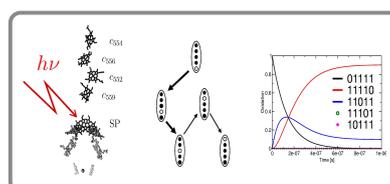
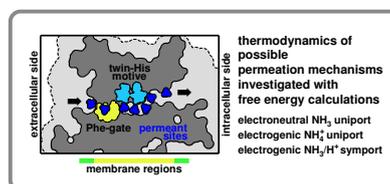
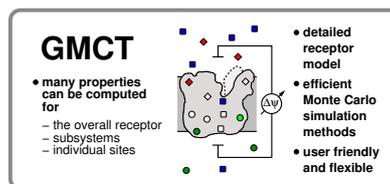
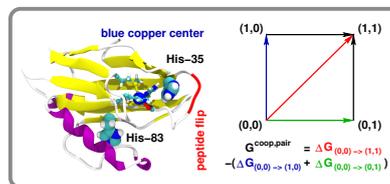
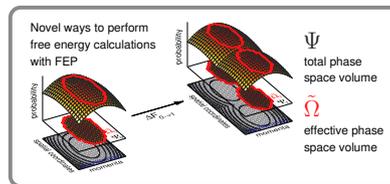


# Monte Carlo Simulation Methods for Studying the Thermodynamics of Ligand Binding & Transfer Processes in Biomolecules

R. Thomas Ullmann





---

Monte Carlo Simulation Methods for  
Studying the Thermodynamics of  
Ligand Binding & Transfer Processes  
in Biomolecules

---

Dissertation

submitted in partial fulfillment of the requirements for  
the degree of Dr. rer. nat. to the Faculty of Biology,  
Chemistry und Geosciences at the University of Bayreuth

by  
Rainer Thomas Ullmann  
from Annaberg, Germany

Bayreuth, July 2012



Die vorliegende Arbeit wurde an der Universität Bayreuth unter der Leitung von Professor Dr. G. Matthias Ullmann erstellt.

Vollständiger Abdruck der von der Fakultät Biologie, Chemie und Geowissenschaften der Universität Bayreuth genehmigten Dissertation zur Erlangung des akademischen Grades Doktor der Naturwissenschaften (Dr. rer. nat.).

Amtierender Dekan: Prof. Dr. Beate Lohnert

Datum der Abgabe: 2012/08/08

Prüfungsausschuss:

Prof. Dr. G. Matthias Ullmann (Erstgutachter)

Prof. Dr. Heinrich Sticht (Zweitgutachter)

Prof. Dr. Andreas Fery (Vorsitzender)

Prof. Dr. Franz X. Schmid

Tag des Kolloquiums: 2012/11/20

# Preface

This thesis belongs to the fields of computational biochemistry and computational biophysics. The topic of the thesis is the development and application of Monte Carlo simulation methods for studying binding and conformational equilibria in biomolecules. Special emphasize is put on biomolecules that function in bioenergetics. The simulated systems are modeled within a continuum electrostatics / molecular mechanics model.

Parts of this work might also be interesting for readers without immediate interest in the mentioned fields or topics, including experimentalists. Manuscript A and the associated open source software might be of value for those interested in the practical software implementation of different free energy calculation methods, e.g., free energy perturbation, thermodynamic integration and non-equilibrium work methods. Readers interested in free energy calculation methods (in particular free energy perturbation) or general statistical mechanics might find Manuscript B worthwhile. Chapter 2 and Manuscript D might find the interest of readers who are interested in the statistical mechanics of ligand binding reactions including the thermodynamic coupling of binding reactions, e.g. in case of ligands with multiple protonation forms. Cooperativity forms the thermodynamic basis of bioenergetic energy transduction. Manuscript C shows that cooperativity between events in molecular systems can be quantified in a thermodynamically meaningful way with cooperativity free energies that have experimental analogues in double and triple mutant cycle experiments. Readers interested in the field of bioenergetics might also find Manuscript D worthwhile which is concerned with the thermodynamics of transmembrane transport and the effect of electrochemical transmembrane potential differences. Finally, Manuscript E might be of value to those interested in the theory and simulation of electron transfer kinetics in proteins.

I hope that you will find this thesis interesting or even useful.

# Acknowledgments

Any major undertaking in life is influenced by the people that surround us, this thesis was no different. I wish to express my recognition and gratitude to anyone who contributed directly or indirectly to the success of this work.

First and foremost I am truly indebted to my advisor Matthias Ullmann for his invaluable guidance, encouragement, and trust throughout my research appointment. He introduced me to the fascinating fields of theoretical biochemistry, bioenergetics and philosophy of science. I am especially grateful for the freedom and encouragement to contribute and pursue my own ideas. His critical assessment of those ideas in many discussions has undoubtedly greatly contributed to shape my own understanding and the publications that have resulted from this thesis.

I also thank the state of Bavaria represented by Uni Bayern e.V. for a scholarship during part of my research appointment.

I sincerely thank Susana Andrade for her collaboration on the ammonium transport protein Amt-1.

I thank the members of the Bioinformatics and Structural Biology Group for the pleasant working atmosphere. I also enjoyed their company at tea times, parties, scientific conferences and meetings. Special thanks go to Edda Kloppmann, Verónica Dumit and Silke Wieninger for proofreading my manuscripts and providing helpful comments. I also thank all members of the group and the staff of the university computing center involved in the maintenance of the computational infrastructure, which I have extensively used for this thesis.

My sincere thanks for their support go to my parents and friends. Many thanks go to Sigrid Ullmann for designing the cover picture of this thesis.

I also want to thank all the interesting people I met at conferences and scientific meetings who made my stays at Vienna, Bremen, Freiburg and Hünfeld especially memorable and sometimes even fun.

# Abstract

The binding and transfer of ligands is of central importance for the function of many biomolecular systems. The main topic of this thesis is the development and application of Monte Carlo (MC) simulation methods for studying complex ligand binding equilibria which can also involve conformational changes. The simulated systems were described by microstates within a continuum electrostatics / molecular mechanics (CE/MM) model of the receptor-ligand system. The CE/MM modeling methodology was improved. The improvements led to more detailed molecular models that enable a more realistic reproduction of system properties and environmental conditions. The developed simulation methods were applied to biomolecular systems whose function involves aspects that are important for the understanding of bioenergetic energy transduction. The results of this thesis are presented in five articles that are published in peer reviewed scientific journals.

Manuscript A presents the Monte Carlo simulation software GMCT which was largely developed in this thesis. The software offers a variety of different simulation methods that allow the user to harness the full potential of CE/MM models in the simulation of complex receptor systems.

Manuscript B presents a novel theoretical framework for free energy calculations with the free energy perturbation method. The novel framework is more broadly applicable and can lead to more efficient simulations than previous formulations. The derivation of the formalism also led to interesting insights into general statistical mechanics. The formalism was implemented in GMCT and could already be used fruitfully for the free energy calculations presented in Manuscripts C and D.

Manuscript C demonstrates the application of free energy measures of cooperativity to study the coupling of protonation, reduction and conformational change in azurin from *Pseudomonas aeruginosa* (PaAz). Such a coupling is prototypic for bioenergetic systems because it forms the thermodynamic basis of their energy transducing function. PaAz

is an experimentally well characterized, small electron transport protein. For this reason, PaAz was used here as model system to demonstrate the usefulness of cooperativity free energies in detecting and quantifying thermodynamic coupling between events in complex biomolecular systems. The results of this study led to new insight that could help to determine the still enigmatic physiological role of PaAz.

In Manuscript D, free energy calculations were applied to study the thermodynamics of transport through the ammonium transporter Amt-1 from *Archaeoglobus fulgidus* (AfAmt-1). Ammonium is the most directly utilizable nitrogen source for plants and microorganisms. AfAmt-1 and its homologues facilitate the transport of ammonia/ammonium across biological membranes in living beings from all domains of life. It is intensely debated how these proteins perform their function and whether ammonia or its protonated form ammonium is actually transported. The study extended upon previous theoretical studies by including the effects of substrate concentration, electrochemical transmembrane gradients, proton-coupled binding equilibria and competitive binding of different ligand species. It was found that the transported species is most likely the ammonium ion. An  $\text{NH}_3/\text{H}^+$  symport mechanism, that involves a pair of coplanar histidine residues at the center of the transmembrane pore as transient proton acceptor, is made plausible by the high genetic conservation of these residues.

Manuscript E presents a first application of the microstate description within a CE/MM model to the simulation of the non-equilibrium dynamics of a molecular system. We simulated the re-reduction kinetics of the primary electron donor in the photocycle of the bacterial photosynthetic reaction center from *Blastochloris viridis*. The simulation results are in very good agreement with experimentally measured data.

Keywords: continuum electrostatics, free energy calculation, Monte Carlo simulation, free energy perturbation, ligand binding, binding free energy, electrochemical potential, membrane potential, cooperativity, transmembrane transport, bioenergetics

# Zusammenfassung

Die Bindung und der Transfer von Liganden ist von zentraler Bedeutung für die Funktion biomolekularer Systeme. Das Hauptthema dieser Dissertation ist die Entwicklung und Anwendung von Monte-Carlo-Simulationsmethoden für das Studium komplexer Ligandenbindungsgleichgewichte die auch Konformationsänderungen beinhalten können. Die simulierten Systeme wurden durch Mikrozustände innerhalb eines Kontinuumselektrostatik/Molekularmechnik-Modells (CE/MM-Modells) beschrieben. Die CE/MM-Modellierungsmethodik wurde verbessert. Das Ergebnis dieser Verbesserung waren detailliertere molekulare Modelle die es ermöglichen, Eigenschaften eines Systems und Umweltbedingungen realistischer darzustellen. Die entwickelten Simulationsmethoden wurden auf einige biomolekulare Systeme angewendet. Die Funktion dieser Systeme weist Aspekte auf die wichtig für das Verständnis der bioenergetischen Energieumwandlung sind. Die Ergebnisse dieser Arbeit sind in fünf Fachartikeln dargestellt die in begutachteten wissenschaftlichen Zeitschriften veröffentlicht wurden.

Manuskript A stellt die größtenteils im Rahmen dieser Arbeit entstandene Monte-Carlo-Simulationssoftware GMCT vor. Die Software bietet viele verschiedene Simulationsmethoden, die es dem Anwender erlauben, das volle Potential der CE/MM Modelle für die Simulation komplexer Rezeptorsysteme zu nutzen.

Manuskript B stellt ein neues theoretisches Grundgerüst für die Berechnung freier Energien mittels thermodynamischer Störungstheorie vor. Das neue Grundgerüst besitzt einen breiteren Anwendungsbereich und kann zu effizienteren Simulationen führen als frühere Formulierungen. Die Herleitung des Formalismus ergab auch interessante neue Einblicke in das Gebiet der allgemeinen Thermodynamik. Der Formalismus wurde in GMCT eingebaut und konnte für die in den Manuskripten C und D dargestellten Berechnungen freier Energien schon nutzbringend eingesetzt werden.

Manuskript C, demonstriert die Anwendung von Kooperativitätsmaßen auf Basis freier Energien zur Untersuchung der Kopplung von Protonierung, Reduktion und Konformationsänderung in Azurin aus

*Pseudomonas aeruginosa* (PaAz). Eine solche Kopplung ist prototypisch für bioenergetische Systeme, weil es die thermodynamische Grundlage ihrer Energieumwandlungsfunktion darstellt. PaAz ist ein experimentell gut charakterisiertes, kleines Elektronentransportprotein. Deshalb wurde PaAz hier als Modellsystem benutzt, um die Nützlichkeit freier Kooperativitätsenergien für die Erkennung und Quantifizierung einer thermodynamischen Kopplung zwischen Ereignissen in komplexen biomolekularen Systemen zu demonstrieren. Die Ergebnisse dieser Studie führten zu neuen Einsichten, die helfen könnten, die immer noch rätselhafte physiologische Rolle von PaAz zu ergründen.

In Manuskript D wurden Berechnungen freier Energien angewendet, um die Thermodynamik des Transports durch den Ammonium-/Ammoniaktransporter Amt-1 aus *Archaeoglobus fulgidus* (AfAmt-1) zu untersuchen. Ammonium ist die am direktesten verwertbare Stickstoffquelle für Pflanzen und Mikroorganismen. AfAmt-1 und seine Homologen erleichtern den Transport von Ammoniak/Ammonium über biologische Membranen in Lebewesen aus allen Reichen des Lebens. Es wird intensiv debattiert, wie diese Proteine ihre Funktion erfüllen und ob Ammonium oder seine protonierte Form Ammoniak das tatsächlich transportierte Substrat ist. Die Studie ging durch die Berücksichtigung der Effekte der Substratkonzentrationen, elektrochemischer Transmembrangradienten, gekoppelter Protonierungsgleichgewichte und der kompetitiven Bindung verschiedener Ligandenarten über vorherige theoretische Studien hinaus. Es wurde festgestellt, dass die transportierte Spezies höchstwahrscheinlich das Ammoniumion ist. Ein  $\text{NH}_3/\text{H}^+$ -Symport-Mechanismus, der ein Paar koplanarer Histidinreste im Zentrum der Transmembranpore als vorübergehenden Protonenakzeptor einbindet, erscheint wegen der hohen genetischen Konservierung dieser Reste plausibel.

Manuskript E stellt eine erste Anwendung der Mikrozustandsbeschreibung innerhalb eines CE/MM-Modells für die Simulation der Nichtgleichgewichtsdynamik eines molekularen Systems dar. Simuliert wurde die Re-Reduktionskinetik des primären Elektronendonors im Photozyklus des bakteriellen photosynthetischen Reaktionszentrums aus

*Blastochloris viridis*. Die Ergebnisse der Simulationen stimmen sehr gut mit experimentell gemessenen Daten überein.

Schlagwörter: Kontinuumselektrostatik, Berechnung freier Energien, Monte-Carlo-Simulation, thermodynamische Störungstheorie, Ligandenbindung, freie Bindungsenergie, elektrochemisches Potenzial, Membranpotenzial, Kooperativität, Transmembran-Transport, Bioenergetik

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Computer Simulation of Binding Equilibria</b>	<b>19</b>
2.1	A Formulation of Binding Theory in Terms of Electrochemical Potentials . . . . .	20
2.2	A Continuum Electrostatics / Molecular Mechanics Model for the Simulation of Receptors Within a Microstate Description . . . . .	28
<b>3</b>	<b>Synopsis of the Manuscripts</b>	<b>49</b>
<b>4</b>	<b>Manuscript A: The Monte Carlo Simulation Package GMCT</b>	<b>81</b>
<b>5</b>	<b>Manuscript B: Generalized Free Energy Perturbation Theory</b>	<b>97</b>
<b>6</b>	<b>Manuscript C: Coupling of Protonation, Reduction and Conformational Change in Azurin</b>	<b>113</b>
<b>7</b>	<b>Manuscript D: Thermodynamics of Transport through the Ammonium Transporter Amt-1</b>	<b>129</b>
<b>8</b>	<b>Manuscript E: Simulation of Electron Transfer in the Bacterial Photosynthetic Reaction Center</b>	<b>145</b>
	<b>List of Abbreviations</b>	<b>159</b>



# Chapter 1

## Introduction

The binding and transfer of ligands is of central importance for the function of many biomolecular systems. The main topic of this thesis is the development and application of Monte Carlo (MC) simulation methods for studying complex ligand binding equilibria which can also involve conformational changes. These simulation methods are applied to continuum electrostatics / molecular mechanics models of biomolecular systems. The developed methods are applied to the small electron transfer protein azurin from the bacterium *Pseudomonas aeruginosa* (PaAz) and to the ammonium transporter Amt-1 from the archaeon *Archaeoglobus fulgidus* (AfAmt-1). The final chapter of this work provides an outlook to non-equilibrium simulations at the example of the bacterial photosynthetic reaction center from *Blaschlochloris viridis* (bRC). The function of these three proteins involves aspects which are of special interest in the context of bioenergetics.

