biol.* **2010**, *398*, 66.
[72] B. M. Martins, M. Blaser, M. Feliks, G. M. Ullmann, W. Buckel, T. Selmer, *J. Am. Chem. Soc.* **2011**, *133*, 14666.
[73] J. E. Nielsen, *J. Mol. Graph. Model* **2007**, *25*, 691.
[74] C. R. Søndergaard, L. P. McIntosh, G. Pollastri, J. E. Nielsen, *J. Mol. Biol.*, **2008**, *376*, 269.
[75] H. Webb, B. M. Tynan-Connolly, G. M. Lee, D. Farrell, F. O'Meara, C. R. Søndergaard, K. Teilum, C. Hewage, L. P. McIntosh, J. E. Nielsen, *Proteins* **2010**, *79*, 685.
[76] E. Bombarda, G. M. Ullmann, *J. Phys. Chem. B* **2010**, *114*, 1994.
[77] A. Onufriev, G. M. Ullmann, *J. Phys. Chem. B* **2004**, *108*, 11157.
[78] A. M. Baptista, C. M. Soares, *Biophys. J.* **1999**, *76*, 2978.
[79] Z. Zheng, M. R. Gunner, *Proteins* **2009**, *75*, 719.
[80] F. Wang, D. P. Landau, *Phys. Rev. Lett.* **2001**, *86*, 2050.
[81] F. Wang, D. P. Landau, *Phys. Rev. E* **2001**, *64*, 056101.
[82] D. L. Beveridge, F. M. DiCapua, *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.
[83] W. L. Jorgensen, *Acc. Chem. Res.* **1989**, *22*, 184.
[84] P. Kollmann, *Chem. Rev.* **1993**, *93*, 2395.
[85] M. R. Shirts, D. L. Mobley, J. D. Chodera, In Annual reports in computational chemistry, Vol. 3; R. A. Wheeler, Elsevier B. V., Amsterdam, **2007**; chapter 4, pp. 41–59.
[86] C. Chipot, A. Pohorille, Free Energy Calculations, Vol. 86 of Springer Series in Chemical Physics; Springer: Berlin, **2007**.
[87] A. Pohorille, C. Jarzynski, C. Chipot, *J. Phys. Chem. B* **2010**, *114*, 10235.
[88] J. P. Valleau, D. N. Card, *J. Chem. Phys.* **1972**, *57*, 5457.
[89] S. Bruckner, S. Boresch, *J. Comput. Chem.* **2011**, *32*, 1303.
[90] S. Bruckner, S. Boresch, *J. Comput. Chem.* **2011**, *32*, 1320.
[91] van Gunsteren, W. F.; X. Daura, A. E. Mark, *Helvet. Chim. Acta.* **2002**, *85*, 3113.
[92] M. R. Shirts, J. W. Pitera, W. C. Swope, V. S. Pande, *J. Chem. Phys.* **2003**, *119*, 5740.
[93] M. R. Shirts, V. Pande, *J. Chem. Phys.* **2005**, *122*, 134508.
[94] W. L. Jorgensen, L. L. Thomas, *J. Chem. Theory Comp.* **2008**, *4*, 869.
[95] G. E. Crooks, *Phys. Rev. E* **2000**, *61*, 2361.
[96] M. R. Shirts, E. Bair, G. Hooker, V. S. Pande, *Phys. Rev. Lett.* **2003**, *91*, 140601.
[97] D. M. Ferguson, *J. Chem. Phys.* **1993**, *99*, 10086.
[98] M. R. Shirts, V. Pande, *J. Chem. Phys.* **2005**, *122*, 134508.
[99] A. Ben-Naim, Cooperativity and regulation in biochemical processes; Kluwer Academic: New York, **2001**.
[100] P. J. Carter, G. Winter, A. J. Wilkinson, A. R. Fersht, *Cell* **1984**, *38*, 835.
[101] A. Horovitz, *Fold Des.* **1996**, *1*, R121.
[102] G. M. Ullmann, L. Noodleman, D. A. Case, *J. Biol. Inorg. Chem.* **2002**, *7*, 632.
[103] S. L. Williams, C. A. F. de Oliveira, J. A. McCammon, *J. Chem. Theory Comp.* **2010**, *6*, 560.
[104] J. Chen, C. L. Brooks III, J. Khandogin, *Curr. Opin. Struct. Biol.* **2008**, *18*, 140.
[105] J. Mongan, D. Case, *Curr. Opin. Struct. Biol.* **2005**, *15*, 157.
[106] S. Donnini, F. Tegeler, G. Groenhof, H. Grubmüller, *J. Chem. Theory Comp.* **2011**, *7*, 1962.

Received: 29 August 2011
Revised: 21 November 2011
Accepted: 2 December 2011
Published online on 25 January 2012