The following parts of this chapter introduce the topics of this thesis and place the thesis and its results in a broader context. The second chapter introduces the statistical thermodynamics of binding equilibria and the principles used for their simulation in the following chapters. The third chapter contains a synopsis of the individual manuscripts that form the main body of this thesis. The following chapters are formed by individual manuscripts which are published in peer reviewed scientific journals. The broader context of these manuscripts

is introduced in the following section of this chapter. The contents and interrelations of the manuscripts and their placement in the previously introduced broader context are the subjects of Chapter 3.

## **Motivation & Context**

There is a growing wealth of structural information on biomolecular systems which has led to important new insights into their function. Often however, structural information alone is not sufficient to understand the details of biomolecular function. The puzzling complexity of many biomolecules makes the principles of their function inaccessible to intuitive inference or reasoning from geometrical criteria. Also information from complementary experiments, e.g., from spectroscopic or voltammetric experiments does often not lead straightforward to understanding of biomolecular function. Theoretical methods, in particular in combination with computer simulations can help to unravel the secrets of biomolecular function by providing information that is not accessible to experiments.

## **Challenges & Questions in the Quest for Understanding of Bioenergetic Energy Transduction**

The work group of G. Matthias Ullmann is especially interested in bioenergetic systems involved in energy transduction like those involved in respiration and photosynthesis. Of particular interest from a human perspective is the function of the mitochondrial respiratory chain that resides in the cristae of the inner mitochondrial membrane (Figure 1.1). The respiratory chain consists of large multi-subunit protein complexes<sup>1,2</sup> that can in turn be organized in loose assemblies termed respiratory supercomplexes.<sup>3-7</sup> The complexes I to IV of the mitochondrial respiratory chain reside in the cristae of the inner mitochondrial membrane and couple the redox reactions to the translocation of protons across the inner mitochondrial membrane. The electron transfer is accomplished with the aid of various redox active cofactors that

are bound by the proteins (shaded in dark gray in the lower panel of Figure. 1.1). The proton transfer across the inner mitochondrial membrane results in the establishment of an electrochemical proton gradient, the so-called proton motive force (pmf).<sup>8-12</sup> The pmf can subsequently be used for secondary active transport processes across the inner mitochondrial membrane<sup>13</sup> and for the synthesis of the energy carrier ATP by ATP synthase, which is often termed complex V of the mitochondrial respiratory chain.<sup>14-16</sup>

Complex III of the respiratory chain, more precisely termed the cytochrome  $bc_1$  complex, is a particularly interesting system, which I will use in the following to illustrate the type of questions and challenges addressed in this thesis. Similar complexes function as electron transfer hubs in many different respiratory systems throughout all domains of life and in the photosynthetic apparatus of bacteria, cyanobacteria and higher plants.<sup>27-30</sup> The cytochrome  $bc_1$  complex is a multi-subunit enzyme which forms a intertwined dimer of two identical sets of three to ten subunits, depending on the source of the enzyme (Figure. 1.2). Three core subunits are involved in the catalytic function of the complex – cytochrome  $b$ , the Rieske iron-sulfur protein, and cytochrome  $c_1$ . The cytochrome  $b$  subunit binds two  $b$ -type heme cofactors in a membrane-intrinsic four-helix bundle. Heme  $b_L$  (for low reduction potential) is located close to the intermembrane side of the inner mitochondrial membrane whereas heme  $b_H$  (for high reduction potential) is located close to the matrix side. Cytochrome  $c_1$  is mostly located in the intermembrane space and binds a single  $c$ -type heme termed heme  $c_1$ . The Rieske iron-sulfur protein consists of a single transmembrane helix which ends in a flexible neck region which is connected to a globular head domain on the intermembrane side of the inner mitochondrial membrane. The globular domain binds a Rieske-type [2Fe-2S] iron-sulfur cluster which is coordinated by two histidine residues with dissociable protons and two cysteine residues.<sup>31-34</sup> Thanks to the flexible neck region, the globular domain of the Rieske subunit can move with a tilting motion between an upper position close to cytochrome  $c_1$  and a lower position close to cytochrome  $b_L$ .<sup>35</sup>

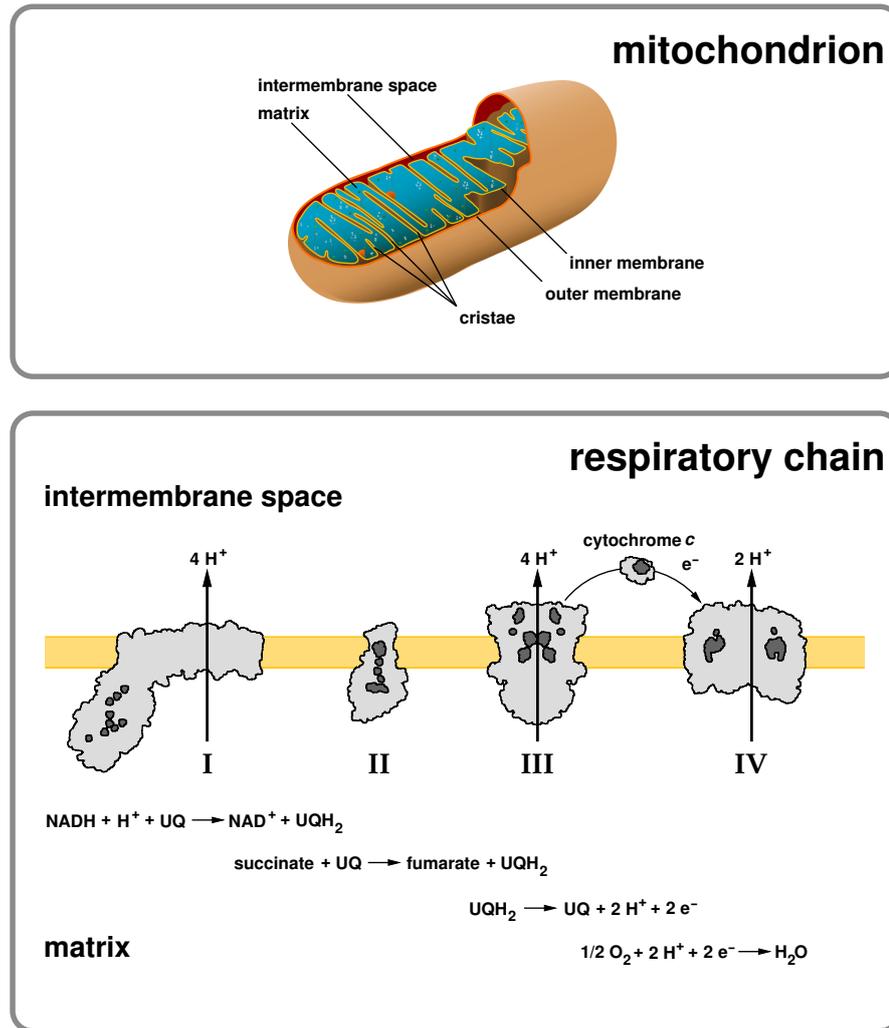


Figure 1.1: Energy transduction at the inner mitochondrial membrane. The protein complexes I to IV of the mitochondrial respiratory chain transduce biochemical energy from redox reactions into electrochemical energy stored in a proton gradient across the inner mitochondrial membrane. top: Schematic view of a mitochondrion. The picture was modified from an initial version taken from Wikimedia Commons. bottom: Schematic view of the mitochondrial respiratory electron transfer chain consisting of complexes I (NADH dehydrogenase), II (succinate dehydrogenase), III (cytochrome *bc*<sub>1</sub>) and IV (cytochrome *c* oxidase). The shapes of the proteins (light grey) and their cofactors (dark grey) are based on experimentally determined structures from the protein databank. Complex I: PDB 3M9S,<sup>17</sup> 2FUG<sup>18</sup> and 3RKO.<sup>19</sup> Complex II: PDB 1YQ3.<sup>20</sup> Complex III: PDB 1KB9<sup>21</sup> and 1EZV.<sup>22</sup> Complex IV: PDB 1V55.<sup>23</sup> cytochrome *c*: PDB 1YFC.<sup>24</sup> The position of the respiratory complexes in the inner mitochondrial membrane (orange) is taken from the OPM database.<sup>25</sup>

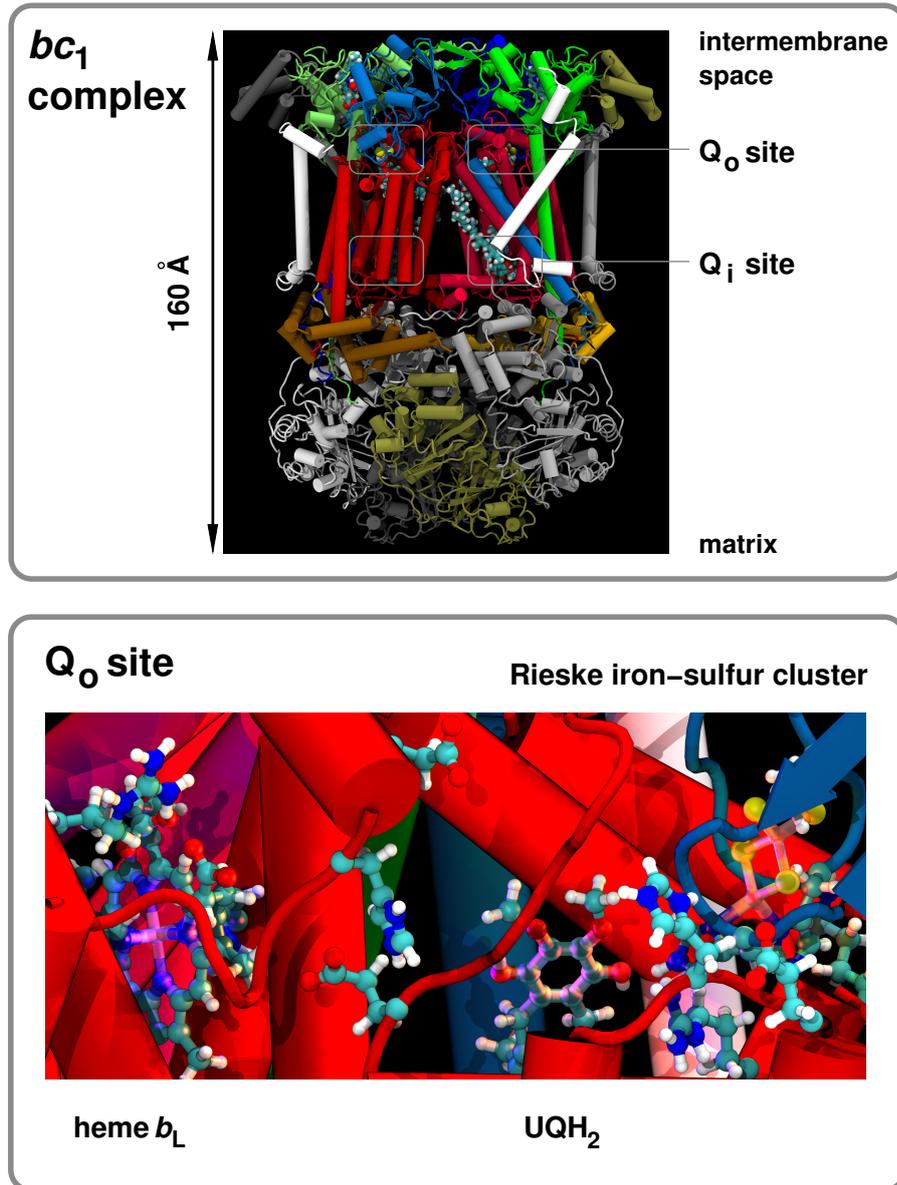
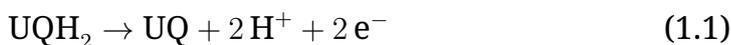


Figure 1.2: The cytochrome *bc*<sub>1</sub> complex (complex III) as example for an energy transducing biomolecular system. top: cartoon representation of the cytochrome *bc*<sub>1</sub> complex. The complex contains two copies of each subunit. The core subunits which are widely conserved throughout all domains of life are cytochrome *c*<sub>1</sub> (green), the Rieske iron-sulfur protein (blue) and cytochrome *b* (red). The Q<sub>o</sub> and Q<sub>i</sub> sites are located at the intermembrane side and the matrix side, respectively. bottom: magnified view of the Q<sub>o</sub> site. The redox-active cofactors and the substrate ubiquinole are indicated. Also shown are protonatable aminoacid sidechains in the vicinity of the active site and the protonatable propionate sidechains of heme *b*<sub>L</sub>. The figure shows the yeast cytochrome *bc*<sub>1</sub> complex (PDB 1KB9).<sup>21</sup> The position of ubiquinole was modeled by Astrid Klingen<sup>26</sup> based on the position of the inhibitor stigmatellin in PDB 1EZV.<sup>22</sup>

The overall mechanism of the complex, termed Q-cycle mechanism, is well understood, while the microscopic details of the mechanism are still enigmatic in many aspects and partly heavily debated.<sup>26–28,36–39</sup> The cytochrome  $bc_1$  complex has two active sites termed  $Q_o$  site and  $Q_i$  site which are located at the outside and the inside of the inner mitochondrial membrane, respectively (upper panel of Figure 1.2). The  $Q_i$  site is located close to heme  $b_H$ . The  $Q_o$  site, extends between the lower position of the Rieske iron-sulfur cluster and heme  $b_L$  (lower panel of Figure 1.2). At the  $Q_o$  site, the substrate ubiquinol ( $UQH_2$ ) is oxidized and deprotonated to ubiquinone (UQ).



One of the electrons is taken up by the Rieske iron-sulfur cluster, and from there, after reorientation of the globular head domain, to heme  $c_1$ . Subsequently, a small soluble cytochrome  $c$  shuttles the electron to respiratory complex IV. The second electron is taken up by heme  $b_L$  and transferred across the membrane to heme  $b_H$  at the  $Q_i$  site. There, the electron is used to regenerate ubiquinol in the reverse reaction. Thus, one half molecule of ubiquinone is re-reduced at the  $Q_i$  site for each molecule of ubiquinol oxidized at the  $Q_o$  site.

Most of the unanswered questions are connected with the microscopic details of the reactions at the  $Q_o$  site. It is unclear how the bifurcation of the electron transfer is achieved while avoiding deleterious side reactions, how the deprotonation of the substrate takes place and how the protons are released.<sup>27,28,38,40,41</sup> In addition, it is not entirely clear how the ubiquinol substrate binds in the  $Q_o$  site and whether the substrate changes its position in the binding pocket in the course of the reaction. The  $Q_o$  site contains a number of protonatable amino acid sidechains the propionate sidechains of heme  $b_L$  and the histidine ligands of the Rieske iron-sulfur cluster all of which could be involved in the proton abstraction and release. Some of the protonatable amino acid sidechains were shown to adopt multiple different rotamers.<sup>42</sup> Mutagenesis experiments on several conserved titratable amino acid sidechains remained inconclusive with regard to the identity of the involved residues and the precise course of the reaction al-

though several promising candidates were identified.<sup>43–47</sup> One more complicating factor is the hindering influence exerted by the pmf on the reactions of the Q-cycle, the so-called back-pressure effect.<sup>48–52</sup>

## **Higher Complexity and Greater System Size – “More is Different”**

Similar complications and questions as those illustrated above at the example of the cytochrome  $bc_1$  complex are also encountered with other bioenergetic systems. The work of former graduate students in our work group on this system and other systems of similar complexity made it clear that the methodology and the corresponding software available at that time lacked several important capabilities.<sup>34,53–58</sup> These features can be divided into two general areas. The first area concerns the level of detail available for the modeling of the systems. The second area concerns the availability of sophisticated methods for the simulation of the created model systems and the analysis of the generated simulation data. The need for such methods was even exacerbated by the more detailed models developed in the course of this thesis and parallel work.<sup>53</sup>

The situation is succinctly described by the title “More is Different” from a much cited work of Philip W. Anderson.<sup>59</sup> In this work, the author gives many examples for the revelation of additional physical phenomena with increasing complexity and size of the system under study. The understanding of these phenomena requires own methods, paradigms or formalism for each level of description.

Some of the problems encountered in previous work can again be illustrated at the example of cytochrome  $bc_1$ . This system contains more than 800 protonatable amino acid sidechains and 12 redox active sites. The identification of the amino acid residues responsible for the coupling of protonation and reduction in a system of this size requires automatization and an appropriate measure of cooperativity. Astrid Klöning<sup>26,34</sup> analyzed the redox-dependent titration behavior of these residues by comparing pH-dependent titration curves of the pro-

tonatable residues for different redox states of the complex. Furthermore, she analyzed the covariances between the protonation states of the amino acid residues with the aim to identify proton transfer pathways that connect the active sites to the bulk solvent phases. Although this work resulted in novel insights into the function of some amino acid residues that supplemented experimental data,<sup>43-47</sup> it did not answer all questions posed by experimental findings. In particular, protonatable sites implicated in the deprotonation of ubiquinole at the  $Q_o$  site because of the data from the above cited mutagenesis experiments were found not to exhibit strongly redox-dependent protonation behavior.

This seemingly contradictory finding may be understandable if one considers that functionally important states of the system are not necessarily frequently occurring in equilibrium. This phenomenon is not unprecedented as, for example, discussed in Manuscript E of this thesis. States of the system that are only rarely and transiently populated, are also not detectable with a covariance analysis as discussed in Manuscript C and the user manual of the program suite GMCT (see below). Free energy calculations can be applied in such cases even if a site does not frequently change its binding form in equilibrium. A free energy calculation method that is particularly well suited to study continuum electrostatics models is presented in Manuscript B. This free energy calculation method can also be employed to study the coupling of different events like the protonation and reduction of two sites as shown in Manuscript C. The novel software for MC simulations presented in Manuscript A provides a number of different MC simulation and free energy calculation methods. The use of these simulation methods is facilitated by the possibility to automate frequently occurring tasks like the calculation of protonation free energies or cooperativity free energies for large numbers of sites. The usefulness of the software is demonstrated in Manuscripts C and D at two example systems PaAz and AfAmt-1.

The first four manuscripts of this work concentrate on equilibrium thermodynamics. Manuscript E provides an outlook to time depen-

dent simulations at the example of the electron transfer dynamics in the bRC. The reader will re-encounter the problems introduced above at the example of cytochrome  $bc_1$  in Manuscripts C, D and E. The insights obtained so far with the methods developed in this thesis are encouraging for future applications to cytochrome  $bc_1$  and other systems of similar complexity.

## Outlook

Bioenergetic energy complexes are admirable not only because of the beauty of their structure but also because of their high efficiency. This efficiency is often far exceeding that of any man-made technology. Current attempts to harness this efficiency for technological applications are typically limited to direct use of isolated parts of biological systems or whole organisms. A more advanced current approach is the development of so-called biomimetics that try to imitate biological systems. Mere imitation of biological systems or even use of whole (micro)organism for technical applications in energy conversion or production of chemicals may, however, not be the best way to go. The reason for this reserve is that biological molecules and systems constructed from them are optimized by evolution to the needs of the life form, but not to the requirements of human technology. Biological systems are permanently resynthesized, recycled and repaired which requires a multitude of proteins. An example is provided by the smallest known unicellular photosynthetic life form *Prochlorococcus marinus*.<sup>60,61</sup> The core genome of this microorganism comprises about 1200 genes which is the most succinct genetic basis of photosynthesis known to date. The major technological disadvantage of using biological systems is that the energy cost of the maintenance processes reduces the overall efficiency of the systems in providing energy or energy rich compounds. This energy cost and of the energy spent in cultivating, harvesting and processing the biological systems necessarily diminishes the total energy efficiency of biotechnological energy conversion. How low the resulting net efficiency can actually be, is illustrated at the example of biofuel production from plants and algae

in reference 62.

For me, the ultimate aim of research on bioenergetics systems would thus be to understand not only the mechanism of these fascinating molecular machines but also of the design principles behind them. As a vision for the future, one could hope that such true understanding of these complex systems can ultimately be used to overcome mere imitation of biological systems by creating something new. The obtained knowledge may thus enable us one day to construct molecular machines which are tailored to human needs. This ability could help us to meet the challenges in technological energy provision posed by the necessity of environmental sustainability.

Progress in the understanding of bioenergetic systems will require the investment of much work in actual research on the systems but also in improvements of the available methodology. In the previous section, I already discussed problems encountered in theoretical studies of bioenergetic systems in equilibrium. Ultimately, the investigation of mechanistic questions will also require non-equilibrium simulation methods that provide time information. Some of the problems encountered with equilibrium simulations are recurring in the non-equilibrium case. One such example is the inability of covariances and time correlation functions to detect coupling of events that involve rarely occurring states of the system. Such events can however play an important mechanistic role (see, e.g., Manuscript E).

Clues to possible new approaches to the problems encountered in non-equilibrium simulations might be available from recently revived developments in non-equilibrium statistical mechanics.<sup>63–68</sup> These developments showed that important equilibrium properties like the partition function and the Boltzmann-Gibbs entropy have analogues in the non-equilibrium case. Therefore, it seems conceivable that other equilibrium properties and simulation methods for their calculation might also find non-equilibrium counterparts. It might thus be worthwhile to explore whether the concepts behind the free energy perturbation theory and the cooperativity free energies<sup>69</sup> presented in this thesis could find application in non-equilibrium simulations, too.

## Bibliography

- [1] Rich, P. R.; Maréchal, A. , “The mitochondrial respiratory chain”. In *Mitochondrial Function*; Brown, G. C., Murphy, M. P., Eds.; Biochemical Society / Portland Press, London, 2010; Vol. 47, pp 1–23. doi.
- [2] Papa, S.; Martino, P. L.; Capitanio, G.; Gaballo, A.; Rasmø, D.; Signorile, A.; Petruzzella, V. , “The Oxidative Phosphorylation System in Mammalian Mitochondria”. In *Advances in Mitochondrial Medicine*; Scatena, R., Bottoni, P., Giardina, B., Eds.; Springer Netherlands, 2012; Vol. 942, pp 3–37. doi.
- [3] Althoff, T.; Mills, D. J.; Popot, J.-L.; Kühlbrandt, W. *EMBO J.* **2011**, *30*, 4652–4664. doi.
- [4] Welchen, E.; Klodmann, J.; Braun, H.-P. , “Biogenesis and supramolecular organization of the oxidative phosphorylation system in plants”. In *Plant Mitochondria*; Kempken, F., Ed.; Springer, New York, 2011; Vol. 1, pp 327–355. doi.
- [5] Stuart, R. *J. Bioenerg. Biomemb.* **2008**, *40*, 411–417. doi.
- [6] Saddar, S.; Dienhart, M. K.; Stuart, R. A. *J. Biol. Chem.* **2008**, *283*, 6677–6686. doi.
- [7] Dudkina, N.; Sunderhaus, S.; Boekema, E.; Braun, H.-P. *J. Bioenerg. Biomemb.* **2008**, *40*, 419–424. doi.
- [8] Mitchell, P. *Biol. Rev.* **1966**, *41*, 445–501. doi.
- [9] Mitchell, P. *J. Theor. Biol.* **1976**, *62*, 327–367. doi.
- [10] Mitchell, P. *Biochim. Biophys. Acta - Bioenergetics* **2011**, *1807*, 1507–1538. doi.
- [11] Skulachev, V. P. *Eur. J. Biochem.* **1992**, *208*, 203–209. doi.
- [12] Demirel, Y.; Sandler, S. *Biophys. Chem.* **2002**, *97*, 87–111. doi.
- [13] Walker, J. E. *Quart. Rev. Biophys.* **1992**, *25*, 253–324. doi.
- [14] Capaldi, R. A.; Aggeler, R. *Trends Biochem. Sci.* **2002**, *27*, 154–160. doi.

- [15] Okuno, D.; Iino, R.; Noji, H. *J. Biochem.* **2011**, *149*, 655–664. doi.
- [16] Watt, I. N.; Montgomery, M. G.; Runswick, M. J.; Leslie, A. G. W.; Walker, J. E. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16823–16827. doi.
- [17] Efremov, R. G.; Baradaran, R.; Sazanov, L. A. *Nature* **2010**, *465*, 441–445. doi.
- [18] Sazanov, L. A.; Hinchliffe, P. *Science* **2006**, *311*, 1430–1436. doi.
- [19] Efremov, R. G.; Sazanov, L. A. *Nature* **2011**, *476*, 414–420. doi.
- [20] Huang, L.-S.; Sun, G.; Cobessi, D.; Wang, A. C.; Shen, J. T.; Tung, E. Y.; Anderson, V. E.; Berry, E. A. *J. Biol. Chem.* **2006**, *281*, 5965–5972. doi.
- [21] Lange, C.; Nett, J. H.; Trumpower, B. L.; Hunte, C. *EMBO J.* **2001**, *20*, 6591–6600. doi.
- [22] Hunte, C.; Koepke, J.; Lange, C.; Roßmanith, T.; Michel, H. *Structure* **2000**, *8*, 669–684. doi.
- [23] Tsukihara, T.; Shimokata, K.; Katayama, Y.; Shimada, H.; Muramoto, K.; Aoyama, H.; Mochizuki, M.; Shinzawa-Itoh, K.; Yamashita, E.; Yao, M.; Ishimura, Y.; Yoshikawa, S. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15304–15309. doi.
- [24] Baistrocchi, P.; Banci, L.; Bertini, I.; Turano, P.; Bren, K. L.; Gray, H. B. *Biochemistry* **1996**, *35*, 13788–13796. doi.
- [25] Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. *Bioinformatics* **2006**, *22*, 623–625. doi.
- [26] Klingen, A. R.; Palsdottir, H.; Hunte, C.; Ullmann, G. M. *Biochim. Biophys. Acta - Bioenergetics* **2007**, *1767*, 204–221. doi.
- [27] Mulkidjanian, A. Y. *Biochim. Biophys. Acta - Bioenergetics* **2010**, *1797*, 1858–1868. doi.
- [28] Hunte, C.; Solmaz, S.; Pálsdóttir, H.; Wenz, T. , “A Structural perspective on mechanism and function of the cytochrome  $bc_1$  complex”. In *Bioenergetics*; Springer Berlin / Heidelberg, 2008; Vol. 45, pp 253–278. doi.
- [29] Crofts, A. R.; Berry, E. A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 501–509. doi.

- [30] Baniulis, D.; Yamashita, E.; Zhang, H.; Hasan, S. S.; Cramer, W. A. *Photochem. Photobiol.* **2008**, *84*, 1349–1358. doi.
- [31] Link, T. A. *FEBS Lett.* **1997**, *412*, 257–264. doi.
- [32] Ullmann, G. M.; Noodleman, L.; Case, D. A. *J. Biol. Inorg. Chem.* **2002**, *7*, 632–639. doi.
- [33] Zu, Y.; Couture, M. M.-J.; Kolling, D. R. J.; Crofts, A. R.; Eltis, L. D.; Fee, J. A.; Hirst, J. *Biochemistry* **2003**, *42*, 12400–12408. doi.
- [34] Klingen, A. R. , “Structure-based theoretical characterisation of the redox-dependent titration behaviour of cytochrome bc1”, Ph.D. thesis, Universität Bayreuth, Germany, 2006. url.
- [35] Zhang, Z.; Huang, L.; Shulmeister, V. M.; Chi, Y.-I.; Kim, K.-K.; Hung, L.-W.; Crofts, A. R.; Berry, E. A.; Kim, S.-H. *Nature* **1998**, *392*, 677–684. doi.
- [36] Castellani, M.; Covian, R.; Kleinschroth, T.; Anderka, O.; Ludwig, B.; Trumpower, B. L. *J. Biol. Chem.* **2010**, *285*, 502–510. doi.
- [37] Khalfaoui-Hassani, B.; Lanciano, P.; Lee, D.-W.; Darrouzet, E.; Daldal, F. *FEBS Lett.* **2012**, *586*, 617–621. doi.
- [38] Crofts, A. R.; Holland, J. T.; Victoria, D.; Kolling, D. R.; Dikanov, S. A.; Gilbreth, R.; Lhee, S.; Kuras, R.; Kuras, M. G. *Biochim. Biophys. Acta - Bioenergetics* **2008**, *1777*, 1001–1019. doi.
- [39] Hong, S.; Victoria, D.; Crofts, A. R. *Biochim. Biophys. Acta - Bioenergetics* **2012**, *1817*, 1053–1062. doi.
- [40] Mulkidjanian, A. Y. *Biochim. Biophys. Acta - Bioenergetics* **2005**, *1709*, 5–34. doi.
- [41] Osyczka, A.; Moser, C. C.; Dutton, P. L. *Trends in Biochemical Sciences* **2005**, *30*, 176–182. doi.
- [42] Hunte, C.; Pálisdóttir, H.; Trumpower, B. *FEBS Lett.* **2003**, *545*, 39–46. doi.
- [43] Ritter, M.; Anderka, O.; Ludwig, B.; Mäntele, W.; Hellwig, P. *Biochem.* **2003**, *42*, 12391–12399. doi.

- [44] Ritter, M.; Pálsdóttir, H.; Abe, M.; Mäntele, W.; Hunte, C.; Miyoshi, H.; Hellwig, P. *Biochem.* **2004**, *43*, 8439–8446. doi.
- [45] Wenz, T.; Hellwig, P.; MacMillan, F.; Meunier, B.; Hunte, C. *Biochem.* **2006**, *45*, 9042–9052. doi.
- [46] Wenz, T.; Covian, R.; Hellwig, P.; MacMillan, F.; Meunier, B.; Trumpower, B. L.; Hunte, C. *J. Biol. Chem.* **2007**, *282*, 3977–3988. doi.
- [47] Covian, R.; Trumpower, B. L. *J. Biol. Chem.* **2009**, *284*, 14359–14367. doi.
- [48] Papa, S.; Lorusso, M.; Boffoli, D.; Bellomo, E. *Eur. J. Biochem.* **1983**, *137*, 405–412. doi.
- [49] Myatt, J.; Cotton, N.; Jackson, J. *Biochim. Biophys. Acta - Bioenergetics* **1987**, *890*, 251–259. doi.
- [50] Murphy, M. P.; Brand, M. D. *Eur. J. Biochem.* **1988**, *173*, 645–651. doi.
- [51] Demin, O.; Kholodenko, B.; Skulachev, V. *Mol. Cell. Biochem.* **1998**, *184*, 21–33. doi.
- [52] Rottenberg, H.; Covian, R.; Trumpower, B. L. *J. Biol. Chem.* **2009**, *284*, 19203–19210. doi.
- [53] Essigke, T. , “A continuum electrostatic approach for calculating the binding energetics of multiple ligands”, Ph.D. thesis, University of Bayreuth, 2008. url.
- [54] Krammer, E.-M. , “Computational Analysis of the Proton Transfer to the Secondary Quinone of Type II Photosynthetic Reaction Centers.”, Ph.D. thesis, Universität Bayreuth, Germany, 2008. url.
- [55] Munusami, P. , “Electrostatic and quantum chemical investigation of the proton pumping mechanism of cytochrome c oxidase”, Ph.D. thesis, Universität Bayreuth, Germany, 2008. url.
- [56] Till, M. , “Proton Transfer Networks and the Mechanism of Long Range Proton Transfer in Proteins”, Ph.D. thesis, Universität Bayreuth, Germany, 2009. url.

- [57] Dumit, V. , “Characterization of Structural Features for the Catalytic Efficiency of Ferredoxin-NADP(H)-Oxidoreductase”, Ph.D. thesis, National University of Rosario, Argentina, 2010.
- [58] Kloppmann, E. , “Structure-Function Relationship of Archaeal Rhodopsin Proteins analyzed by Continuum Electrostatics”, Ph.D. thesis, Universität Bayreuth, Germany, 2010. url.
- [59] Anderson, P. W. *Science* **1972**, *177*, 393–396. doi.
- [60] Dufresne, A. et al. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 10020–10025. doi.
- [61] Dufresne, A.; Ostrowski, M.; Scanlan, D.; Garczarek, L.; Mazard, S.; Palenik, B.; Paulsen, I.; de Marsac, N.; Wincker, P.; Dossat, C.; Ferriera, S.; Johnson, J.; Post, A.; Hess, W.; Partensky, F. *Genome Biology* **2008**, *9*, R90. doi.
- [62] Michel, H. *Angew. Chem. Int. Edit.* **2012**, *12*, 2516–2518. doi.
- [63] Jaynes, E. T. *Phys. Rev.* **1957**, *106*, 620–630. doi.
- [64] Jaynes, E. T. *Phys. Rev.* **1957**, *108*, 171–190. doi.
- [65] Jaynes, E. T. In *Information Theory and Statistical Mechanics*; W. A. Benjamin, New York, 1963; Vol. 3, pp 181–218.
- [66] Jaynes, E. T. In *The Gibbs Paradox*; Kluwer Academic Publishers, 1992; Chapter 1, pp 1–22.
- [67] Jaynes, E. T. , “Macroscopic Prediction”. In *Complex Systems – Operational Approaches in Neurobiology, Physics and Computers, Proceedings of the International Symposium on Synergetics at Schloss Elmau, Bavaria, May 6-11, 1985*; Haken, H., Ed.; Springer, Berlin, 1985; Vol. 32, pp 254–269.
- [68] Ge, H.; Presse, S.; Ghosh, K.; Dill, K. A. *J. Chem. Phys.* **2012**, *136*, 064108. doi.
- [69] Ben-Naim, A. *Cooperativity and Regulation in Biochemical Processes*; Kluwer Academic / Plenum Publishers, New York, 2001.



## Chapter 2

# Computer Simulation of Binding Equilibria: Theory and Practical Implementation

This section contains a brief description of the theoretical basis for the description of binding equilibria within the program suite GMCT and the program GCEM. GMCT is a program package for MC simulations on complex (bio)molecular receptors (Manuscript A). GCEM automates the preparation of the necessary program input for GMCT from a continuum electrostatics / molecular mechanics (CE/MM) model. The theoretical framework behind this software extends earlier work considering protonation and redox equilibria in our<sup>1-6</sup> and other groups.<sup>7-14</sup> In the next section, I present the theoretical basis of the formalism. The formalism is derived in terms of the (electro)chemical potentials of all system constituents. This approach leads to a formalism that is especially simple and transparent which facilitates the interpretation of simulation results. Ultimately, the novel formalism makes it easier to understand the binding behavior of complex receptors. The novel formalism leads to the energy function used in GMCT. The CE/MM model for the practical implementation of the energy function, as implemented in GMCT and GCEM, is presented in Section 2.2.

## 2.1 A Formulation of Binding Theory in Terms of Electrochemical Potentials

In this section, we show how binding equilibria can be described quantitatively starting from basic thermodynamic quantities and arguments. We start from the electrochemical potentials of the different compound species, derive the (semi-)grand canonical partition function of the system employing basic thermodynamic arguments and finally recover the classical binding polynomial. We conclude the section with a discussion of possible limitations of the current theory and point out largely unexplored areas of binding theory which could be interesting for future studies.

### Introduction of the Basic Quantities used in the Theoretical Description of the Receptor

The thermodynamics of binding processes can be described in terms of the electrochemical potentials of ligands and receptor species. The chemical potential  $\mu_i$  of a chemical species  $i$  is given by

$$\mu_i = \mu_i^\circ + \beta^{-1} \ln a_i \quad (2.1)$$

where  $\mu_i^\circ$  is the standard chemical potential of the species. The activity is defined via Eq. (2.1) as  $a_i \equiv \exp[-\beta(\mu_i^\circ - \mu_i)]$ . Different measures of concentration can be used to state an expression for the activity. Here, we use a statement in terms of molarity or molar concentration  $c$  (in mol/l) which is the most commonly used convention for compounds in solution. The activity is given by

$$a_i = \gamma_i \frac{c_i}{c^\circ} \quad (2.2)$$

where  $c_i^\circ$  is the standard concentration (1 mol/l) and  $\gamma_i$  is the activity coefficient defined as  $\gamma_i \equiv a_i \frac{c^\circ}{c_i}$ . Under standard conditions, the chemical potential is equal to the standard chemical potential. The activity coefficient corrects for non-ideal behavior of the solution and is a function of the concentration that is specific to the chosen measure of con-

centration and to the specific choice of the standard state. Such non-ideal behavior can arise from interactions among the molecules that form the solution. Specific interactions, for example those involved in binding reactions, should ideally be modeled explicitly by explicitly including all possible, relevant compound species in the model. Unspecific interactions, such as those between mobile ions, can either be included in the standard chemical potential with an appropriate choice of the standard state or included in the activity coefficients. If all non-specific interactions are included in the standard chemical potential and all specific interactions are explicitly accounted for, all the activity coefficients should be equal to one.

The electrochemical potential extends the chemical potential by the interaction of an electrostatic potential with the charge distribution of the species

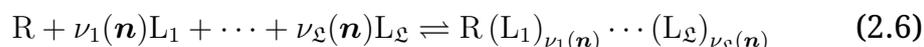
$$\bar{\mu}_i = \bar{\mu}_i^\circ + \beta^{-1} \ln a_i \quad (2.3)$$

$$\bar{\mu}_i^\circ = \mu_i^\circ + z_i F \phi_i \quad (2.4)$$

written here for an ion with a formal charge  $z_i$ , where  $F$  is the Faraday constant and  $\phi_i$  is the electrostatic potential at the spatial location species  $i$ . The Gibbs free energy  $G$  of the system is related to the electrochemical potentials (partial molar free energies)  $\bar{\mu}_i$  and the stoichiometric coefficients  $\nu_i$  of all components  $i$  present in the system via<sup>15-17</sup>

$$G = \sum_i \nu_i \bar{\mu}_i \quad (2.5)$$

where the electrochemical potentials of the compounds are mutually interdependent, that is, adding a compound to the system or altering the configuration of the system can shift the electrochemical potentials of all species. The reaction equation for the formation of a complex species or microstate  $\mathbf{n}$  from an unbound receptor  $R$  to  $\mathcal{L}$  different types  $m$  of ligands  $L_m$  with varying stoichiometric coefficients  $\nu_m(\mathbf{n})$  can be written as



There is a number of different microstates  $n$  of the receptor. These microstates may differ in the number of bound ligands of each type, their conformation, their arrangement in the complex or in the conformation of the receptor scaffold itself.

The molar Gibbs energy Eq. (2.5) of the receptor in its environment is calculated for a specific microstate. Here, the term microstate indicates, that all explicitly considered variables have a specific value. Thus the precise definition of what a microstate is depends on how detailed the model system describes the actual system. The detail of traditional protonation state calculations was, for example, often limited to the total number of protons bound to each protonatable site and thus had to use an average, “smeared out” charge distribution for carboxylic acids. A more detailed model, like the one used in this thesis, might differentiate between specific positions of the bound protons within the binding site, e.g., between the possible proton positions at the two oxygen atoms of a carboxylic acid sidechain. The energetic contributions of the remaining degrees of freedom which are not considered explicitly are assumed to be separable from those of the explicitly considered degrees of freedom. Similar approximations are commonly made in the formulation of molecular partition functions.<sup>18,19</sup> The thermodynamic average over the degrees of freedom which are not considered explicitly is included in the chemical potential of the receptor microstate. Electrostatic energies calculated from the Poisson-Boltzmann equation contain, for example, thermodynamic averages over the orientation of dipoles in receptor and solvent and over the distribution of mobile ions in the solvent. Thus, the microstate energy derived from a continuum electrostatic model like the one described in the following sections can with some right already be seen as a free energy as done in previous work<sup>1</sup> and the related work of other authors.<sup>20-24</sup> In the following, we will use the neutral symbol  $E$  for microstate energies to better distinguish them from a fully qualified free energy which contains also an average over the receptor microstates (see below).

## Equilibrium distribution of the receptor forms

We consider a system containing a receptor and a number of different ligand species that can be bound by the receptor. The receptor in various binding forms and the ligands are considered explicitly through their chemical potentials. All other constituents of the system are considered implicitly through their influence on the electrochemical potentials of the receptor and the ligands. For each ligand, we consider a single electrochemical potential without further distinction of different forms of the ligand. The ligand chemical potentials are held constant, which can be achieved experimentally by combining methods for monitoring and adjusting the ligand activity. The proton activity can, for example, be monitored with a pH electrode and adjusted through the addition of acid or base. The receptor with any bound ligands is considered in greater detail, where the detail level of the description can be chosen as needed. Each explicitly considered receptor form is characterized by its electrochemical potential. The activity coefficients of all compounds are assumed to be constant within the range of possible concentrations. This assumption will usually be well fulfilled if two conditions are met. The first condition is that all relevant specific interactions must be considered by explicitly including the corresponding bound species in the model. The second condition is that all unspecific interactions of a compound, e.g., unspecific electrostatic interactions with mobile ions in a solution, must be independent of the compounds concentration. This condition is usually described by requiring a dilute solution of the compound, where the rationale is that the compound itself is so rare that its non-specific interactions with compounds of the same type and other explicitly considered compounds are so rare as to be negligible. In dilute solution, the compounds will, for example, not contribute significantly to the ionic strength of the solution.

Our knowledge on the composition of the receptor forms and their electrochemical potentials can be combined with constraints imposed by the notion of chemical equilibrium and mass conservation to deter-

mine the probabilities of finding each receptor species in equilibrium. Each receptor form  $n$  can be formed from the reference form of the receptor and any additionally bound ligands as defined by the reaction Eq. (2.1). In thermodynamic equilibrium, the sum of the electrochemical potentials of all compounds appearing on both sides of the reaction equation are equal<sup>15,16</sup>

$$\sum_i^{\text{reactants}} \nu_i \bar{\mu}_i = \sum_j^{\text{products}} \nu_j \bar{\mu}_j \quad (2.7)$$

Writing this equation for each receptor forms gives us a system of  $\mathfrak{N}$  equations with their activities as unknowns

$$\bar{\mu}_{\text{ref}} + \sum_m^{\mathfrak{L}} \nu_m(\mathbf{n}) \bar{\mu}_m = \bar{\mu}_n \quad (2.8)$$

where  $\bar{\mu}_{\text{ref}}$  and  $\bar{\mu}_n$  are the electrochemical potentials of the receptor in the reference form and form  $n$ , respectively. The electrochemical potential of the unbound ligand of type  $m$  and the number of this ligand bound by the receptor form  $n$  are given by  $\bar{\mu}_m$  and  $\nu_m(\mathbf{n})$ , respectively. Here we again have to consider all ligand types  $m$  numbered consecutively from 1 to  $\mathfrak{L}$ , where the stoichiometric coefficient can also adopt a value of zero. In the next step, we rearrange Eq. (2.8) by expanding the chemical potentials using Eqs. (2.2) and (2.3), and collecting terms such that all constant terms are gathered on the right-hand side while all variable terms are gathered on the left-hand side of the equation

$$\ln \left[ \frac{c_n}{c_{\text{ref}}} \right] = -\beta \left( \bar{\mu}_n^\circ - \bar{\mu}_{\text{ref}}^\circ + \ln \left[ \frac{\gamma_n}{\gamma_{\text{ref}}} \right] - \sum_m^{\mathfrak{L}} \nu_m(\mathbf{n}) \bar{\mu}_m \right) \quad (2.9)$$

The final constraint equation is obtained by taking the natural exponential of both sides of the equation and making two rearrangements. First, we make the substitution

$$E_n^{\text{int}} = \bar{\mu}_n^\circ - \bar{\mu}_{\text{ref}}^\circ + \ln \left[ \frac{\gamma_n}{\gamma_{\text{ref}}} \right] \quad (2.10)$$

where  $E_n^{\text{int}}$  is the intrinsic energy of the receptor form  $n$ .<sup>25</sup> Second, we multiply both sides of the equation with  $c_{\text{ref}}$  and obtain

$$c_n = c_{\text{ref}} \exp \left[ -\beta \left( E_n^{\text{int}} - \sum_m^{\mathfrak{L}} \nu_m(\mathbf{n}) \bar{\mu}_m \right) \right] \quad (2.11)$$

An additional constraint equation comes from the requirement of mass conservation<sup>15,16</sup> which demands that the total concentration of the receptor in all of its forms is constant

$$c_{\text{R}}^{\text{tot}} - \sum_{\mathbf{n}} c_{\mathbf{n}} = 0 \quad (2.12)$$

Eqs. (2.11) and Eq. (2.12) form a completely determined system of equations which we will solve now to obtain the equilibrium concentrations of the individual receptor forms. The solution is obtained by summation of the set of equations obtained from Eq. (2.11) written for each of the  $\mathfrak{N}$  forms of the receptor and Eq. (2.12) to obtain

$$c_{\text{R}}^{\text{tot}} = c_{\text{ref}} \sum_{\mathbf{n}} \exp \left[ -\beta \left( E_{\mathbf{n}}^{\text{int}} - \sum_m^{\mathfrak{L}} \nu_m(\mathbf{n}) \bar{\mu}_m \right) \right] \quad (2.13)$$

The sum on the right-hand side of Eq. (2.13) can be identified as the (semi)-grand canonical partition function of the receptor-ligand system denoted by the symbol  $\mathcal{Z}$

$$\mathcal{Z} = \sum_{\mathbf{n}} \exp [-\beta E_{\mathbf{n}}] \quad (2.14)$$

where

$$E_{\mathbf{n}} = E_{\mathbf{n}}^{\text{int}} - \sum_m^{\mathfrak{L}} \nu_m(\mathbf{n}) \bar{\mu}_m \quad (2.15)$$

is the state energy of the receptor in form  $\mathbf{n}$ . Solving Eq. (2.13) for the concentration of the reference form and substituting according to Eq. (2.14) gives

$$c_{\text{ref}} = \frac{c_{\text{R}}^{\text{tot}}}{\mathcal{Z}} \quad (2.16)$$

Equation (2.16) can be used to substitute for  $c_{\text{ref}}$  in Eq. (2.11) to finally obtain the equilibrium concentration of any form  $\mathbf{n}$  of the receptor

$$c_{\mathbf{n}} = c_{\text{R}}^{\text{tot}} \frac{\exp [-\beta E_{\mathbf{n}}]}{\mathcal{Z}} \quad (2.17)$$

In classical binding theory, the partition function given by Eq. (2.14) is written as polynomial in the ligand activities, hence called the binding polynomial.<sup>26,27</sup> To obtain the binding polynomial, we group all

receptor forms into binding macrostates  $a$  which are defined by the having bound the same number of ligands of each type, i.e., by having the same set of stoichiometric coefficients. The standard free energy of a receptor form  $n$  relative to the reference state of the receptor is given by

$$E_n^\circ = E_n^{\text{int}} - \sum_m^{\mathcal{L}} \nu_m(\mathbf{n}) \bar{\mu}_m^\circ \quad (2.18)$$

which is identical to the state energy  $E_n$  of the form except for the omission of the activity terms of the ligand electrochemical potentials. The standard free energy of a binding macrostate  $a$  relative to the reference state of the receptor is then given by

$$E_a^{\circ, \text{macro}} = \beta^{-1} \ln \left[ \sum_n^{\mathfrak{N}_a} \exp[-\beta E_n^\circ] \right] \quad (2.19)$$

where the sum runs over all  $\mathfrak{N}_a$  forms of the receptor that belong to a given binding macrostate  $a$ . The binding polynomial is recovered by writing the partition function of the system in terms of the standard free energies of the binding macrostates

$$\mathcal{Z} = \sum_a^{\mathfrak{N}^{\text{macro}}} \left( \exp[-\beta E_a^{\circ, \text{macro}}] \prod_m^{\mathcal{L}} a_m^{\nu_m(a)} \right) \quad (2.20)$$

where the sum runs over all  $\mathfrak{N}^{\text{macro}}$  macroscopic binding forms.

## Calculation of Observables and Reaction Free Energies

The thermodynamic average of an observable  $O$  can be computed from

$$\langle O \rangle = \mathcal{Z}^{-1} \sum_n^{\mathfrak{N}} O_n \exp[-\beta E_n^{\text{micro}}] \quad (2.21)$$

where  $O_n$  is the value of the observable in microstate  $n$ . 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. The number  $\mathfrak{N}_a$  of selected microstates of

a substate  $a$  form again an ensemble of microstates with the partition function

$$Z_a = \sum_n \delta_{n,a} \exp[-\beta E_n^{\text{micro}}] \quad (2.22)$$

where  $\delta_{n,a}$  is equal to 1 if the microstate  $n$  matches the definition of the substate  $a$  and equal to 0 otherwise. The probability of finding the receptor within this substate is given by

$$p_a = \frac{Z_a}{Z} \quad (2.23)$$

Often, the free energy change for a transformation between two substates will be of interest. Such a transformation can for example be used to describe a binding reaction, or a conformational change. The free energy change for the transformation of substate  $a$  to substate  $b$  is given by

$$\Delta G_{a \rightarrow b} = -\beta^{-1} \ln \left[ \frac{Z_b}{Z_a} \right] \quad (2.24)$$

The formalism given in this section applies also to binding equilibria in complex receptors with many binding sites. In practice, however, there is a major problem in applying the formalism to complex binding equilibria in macromolecular systems of interest. The problem arises from the combinatorial explosion of the number of microstates which rapidly occurs with a growing number of binding sites. The energy of each microstate needs to be known, and the evaluation of the thermodynamic averages or partition functions require summations over all the relevant microstates. These calculations are computationally prohibitive if the number of microstates is large, which can already occur in systems of relatively small size. Properties of large systems can be computed with the aid of simulation methods as described in Manuscript A.

## 2.2 A Continuum Electrostatics / Molecular Mechanics Model for the Simulation of Receptors Within a Microstate Description

Continuum electrostatics/molecular mechanics models provide insights into biomolecular function at a much lower computational cost than all-atom MD simulations.<sup>1,20,28,29</sup> Our conceptual model for the description of binding equilibria is based on a microstate description of the receptor-ligand system.<sup>5,14,20,30</sup> The receptor with the bound ligands is modeled in atomic detail while the environment of the receptor including the unbound ligands is described by average statistical mechanics within a continuum electrostatics model.<sup>1,20,28,29</sup> The continuum electrostatics part of our model can be augmented with terms from a molecular mechanics force field to account for the conformational energy of receptor parts and apolar interactions between parts of the receptor. The CE/MM model can be extended such that parts of the receptor which are of special interest, like the active site of an enzyme, can be treated with a quantum mechanical method. Since this possibility is not used in this thesis, it will in the following parts of this chapter only be mentioned where it is necessary to document the corresponding features offered by our software.

Specifically, the continuum electrostatics part of our model uses the linearized Poisson-Boltzmann equation (LPBE)

$$\nabla [\varepsilon(\mathbf{r}) \nabla \phi] = -4\pi \rho(\mathbf{r}) + \bar{\kappa}^2(\mathbf{r}) \phi(\mathbf{r}) \quad (2.25)$$

which relates the electrostatic potential  $\phi$  at a spatial coordinate  $\mathbf{r}$  to the absolute dielectric permittivity  $\varepsilon$  and the inverse Debye length  $\bar{\kappa}^2$  that both also depend on the spatial position. The inverse Debye length is given by

$$\bar{\kappa}^2 = 8\pi\beta N_A e_o^2 I \quad (2.26)$$

where  $\beta = (k_B T)^{-1}$  is the inverse thermal energy,  $N_A$  is the Avogadro constant,  $e_o$  is the elementary charge, and  $I$  is the ionic strength which

is given by

$$I = \frac{1}{2} \sum_i c_i z_i^2 \quad (2.27)$$

where the sum runs over all types of mobile ions  $i$  with the concentration  $c_i$  and the formal charge  $z_i$ . The absolute dielectric permittivity is given by the product of the dielectric permittivity of the vacuum  $\varepsilon_0$  and the relative dielectric permittivity  $\varepsilon_r$  of the material at the respective spatial position

$$\varepsilon(\mathbf{r}) = \varepsilon_0 \varepsilon_{\text{rel}}(\mathbf{r}) \quad (2.28)$$

The relative dielectric permittivity is often also referred to as relative dielectric constant or simply dielectric constant of the material.

An important property of the LPBE is the additivity of electrostatic potentials for multiple sets of partial charges. This property allows us to decompose the microstate energy into multiple terms that can be precomputed. The full microstate energy can then be computed by a simple summation of the terms during a simulation, which avoids a constant repetition of the computationally relatively expensive continuum electrostatics calculations.

## The Calculation of the Electric Transmembrane Potential Distribution

Two equivalent ways exist to calculate the spatial distribution of the transmembrane potential from a finite difference solution to the LPBE. One possibility is the adding of a constant electrostatic potential offset equal to the net trans-membrane potential  $\Psi$  for all ion accessible grid points at the inner membrane side as described by Roux.<sup>31</sup> The second possibility is the assignment of an effective uniform charge distribution which represents the potential offset to all ion accessible points at the inner membrane side or both membrane sides.

The second possibility is the one actually implemented in the PBEQ module<sup>31</sup> of the CHARMM program,<sup>32</sup> APBSmem<sup>33,34</sup> and the modified version of the MEAD library<sup>20,35</sup> that originated from this thesis. The MEAD implementation for the calculation of the electric transmembrane potential constructs the effective charge distribution by adding

a charge

$$q^{\text{eff}} = \pm \frac{\bar{\kappa}^2 h^3 \Psi}{8\pi} \quad (2.29)$$

to each ion accessible grid point at the inner or outer membrane side, respectively. The symbol  $h$  denotes the grid spacing of the cubic grid used in the finite difference scheme and  $\bar{\kappa}^2$  is the inverse Debye length (see Eq. (2.26)). The electric transmembrane potential ranges from  $-0.5 \Psi$  in the bulk outer phase to  $+0.5 \Psi$  in the bulk inner phase. The linear dependence of  $\Delta E^{\text{gc}, \Psi}(c)$  on the magnitude of the membrane potential results from the properties of the LPBE.<sup>31</sup> The membrane potential dependent energy terms computed by GCEM are expressed according to the electrophysiological convention.<sup>36</sup> According to the electrophysiological convention, the magnitude of the transmembrane potential is measured relative to the electrostatic potential in the bulk outer phase, where bulk means far from the membrane. According to this convention, the electric transmembrane potential ranges from a value of zero in the bulk outer solvent phase to a value of  $\Psi$  in the bulk inner solvent phase.

## The Microstate Energy Function

A schematic view of the receptor model described herein is depicted in Figure 1 of Manuscript A. The receptor is partitioned into a number of variable regions that are termed sites and the largely invariable remainder of the receptor that is termed background. The background can adopt different global conformations denoted by the symbol  $c$ . The form adopted by a site is termed instance in GMCT and GCEM. An instance is characterized by the number of bound ligands of each type, the binding topology, the charge distribution, and the local conformation. 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 instances adopted by the sites are arranged in the so-called state vector  $\vec{s}$ . The specific instance adopted by a site  $i$  is given by the the corresponding entry of the state vector  $s_i$ .

The term microstate description means that each constituent of the system adopts a well defined form at any given point of a simulation. The microstate of the receptor is thus defined by

$$\mathbf{n} = \{c, \vec{\mathbf{s}}\} \quad (2.30)$$

The partitioning of the receptor allows us to express the energy of a receptor microstate as a sum of precomputed energy terms (see above). The microstate energy Eq. (2.15) is rewritten for our partitioned receptor as

$$E_{\mathbf{n}}^{\text{micro}} = E_c^{\text{conf}} + \sum_{i=1}^{\text{Nsites}} \left( E^{\text{int,r}}(c, \mathbf{s}_i) - \sum_m^{\mathcal{L}} \nu_m(c, \mathbf{s}_i) \bar{\mu}_m \right) + \sum_{i=1}^{\text{Nsites}} \sum_{j=1}^{j < i} W(c, \mathbf{s}_i, \mathbf{s}_j) \quad (2.31)$$

This microstate energy function can be physically interpreted in terms of a step-wise assembly process of the receptor-ligand system from the separate constituents as described in Manuscript A. The contribution of the background to the microstate energy is given by the global conformational energy  $E_c^{\text{conf}}$ . The term  $E^{\text{int,r}}(c, \mathbf{s}_i)$  is the intrinsic energy of instance  $\mathbf{s}_i$  of site  $i$  in the receptor environment. The intrinsic energies depend on the global conformation  $c$  adopted by the receptor because the sites may interact with the background. The first sum on the right-hand side of Eq. (2.31) accounts for the energy cost for the removal of the bound ligands from the respective solvent phase. Here,  $\nu_m(c, \mathbf{s}_i)$  is the number of ligands of type  $m$  bound by the site  $i$  in instance  $\mathbf{s}_i$ . The term  $\bar{\mu}_m$  denotes the electrochemical potential of the ligand type  $m$ . The ligand type does also distinguish between ligands bound from two separate compartments with different electrochemical potential. In this way, the model can account for chemical potential differences of ligands between two compartments separated by a membrane.<sup>3,4</sup> The last term on the right hand side of Eq. (2.31) accounts for the mutual influence of the sites that shifts their electrochemical potentials. Here, restriction of the summation to  $j < i$  ensures that each interaction is counted only once. The mutual influence of the sites that shifts their electrochemical potentials is modeled with pair-wise interaction energies between the sites. The symbol  $W(c, \mathbf{s}_i, \mathbf{s}_j)$  denotes the inter-

action energy between sites  $i$  and  $j$  each of which adopts a particular instance  $s_i, s_j$ , respectively. The interaction energies depend on  $c$  because an alteration of the global conformation can change the spatial arrangement of the sites and their accessibility to solvent or lipid molecules in the receptor environment.

## **The Calculation of the Microstate Energy Terms from a CE/MM Model**

The CE/MM model described herein is implemented in the program GCEM which is based on a modified version of the MEAD library.<sup>20,35</sup> This program implements the full CHARMM force field for the molecular mechanics part<sup>32</sup> and the LPBE (Eq. (2.25)) for the continuum electrostatics part. Although the full CHARMM force field is implemented, the molecular mechanics part does not use bond and angle energy terms by default. This approach was chosen to avoid spurious energy differences that can arise from insignificantly small geometrical differences between the static structures of different forms of a receptor constituent.

The LPBE is solved numerically with a finite-difference method on cubic grids.<sup>37-40</sup> Detailed derivations of the method are, e.g., given in references 41 and 42. The calculation of the electric transmembrane potential distribution with the LPBE is discussed above. The continuum electrostatics terms that depend on the electric transmembrane potential apply of course only to receptor systems that contain a membrane, e.g., to integral membrane proteins. These terms adopt values of zero for other systems. The physical significance of these terms is discussed separately below.

There are two principal types of membrane potential independent continuum electrostatics terms used in the CE/MM model. The first type are interaction energies between two sets of atomic partial charges. These interactions can be weakened through the response of the dielectric material that surrounds the interacting charge distributions. The dielectric response results in a so-called reaction field, which is

the electrostatic potential exerted by the dielectric material on an embedded charge distribution. The microscopic origins of the reaction field are the reorientation of permanent electric dipoles within the environment and the induction of electric dipoles through electronic polarization of the environment. Permanent electric dipoles in the environment are dipolar solvent molecules and dipolar moieties of the receptor and other molecules (e.g., lipids). The second type of continuum electrostatics terms used herein is the energy of the charge distribution that induced the reaction field within the reaction field. This type of energy term is commonly referred to as reaction field energy or Born energy. Microscopically, this type of continuum electrostatics terms describes the interaction of a set of atomic partial charges with the induced electric dipoles or reoriented permanent electric dipoles in the environment.

### **The Intrinsic Energy of a Site**

The intrinsic energy of a site in the receptor environment  $E^{\text{int,r}}(c, s_i)$ , is computed relative to the intrinsic energy of the model compound in a reference environment  $E^{\text{int,m}}(c, s_i)$ .<sup>7,43</sup> The quantity  $E^{\text{int,m}}(c, s_i)$ , is called model energy, because it is known from experiments or calculations on model compounds in a reference environment. The shift of the intrinsic energy upon transfer of the model compound from the reference environment to the receptor environment can be computed from the thermodynamic cycle depicted in Figure 2.1.<sup>7,43</sup> According to Hess's law, the sum of the energy differences in one direction around the cycle is equal to zero. This property can be exploited to compute the intrinsic energy shift. The intrinsic energy shift can be decomposed into different energy terms whose functional form and physical meaning are discussed in the following.

First, we split the thermodynamic cycle for the calculation of the intrinsic energy shift into two thermodynamic cycles (Figure 2.2). The rationale for this splitting is that the model energy does not always include the energy difference between different conformers of the site. An experimentally determined model energy will normally not dis-

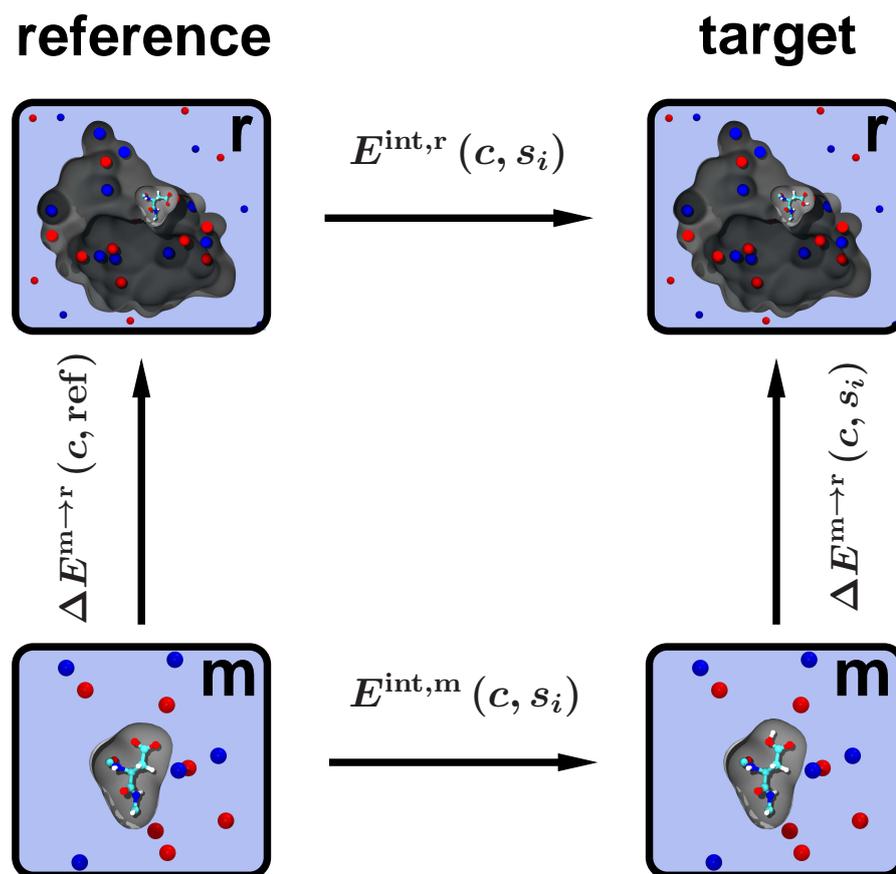


Figure 2.1: Thermodynamic cycle for the calculation of the intrinsic energy  $E^{\text{int},r}(c, s_i)$  in the receptor environment relative to the intrinsic energy  $E^{\text{int},m}(c, s_i)$  for an appropriate model compound in a reference environment (model energy). The model energy can be computed from experimental data (e.g., a  $\text{p}K_a$  value measured in aqueous or another solvent<sup>1</sup>) or computed with a quantum chemical approach (e.g., with density functional theory<sup>44</sup>). The transfer of the site from the reference environment of the model compound (m) to the receptor environment (r) leads to altered interactions with the surrounding environment. For simplicity, the receptor environment is illustrated here at the example of a small soluble protein in aqueous solution. The model compound can also be surrounded by some solvent (usually water). The solvents surrounding the receptor and the model compound can contain mobile ions symbolized by the small red and blue spheres. The inserted site also interacts with the receptor background. This interaction is usually dominated by electrostatic interactions of fixed charges within the site and within the receptor background (larger red and blue spheres). Interactions with other sites are treated separately.

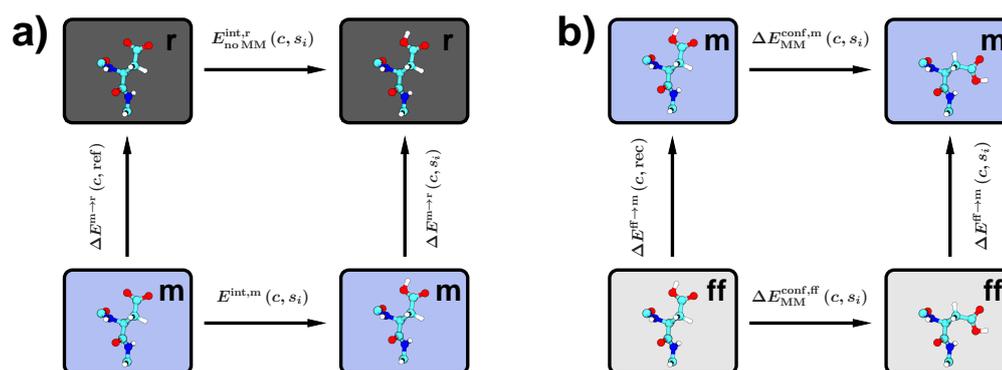


Figure 2.2: The thermodynamic cycle in Figure 2.1 is split into two thermodynamic cycles for the calculation of the intrinsic energy shift. The labels within the boxes denote the respective environment (m: reference environment of the model compound, r: receptor environment, ff: force field environment). a) Intrinsic energy shift due to the transfer of the site from the reference environment of the model compound to the receptor environment. The reference form (ref) and is shown on the left-hand side of the thermodynamic cycle and the target form ( $s_i$ ) is shown on the right-hand side. b) Intrinsic energy shift due to a conformational change between different instances that belong to the same binding form (if not already included in the model energy as in the case of a quantum chemically computed model energy). The reference conformer (rec) is shown on the left-hand side of the thermodynamic cycle and the target conformer ( $s_i$ ) is shown on the right-hand side.

tinguish between different conformers of the site. In this case, the conformational energy can be computed within the CE/MM approach via the thermodynamic cycle depicted in Figure 2.2b. A model energy that was computed with a quantum chemical approach can account directly for the conformational energy and thus will not require an additional calculation of this energy contribution. The model energy is assumed here to define the energy between two binding forms of a site, where the thermodynamic average over the conformational ensemble of each binding form is implicitly included. Here, a binding form is described by all possible attributes of an instance (see above) except for the local conformation. In contrast, the binding form becomes synonymous to the instance if the model energies of a site already account for the relative conformational energy of its instances, e.g. in case of a quantum chemically treated site. For simplicity, we assume here that the conformational energy of the conformer with the lowest energy (reference conformer) is equal to zero. This assumption will not affect the relative intrinsic energies of the instances of a site. Thus, the equilibrium distribution of the instances will be preserved.

In terms of the thermodynamic cycles depicted in Figure 2.2, the intrinsic energy of an instance  $s_i$  of a site  $i$  is given by

$$E^{\text{int,r}}(c, s_i) = E_{\text{noMM}}^{\text{int,r}}(c, s_i) + \Delta E_{\text{MM}}^{\text{conf,m}}(c, s_i) \quad (2.32)$$

where the  $s_i$  denotes the considered instance of the site  $i$  and  $\text{rec}$  denotes the reference conformer of the same binding form. The symbol  $E_{\text{noMM}}^{\text{int,r}}(c, s_i)$  denotes the intrinsic energy of  $s_i$  within the receptor environment excluding its relative conformational energy  $\Delta E_{\text{MM}}^{\text{conf,m}}(c, s_i)$ . The energy difference  $\Delta E^{\text{m} \rightarrow \text{r}}(c, x)$  accounts for the transfer of instance  $x$  from the reference environment of the model compound to the receptor environment. The labels  $\text{m}$ ,  $\text{r}$  and  $\text{ff}$  in the superscripts of the energies in Figures 2.1 and 2.2 denote the model environment, the receptor environment and the force field parametrization environment, respectively. The superscripts  $a \rightarrow b$  denote the transfer of an instance from an environment  $a$  to an environment  $b$ . The subscript  $\text{noMM}$  indicates that the respective energy does not contain a separately computed relative conformational energy contribution (see the thermody-

dynamic cycle depicted in Figure 2.2a). The subscript MM indicates an energy difference that is used in the calculation of the relative conformational energy with the aid of a molecular mechanics force field (Figure 2.2b).

The addition of a constant offset to the intrinsic energies of all instances of a site does not affect their equilibrium distribution which is only determined by the relative values of the intrinsic energies. This property is exploited in GCEM to express the intrinsic energy as

$$E^{\text{int},r'}(c, \mathbf{s}_i) = E^{\text{int},r}(c, \mathbf{s}_i) + \Delta E^{\text{m} \rightarrow \text{r}}(c, \text{ref}) \quad (2.33)$$

The use of  $E^{\text{int},r'}(c, \mathbf{s}_i)$  is advantageous because one can in this case retrace the origin of the intrinsic energy shifts to the instance of interest or the reference instance, which is not possible with  $E^{\text{int},r}(c, \mathbf{s}_i)$ . This possibility facilitates the analysis of the physical origins of the binding behavior exhibited by a site. Therefore, the intrinsic energies calculated by GCEM correspond to  $E^{\text{int},r'}(c, \mathbf{s}_i)$ . One can easily obtain  $E^{\text{int},r}(c, \mathbf{s}_i)$  from  $E^{\text{int},r'}(c, \mathbf{s}_i)$  via

$$E^{\text{int},r}(c, \mathbf{s}_i) = E^{\text{int},r'}(c, \mathbf{s}_i) - E^{\text{int},r'}(c, \mathbf{r}_i) \quad (2.34)$$

where the reference instance  $\mathbf{r}_i$  of each site  $i$  can be chosen freely. According to Eqs. (2.32), and (2.33), and the thermodynamic cycle depicted in 2.2a, the primed intrinsic energy is given by

$$E^{\text{int},r'}(c, \mathbf{s}_i) = E^{\text{int},\text{m}}(c, \mathbf{s}_i) + \Delta E^{\text{m} \rightarrow \text{r}}(c, \mathbf{s}_i) + \Delta E_{\text{MM}}^{\text{conf},\text{m}}(c, \mathbf{s}_i) \quad (2.35)$$

In the following, the energy terms on the right-hand side this equation will be broken down further to the separate energy terms computed by GCEM.

According to the thermodynamic cycle depicted in Figure 2.2b, the relative conformational energy can be computed from

$$\Delta E_{\text{MM}}^{\text{conf},\text{m}}(c, \mathbf{s}_i) = \Delta E_{\text{MM}}^{\text{conf},\text{ff}}(c, \mathbf{s}_i) + \Delta E^{\text{ff} \rightarrow \text{m}}(c, \mathbf{s}_i) - \Delta E^{\text{ff} \rightarrow \text{m}}(c, \text{rec}) \quad (2.36)$$

where  $\Delta E_{\text{MM}}^{\text{conf},\text{ff}}(c, \mathbf{s}_i)$  is the force field energy difference between the conformer of interest and the reference conformer  $\text{rec}$  of the same binding form as instance  $\mathbf{s}_i$ . The relative force field energy applies

to the environment in which the force field was parametrized (usually vacuum). The change in the influence of the environment on the relative conformational energy upon transfer from the force field environment to the model environment is given by the last two terms on the right-hand side of Eq. (2.36). Within the CE/MM model, this change is given by the change in the energy of the site's charge distribution within its reaction field. The energy difference associated with the transfer of instance  $x$  of site  $i$  from the force field environment to the model environment is given by

$$\Delta E^{\text{ff} \rightarrow \text{m}}(c, x) = \frac{1}{2} \sum_{o=1}^{N_{c,x}^{\text{ats}}} q_o^{\text{site}} [\phi^{\text{m}}(\mathbf{r}_o, q^{\text{site}}(c, x)) - \phi^{\text{ff}}(\mathbf{r}_o, q^{\text{site}}(c, x))] \quad (2.37)$$

where the counter variable  $o$  denotes a specific atomic partial charge of the site in instance  $x$  and the sum runs over all  $N_{c,x}^{\text{ats}}$  atomic partial charges of the instance. The symbols  $q_o^{\text{site}}$  and  $\mathbf{r}_o$  denote the value of the atomic partial charge and its spatial coordinate, respectively. The symbols  $\phi^{\text{m}}(\mathbf{r}_o, q^{\text{site}}(c, x))$  and  $\phi^{\text{ff}}(\mathbf{r}_o, q^{\text{site}}(c, x))$  denote the electrostatic potentials that originate from the charge distribution of the site in the model environment and the force field parametrization environment, respectively.

The remaining contributions to the intrinsic energy shift, apart from the conformational energy contribution, are collected in the energy term  $\Delta E^{\text{m} \rightarrow \text{r}}(c, \mathbf{s}_i)$  (Eq. (2.35)). This energy term corresponds to the shift of the intrinsic energy of the site  $i$  in its instance  $\mathbf{s}_i$  upon transfer from the model environment to the receptor environment. This energy contribution is given by

$$\Delta E^{\text{m} \rightarrow \text{r}}(c, \mathbf{s}_i) = \Delta E^{\text{rf}}(c, \mathbf{s}_i) + \Delta E^{\text{back}}(c, \mathbf{s}_i) + \Delta E^{\Psi}(c, \mathbf{s}_i) \quad (2.38)$$

The individual energy terms on the right-hand side of this equation are discussed in the following.

The change in the reaction field energy of the instance upon transfer from the model environment to the receptor environment,  $\Delta E^{\text{rf}}(c, \mathbf{s}_i)$ , is given by

$$\Delta E^{\text{rf}}(c, \mathbf{s}_i) = \frac{1}{2} \sum_{o=1}^{N_{c,\mathbf{s}_i}^{\text{ats}}} q_o^{\text{site}} [\phi^{\text{r}}(\mathbf{r}_o, q^{\text{site}}(c, \mathbf{s}_i)) - \phi^{\text{m}}(\mathbf{r}_o, q^{\text{site}}(c, \mathbf{s}_i))] \quad (2.39)$$

where the counter variable  $o$  denotes a specific atomic partial charge of the site in instance  $s_i$  and the sum runs over all  $N^{\text{ats}}(c, s_i)$  atomic partial charges of the instance. The symbols  $q_o^{\text{site}}$  and  $\mathbf{r}_o$  denote the value of the atomic partial charge and its spatial coordinate, respectively. The electrostatic potentials  $\phi^{\text{r}}, \phi^{\text{m}}$  of the site's charge distribution  $q_o^{\text{site}}$  are obtained from the numerical solutions of the LPBE, stated in Eq. (2.25), at the positions  $\mathbf{r}_o$  of the respective charges. The potential  $\phi^{\text{r}}$  is obtained within the dielectric and ion accessible regions of the receptor, while  $\phi^{\text{m}}$  is obtained within the the respective boundaries of the model compound in the reference environment.

The change in the background energy of the instance upon transfer from the model environment to the receptor environment,  $\Delta E^{\text{back}}(c, s_i)$ , is given by

$$\Delta E^{\text{back}}(c, s_i) = \Delta E^{\text{back,MM}}(c, s_i) + \Delta E^{\text{back,CE}}(c, s_i) \quad (2.40)$$

where  $\Delta E^{\text{back,MM}}(c, s_i)$  is the molecular mechanics part of the background energy that comprises all force field energy terms that involve atoms of the instance and atoms of the background. The continuum electrostatics part of the background energy  $\Delta E^{\text{back,CE}}(c, s_i)$  describes the electrostatic interaction energy between the charge distributions of the instance and the background. This energy term is given by

$$\begin{aligned} \Delta E^{\text{back,CE}}(c, s_i) = & \sum_{o=1}^{N^{\text{ats,r}}(\text{back})} q_o^{\text{back}} \phi^{\text{r}}(\mathbf{r}_o, q^{\text{site}}(c, s_i)) \\ & - \sum_{p=1}^{N^{\text{ats,m}}(\text{back})} q_p^{\text{back}} \phi^{\text{m}}(\mathbf{r}_p, q^{\text{site}}(c, s_i)) \end{aligned} \quad (2.41)$$

where the counter variable  $o, p$  denotes a specific atomic partial charge of the background in the receptor and model environment, respectively. The sums run over all  $N^{\text{ats,r}}(\text{back}), N^{\text{ats,m}}(\text{back})$  atomic partial charges of the background in the receptor and model environment, respectively. The background in the model environment can, e.g., correspond to blocking groups added to the backbone of an aminoacid. The symbols  $q_x^{\text{back}}$  and  $\mathbf{r}_x$  denote the value of an atomic partial charge and its spatial coordinate, respectively. The electrostatic potentials

$\phi^r(\mathbf{r}_o, q^{\text{site}}(c, \mathbf{s}_i))$  and  $\phi^m(\mathbf{r}_p, q^{\text{site}}(c, \mathbf{s}_i))$  are obtained as described above for  $\Delta E^{\text{rf}}(c, \mathbf{s}_i)$  at the positions  $\mathbf{r}_o, \mathbf{r}_p$  of the respective charges.

The symbol  $\Delta E^\Psi(c, \mathbf{s}_i)$  denotes the potential energy of the instance's charge distribution in the electrostatic membrane potential gained upon transfer from the model environment to the receptor environment. This energy term is given by

$$\Delta E^\Psi(c, \mathbf{s}_i) = \Gamma^{\text{site}}(c, \mathbf{s}_i) \Psi \quad (2.42)$$

where  $\Gamma_{c,i,k}^{\text{site}}$  is the energy of the instance's charge distribution within the membrane potential normalized to a membrane potential of 1 V. The symbol  $\Psi$  denotes the magnitude of the electric transmembrane potential. The energy term  $\Delta E^\Psi(c, \mathbf{s}_i)$  accounts for the electrostatic interaction of the fixed charges of the instance with the mobile source charge distribution of the electrostatic membrane potential in the two solvent phases that are separated by the membrane. The normalized energy term is given by

$$\Gamma^{\text{site}}(c, \mathbf{s}_i) = \sum_{o=1}^{N^{\text{ats}}(c, \mathbf{s}_i)} q_o \chi_o \quad (2.43)$$

where  $\chi_o$  is the fraction of the membrane potential felt by charge  $q_o$  and the sum runs over all  $N^{\text{ats}}(c, \mathbf{s}_i)$  atomic partial charges of the instance. The electric distance function  $\chi_o = \psi(\mathbf{r}_o, q^\Psi) \Psi^{-1}$  is the fraction of the membrane potential felt by charge  $q_o$  where  $q^\Psi$  is the distribution of mobile ions from which the membrane potential originates. The calculation of the electric transmembrane potential is described above. The electric transmembrane potential distribution, and its source charge distribution are visualized for the transmembrane transport protein AfAmt-1 in Manuscript D and the corresponding Supporting Information.

### The Interaction Energy Between Two Sites

The interaction energy between two sites, as computed by GCEM, consists of an continuum electrostatics part and a molecular mechanics part

$$W(c, \mathbf{s}_i, \mathbf{s}_j) = W^{\text{CE}}(c, \mathbf{s}_i, \mathbf{s}_j) + W^{\text{LJ}}(c, \mathbf{s}_i, \mathbf{s}_j) \quad (2.44)$$

where the molecular mechanics part  $W^{\text{LJ}}(c, \mathbf{s}_i, \mathbf{s}_j)$  is computed from the standard Lennard-Jones interaction term of the CHARMM force field summed up for all pairs of atoms from the instances  $\mathbf{s}_i$  and  $\mathbf{s}_j$  of both sites. The continuum electrostatics contribution to the interaction energy is given by

$$W^{\text{CE}}(c, \mathbf{s}_i, \mathbf{s}_j) = \sum_{o=1}^{N^{\text{ats}}(c, \mathbf{s}_j)} q_o^{\text{site}} \phi^{\text{r}}(\mathbf{r}_o, q^{\text{site}}(c, \mathbf{s}_i)) \quad (2.45)$$

denotes the electrostatic energy of the charge distribution of instance  $\mathbf{s}_j$  of site  $j$  in the electrostatic potential that originates from the charge distribution of site  $i$  in instance  $\mathbf{s}_i$ . The sum runs over all  $N^{\text{ats}}$  charges  $q_o$  of instance  $\mathbf{s}_j$  of site  $j$ . The electrostatic potential  $\phi^{\text{r}}(\mathbf{r}_o, q^{\text{site}}(c, \mathbf{s}_i))$  is computed at the respective coordinates  $r_o$  of the charges  $q_o$  as described above for  $\Delta E^{\text{rf}}(c, \mathbf{s}_i)$ .

### The Global Conformational Energy of the Background

The global conformational energy of the background can be partitioned into two membrane potential dependent terms  $s$  and a membrane potential independent term  $\Delta E_{\text{MM}}^{\text{gc}, \text{r}}(c)$

$$E_c^{\text{conf}} = \Delta E_{\text{MM}}^{\text{gc}, \text{r}}(c) + \Delta E^{\text{gc}, \Psi}(c) + \Delta E^{\text{gc}, \text{C}}(c) \quad (2.46)$$

The individual terms of this equation are discussed in the following.

The electrostatic energy of the background charge distribution in conformation  $c$  within the electric transmembrane potential is given by

$$\Delta E^{\text{gc}, \Psi}(c) = \Gamma_c^{\text{back}} \Psi \quad (2.47)$$

where  $\Gamma_c^{\text{back}}$  is the interaction energy of the background charge distribution with the membrane potential normalized to a membrane potential of 1V and  $\Psi$  is the magnitude of the electric transmembrane potential. This energy term accounts for the electrostatic interaction of the fixed background charges with the source charge distribution of the electrostatic membrane potential

$$\Gamma_c^{\text{back}} = \sum_{o=1}^{N^{\text{ats}, \text{gc}}(c)} q_o \chi_o \quad (2.48)$$

where  $\chi_o = \psi(\mathbf{r}_o, q^\Psi) \Psi^{-1}$  is the fraction of the membrane potential felt by charge  $q_o$  and the sum runs over all charges of the background. The spatial distribution of the electrostatic transmembrane potential  $\psi$  is computed as described above.

The last energy contribution which depends on the global conformation is the capacitive energy

$$\Delta E^{\text{gc,C}}(c) = \frac{1}{2} C \Psi^2 \quad (2.49)$$

where  $C$  is the capacitance of the membrane-protein system given by

$$C = \frac{Q^\Psi}{\Psi} \quad (2.50)$$

Here,  $Q^\Psi$  is the net charge on the positively charged membrane side and  $-Q^\Psi$  is the net charge on the negatively charged membrane side. This energy contribution accounts for the energy expense spent in charging the system, in absence of the receptor charges, by applying the an electric membrane potential with the magnitude  $\Psi$ .<sup>31</sup> The membrane protein system is equivalent to an electric capacitor, i.e., positive and negative charges are separated by voltage driven translocation across the membrane. The fraction of the unbalanced charge  $q^\Psi(\mathbf{r})$  causing the membrane potential at a certain ion accessible spatial coordinate can be expressed as the sum of the effective charge, defined by Eq. (2.29), and a charge of opposite sign that results from the redistribution of mobile ions in response to the membrane potential.

$$q^\Psi(\mathbf{r}) = q^{\text{eff}}(\mathbf{r}) + q^{\text{ind}}(\mathbf{r}) \quad (2.51)$$

where  $q^{\text{eff}}(\mathbf{r})$ , given by Eq. (2.29), is the effective charge used in the calculation of the membrane potential distribution. The charge  $q^{\text{ind}}(\mathbf{r})$  induced by the membrane potential has the opposite sign of the effective charge and cancels it exactly in the bulk inner and outer phases (far away from the membrane) where the membrane potential reaches  $\psi(\mathbf{r}) = \pm 0.5 \Psi$ . The induced charge at a certain spatial coordinate is given by

$$q^{\text{ind}}(\mathbf{r}) = -\frac{\bar{\kappa}^2 h^3 \psi(\mathbf{r})}{4\pi} \quad (2.52)$$

where  $\psi(\mathbf{r})$  is the electrostatic membrane potential at the spatial coordinate which is calculated as described above. Most of the charge from which the membrane potential originates is concentrated close to the ion-accessible surface of the membrane and the membrane intrinsic receptor. The average net charge on each membrane side is calculated by summing up  $q^\Psi(\mathbf{r})$  over all ion accessible grid points on the respective membrane side. Besides the distribution of dielectric regions and electrolyte, the capacitance and hence the capacitive energy depend on the size of the membrane protein system (primarily on its area in the membrane plane). Hence the capacitance of different global conformations  $c$  has to be computed for the same grid size to be comparable. The net unbalanced charge on either membrane side is very small for physiological membrane potentials on a molecular scale and is not very sensitive to the conformation of an integral membrane protein.<sup>31</sup> Typically values of  $Q^\Psi \approx 0.4e_0$  are reached for physiological membrane potentials of -100 mV, an ionic strength of 0.15 mol/l and typical system sizes for integral membrane proteins of approximately  $(100 \text{ \AA})^3$ . Consequently, the relative capacitive energy of different global conformations is negligibly small under physiological conditions.<sup>31</sup>

Currently, GCEM computes only the membrane potential dependent contributions to the global conformational energy. The remaining part of the global conformational energy, denoted by  $\Delta E_{MM}^{\text{gc,r}}(c)$ , has to be provided by the user because there are different approaches to calculate this quantity. A choice between the different possibilities was not yet required, since the applications presented in this thesis did not require the use of multiple global conformations.

The global conformational energy can in principle be computed in the same way as the relative conformational energy of the sites (Figure 2.2b). The difference in the calculation of the global conformational energy is that one would transfer the background from the force field environment directly to the receptor environment. With this approach one might need a large number of global conformations to represent the conformational ensemble properly and to obtain the cor-

rect equilibrium distribution. A large number of global conformations would impair the efficiency of the method because the whole set of continuum electrostatics calculations has to be repeated for each global conformation. In addition, also the subsequent MC simulations of the resulting model would become more computationally demanding with a growing number of global conformations. However, Kieseritzky and Knapp reported good experiences in practical calculations with a similar approach using a limited number of global conformations.<sup>21</sup>

A more empirical approach uses global conformational energies that are fitted to experimentally observed populations of the conformers.<sup>14,45</sup> The major disadvantage of this approach is that it is only applicable if experimentally determined populations of the global conformations and representative structures for these conformations are available.

In a different approach, one could use relative free energies of conformational substates identified with a Markov model built from an equilibrium molecular dynamics simulation of the receptor.<sup>46,47</sup> Structures of the global conformations which are representative of each conformational substate and relative conformational energies could be extracted from such a simulation. There are two drawbacks of this approach. The first drawback is the need to run computationally relatively expensive molecular dynamics simulations. The second drawback is the necessity to remove the influence of the sites on the relative conformational energies to avoid a double counting of these energy contributions which would impair the internal consistency of the microstate energy function.

## Bibliography

- [1] Ullmann, G. M.; Knapp, E. W. *Eur. Biophys. J.* **1999**, *28*, 533–551. doi.
- [2] Ullmann, G. M. *J. Phys. Chem. B* **2000**, *104*, 6293–6301. doi.
- [3] Calimet, N.; Ullmann, G. M. *J. Mol. Biol.* **2004**, *339*, 571–589. doi.
- [4] Bombarda, E.; Becker, T.; Ullmann, G. M. *J. Am. Chem. Soc.* **2006**, *128*, 12129–12139. doi.
- [5] Becker, T.; Ullmann, R. T.; Ullmann, G. M. *J. Phys. Chem. B* **2007**, *111*, 2957–2968. doi.
- [6] Essigke, T. , “A continuum electrostatic approach for calculating the binding energetics of multiple ligands”, Ph.D. thesis, University of Bayreuth, 2008. url.
- [7] Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219–10225. doi.
- [8] Beroza, P.; Fredkin, D. R.; Okamura, M. Y.; Feher, G. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5804–5808. doi.
- [9] Yang, A.-S.; Gunner, M. R.; Sampogna, R.; Honig, B. *Proteins* **1993**, *15*, 252–265. doi.
- [10] Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 9557–9561. doi.
- [11] Antosiewicz, J.; Briggs, J. M.; Elcock, A. H.; K., G. M.; McCammon, J. A. *J. Comput. Chem.* **1996**, *17*, 1633–1644. doi.
- [12] Beroza, P.; Fredkin, D. R. *J. Comput. Chem.* **1996**, *17*, 1229–1244. doi.
- [13] Spassov, V. Z.; Bashford, D. *J. Comput. Chem.* **1999**, *20*, 1091–1111. doi.
- [14] Ferreira, A. M.; Bashford, D. *J. Am. Chem. Soc.* **2006**, *128*, 16778–16790. doi.
- [15] Gibbs, J. W. *Trans. Conn. Acad. Arts Sci.* **1876**, *3*, 108–248.
- [16] Gibbs, J. W. *Trans. Conn. Acad. Arts Sci.* **1878**, *3*, 343–524.
- [17] Becker, R. *Theorie der Wärme*, 3rd ed.; Springer, Heidelberg, 1985.

- [18] Herschbach, D. R.; Johnston, H. S.; Rapp, D. *J. Chem. Phys.* **1959**, *31*, 1652–1661. doi.
- [19] Zhou, H.-X.; Gilson, M. K. *Chem. Rev.* **2009**, *109*, 4092–4107. doi.
- [20] Bashford, D. *Frontiers in Bioscience* **2004**, *9*, 1082–1099. doi.
- [21] Kieseritzky, G.; Knapp, E.-W. *Proteins* **2008**, *71*, 1335–1348. doi.
- [22] Piłat, Z.; Antosiewicz, J. M. *J. Phys. Chem. B* **2010**, *114*, 1393–1406. doi.
- [23] Song, Y.; Mao, J.; Gunner, M. R. *J. Comput. Chem.* **2009**, *30*, 2231–2247. doi.
- [24] Nielsen, J. E.; Vriend, G. *Proteins* **2001**, *43*, 403–412. doi.
- [25] Gilson, M. K. *Proteins* **1993**, *15*, 266–282. doi.
- [26] Schellman, J. A. *Biopolymers* **1975**, *14*, 999–1018. doi.
- [27] Wyman, J.; Gill, S. J. *Binding and Linkage: The Functional Chemistry of Biological Macromolecules*; University Science Books, Mill Valley, CA, 1990.
- [28] Gunner, M. R.; Mao, J.; Song, Y.; Kim, J. *Biochim. Biophys. Acta* **2006**, *1757*, 942–968. doi.
- [29] Antosiewicz, J. M.; Shugar, D. *Mol. BioSyst.* **2011**, *7*, 2923–2949. doi.
- [30] Ullmann, G. M. *J. Phys. Chem. B* **2003**, *107*, 1263–1271. doi.
- [31] Roux, B. *Biophys. J.* **1997**, *73*, 2980–2989. doi.
- [32] MacKerell, A. D. et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616. doi.
- [33] Grabe, M.; Lecar, H.; Jan, Y. N.; Jan, L. Y. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17640–17645. doi.
- [34] Callenberg, K. M.; Choudhary, O. P.; de Forest, G. L.; Gohara, D. W.; Baker, N. A.; Grabe, M. *PLoS ONE* **2010**, *5*, e12722. doi.
- [35] Bashford, D. , “An object-oriented programming suite for electrostatic effects in biological molecules An experience report on the MEAD project”. In *Scientific computing in object-oriented parallel environments*;

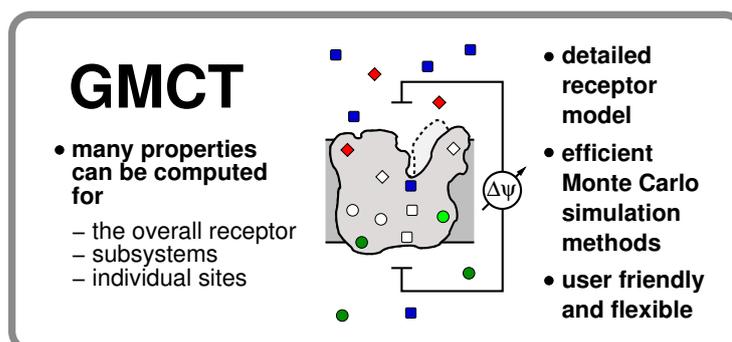
- Ishikawa, Y., Oldehoeft, R., Reynders, J., Tholburn, M., Eds.; Springer : Berlin / Heidelberg, Germany, 1997; Vol. 1343, pp 233–240. doi.
- [36] Azzone, G. et al. *Biochim. Biophys. Acta - Bioenergetics* **1993**, *1183*, 1–3. doi.
- [37] Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *186*, 671–679. doi.
- [38] Warwicker, J. *J. Theor. Biol.* **1986**, *121*, 199–210. doi.
- [39] Klapper, I.; Fine, R.; Sharp, K. A.; Honig, B. H. *Proteins* **1986**, *1*, 47–59. doi.
- [40] Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1987**, *9*, 327–335. doi.
- [41] Li, C.; Li, L.; Zhang, J.; Alexov, E. *J. Comput. Chem.* **2012**, *33*, 1960–1966. doi.
- [42] Ullmann, R. T. , “Calculation of Reorganization Energies and Electron Transfer Rates in Azurin Dimers and Complexes”, *Diplomarbeit*, Universität Bayreuth, Germany, 2005.
- [43] Warshel, A. *Biochemistry* **1981**, *20*, 3167–3177. doi.
- [44] Ullmann, G. M.; Noodleman, L.; Case, D. A. *J. Biol. Inorg. Chem.* **2002**, *7*, 632–639. doi.
- [45] Rabenstein, B.; Knapp, E. W. *Biophys. J.* **2001**, *80*, 1141–1150. doi.
- [46] Bowman, G. R.; Huang, X.; Pande, V. S. *Methods* **2009**, *49*, 197–201. doi.
- [47] Senne, M.; Trendelkamp-Schroer, B.; Mey, A. S.; Schütte, C.; Noé, F. *J. Chem. Theory Comp.* **2012**, *8*, 2223–2238. doi.



# Chapter 3

## Synopsis of the Manuscripts

### Manuscript A: GMCT : A Monte Carlo Simulation Package for Macromolecular Receptors



**Synopsis:** This manuscript presents the MC simulation package GMCT which enables the practical use of the improved methodology discussed in this thesis for the simulation of complex receptor systems. Improvements of GMCT over similarly purposed software were made along two lines. First, GMCT offers an increased level of detail in the description of the receptor-ligand system. Second, our software offers an advanced physicochemical machinery for the prediction of molecular properties based on the improved model. 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 like those encountered in bioenergetics can be studied with a variety of efficient simulation methods. GMCT offers an analytical counterpart for all of its simulation methods which can be used for validation and testing purposes at small systems that are analytically tractable. This possibility was intensely used during the programming of GMCT to ensure the correct function of the software. GMCT allows the user to utilize the full potential of continuum electrostatics/molecular mechanics models in the description of the equilibrium thermodynamics of complex receptor systems. The potential of this methodology is harnessed in Manuscripts C and D to study two biomolecular systems that feature important aspects for the understanding of bioenergetic systems. In the following, I give a more detailed summary of the novel capabilities implemented in GMCT and their application in this thesis.

A theoretical cornerstone of GMCT is the so-called microstate description of the receptor-ligand system.<sup>1-4</sup> The receptor with the bound ligands is modeled explicitly in atomic detail while the free ligands and the surroundings of the receptor are modeled implicitly. The receptor is further subdivided into a number of variable regions that are termed sites and the largely invariable remainder of the receptor that is termed background. The model comprises also the possibility that the background can possess multiple conformations termed global conformations in GMCT. The essence of the term microstate model is, that the background adopts a particular global conformation and each site adopts a particular form at any given point in the simulation. The interaction between any given pair of individual parts is thus a truly microscopic interaction energy. This is in contrast to earlier receptor models like that described by Tanford and Roxby<sup>5</sup> and its successors<sup>6-9</sup> in which each part of the receptor experiences only an average interaction with the other parts. These so-called mean-field approximations can make the calculation of receptor properties highly efficient but have a major disadvantage. Mean-field approximations are only valid for weakly interacting systems but do not provide a realistic description of strongly coupled systems.<sup>7,8</sup> Many systems

of biological interest like membrane intrinsic proteins are, however, very strongly coupled, so that mean-field approximations have a limited range of applicability. Note that a method which uses a mean-field approximation is also available in GMCT for comparison purposes and use with weakly coupled systems (see the GMCT user manual for details). It should also be noted that the microstate description can be applied in the non-equilibrium case as well. Manuscript E shows that our microstate model can be used to simulate the kinetics of electron transfer in proteins in very good agreement with experiment. Perspectives for further applications of the microstate model to the simulation of biomolecular dynamics and the associated challenges are briefly discussed in Section 1.

A second theoretical cornerstone of GMCT is the chemical potential based formalism that is derived in Chapter 2 and practically used in GMCT as described in Manuscript A. This formulation results in an energy function in which all terms have a common energy unit which facilitates understanding and interpretation of simulation results. Besides the increased transparency provided by this formalism, it has also the practical advantage that the variability of sites can easily be modeled. Sites can possess more than two forms, where each form can bind multiple ligands of different types. The traditional, widely used formalism as presented in reference 10, considered only two different ligand types (protons and electrons) and allowed only one ligand type to be bound per site. Also the inclusion of more than two forms per site was not easily possible. These restrictions seriously limited the formalism's range of applicability since many binding sites of biological interest occur in more than two binding forms and bind multiple different ligand types. Prime examples are protein cofactors like quinones or flavines that bind protons and electrons and can occur in many different binding forms, whose number is often further increased by tautomerism. Besides the modeling of cofactors, the formalism does also offer increased flexibility for the modeling of protonatable aminoacid sidechains. Traditionally, these sites were modeled with only two forms – protonated and deproto-

nated. A more detailed description was shown to be indispensable for a realistic description of short-range interactions as, for example, encountered in the simulation of proton transfer reactions.<sup>3</sup> With GMCT, one can model different binding positions of titratable protons, e.g. at the two oxygen atoms of carboxylic acids or the two nitrogen atoms of histidine. Also the rotameric flexibility of hydroxyl and sulfhydryl groups or even whole side chains can be considered. The novel formalism also naturally accounts for electrochemical transmembrane potential differences whose treatment was more complicated within the traditional formalism.<sup>11,12</sup> This capability is used in Manuscript D to investigate the influence of electrochemical transmembrane gradients on the function of the ammonium transporter Amt-1. The usefulness of the increased flexibility for a more realistic modeling of biomolecules is demonstrated in Manuscripts C and D. The molecular model of PaAz simulated in Manuscript C, comprises amino acid side chain rotamers and a detailed charge model with all possible proton positions for the protonatable amino acid side chains including hydrogen rotamers for hydroxyl and sulfhydryl groups. The model of the ammonium transporter AfAmt-1 presented in Manuscript D does also use this detailed charge model for titratable amino acid sidechains. In addition, GMCT allowed me to model the competitive binding of five different permeant species (water, hydroxyl ion, hydronium ion, ammonia and ammonium) to the transmembrane pore of the transporter including the coupled protonation equilibria connecting these permeant species. The model of AfAmt-1 does also account for the influence of the proton-motive force (electrochemical transmembrane gradient of the proton) on the thermodynamics of AfAmt-1.

GMCT uses precomputed energy terms in its microstate energy function. These energies need to be provided by the user as program input for GMCT. I wrote the software GCEM which automates the preparation of the necessary program input files. GCEM enables the user to utilize all the novel features of GMCT for a detailed modeling of biomolecular systems. The program is based on a continuum electrostatics / molecular mechanics model and makes use of a modified version of

the MEAD library.<sup>13</sup>

The increased detail of the molecular models enabled by GMCT provides a more realistic description of the simulated systems but inevitably leads to more complex and more computationally demanding models. Therefore, it was necessary to implement more sophisticated and efficient methods for the simulation of these models and for the setup and analysis of the simulations. An increased efficiency of the simulation setup and analysis was achieved by automating frequently occurring tasks. An example for such an automated task is the possibility to set up binding free energy calculations for all stepwise binding reactions of all sites or a user defined subset of the sites.

Previously, Metropolis MC<sup>14</sup> was the only basic simulation method available in simulation software for related continuum electrostatics models.<sup>15-20</sup> Metropolis MC is also available in GMCT for the computation of equilibrium observables and as basis for the newly implemented free energy calculation methods. Besides Metropolis MC, GMCT offers a second basic MC simulation method. The Wang-Landau MC method<sup>21,22</sup> can be used to compute the partition function and global thermodynamic properties very accurately, even for large systems. Manuscript C demonstrates the practical utility of Wang-Landau MC for studying the overall binding behavior of biomolecules.

Unique features of GMCT are accurate and 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, the free energy perturbation (FEP) method,<sup>23,24</sup> thermodynamic integration,<sup>25</sup> the non-equilibrium work method<sup>26</sup> and the Bennett acceptance ratio method<sup>27</sup> have been implemented. The only free energy calculation method available in previous, related software was the biased Metropolis MC method<sup>15</sup> which is also implemented in GMCT for comparison purposes. This method is, however, ineffective in many cases and not generally applicable because of numeric problems. The implementation of all free energy calculation methods in GMCT follows the principles and concepts developed in Manuscript B. In that manuscript, I also developed

the generalized FEP theory implemented in the software package. The novel FEP theory offers an improved performance and more general applicability than traditional FEP as required by the molecular models utilized in GMCT. The free energy calculation methods formed an important prerequisite of the simulations presented in in Manuscripts C and D.

Cooperativity often forms the basis of biomolecular function, for example, in bioenergetic energy transduction or signal transduction. Covariances between pairs of observables are widely used to investigate cooperative phenomena. The computation of covariances from Metropolis MC simulations is also available in GMCT. Covariances as cooperativity measure have, however, drawbacks as discussed in detail in the user manual of GMCT and the Supporting Information of Manuscript C. Cooperativity between events in molecular systems can be quantified in a thermodynamically meaningful way with cooperativity free energies. Moreover, cooperativity free energies have experimental analogues in the effective interaction energies obtained from double-mutant and triple-mutant cycle experiments. This analogy enables a direct comparisons between simulation and experiment. The usefulness of cooperativity free energies is demonstrated in Manuscript C. In that manuscript, I applied cooperativity free energies to investigate the thermodynamic coupling of protonation, reduction and conformational change in PaAz.

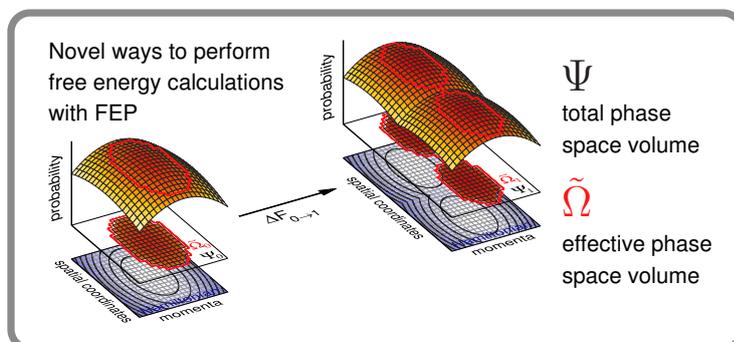
GMCT has already been used in multiple publications for the calculation of binding probabilities and to investigate the thermodynamics of enzyme function in combination with other methods.<sup>28-35</sup> The more advanced features of GMCT presented and applied in this thesis provide a wide range of opportunities for future applications in the investigation of (bio)molecular function.

**Author Contributions:** G. Matthias Ullmann wrote the initial version of GMCT and contributed ideas for improvements. I extended the software with many new features including the treatment of electrochemical transmembrane potentials, the free energy calculation capabilities, the Wang-Landau MC method, the approximate clustering

method, automatization of frequently occurring tasks, and improvements for increased performance and general user friendliness. I wrote the companion software GCEM on the basis of the MEAD library<sup>13</sup> which I extensively modified for this purpose. I tested the software, wrote the user manual and designed example calculations to provide an easy entry point for potential new users.

The current form of the formalism delineated in this manuscript and its derivation presented in Chapter 2 were contributed by me. G. Matthias Ullmann verified the formal correctness of the formalism and took part in optimizing its presentation. I conceived, designed and performed the example calculations, analyzed the results and wrote the paper. G. Matthias Ullmann commented on the manuscript and contributed ideas for the design of some figures.

## Manuscript B: A Generalized Free Energy Perturbation Theory Accounting for End States with Differing Configuration Space Volume



**Synopsis:** This manuscript presents a novel theoretical framework for free energy calculations with the free energy perturbation method. Free energy is the central quantity of thermodynamics and thus of outstanding importance in physics, chemistry and biology. Free energy calculations were used in this thesis to study the thermodynamics of binding reactions and conformational changes with special emphasis on the cooperativity between these events.

In this section, I will summarize the essential features of the novel FEP formalism and its benefits relative to previous formulations. I will limit the discussion to systems defined in terms of discrete microstates because such systems are most intuitively comprehensible and because this class of systems is actually simulated in this thesis. Herein, each microstate is defined by a given system configuration without explicitly accounting for the particle momenta. A more general discussion can be found in Manuscript B. A macrostate, often simply referred to as state, comprises one or more microstates. Examples of such macrostates are the initial and final states of a transformation (or reaction in a general sense<sup>36</sup>) and possible intermediate states connecting them. The range of microstates accessible to a system in a certain macrostate is measured by the configuration space volume of this macrostate. The configuration space volume of a macrostate in case of a discrete microstates is given by the number of microstates

comprised by the macrostate.

The purpose of FEP is to calculate the free energy difference between two macrostates usually termed initial and final state or end states of the transformation. Free energy perturbation theory uses an exponential average of energy difference samples between pairs of microstates  $o, n$  sampled from the initial and final macrostate, respectively. The microstates  $o$  are sampled from an equilibrium simulation of the initial macrostate. In GMCT, this equilibrium simulation is carried out with Metropolis MC. The free energy difference for the transformation of the initial macrostate 0 to the final macrostate 1 is then obtained as

$$\exp[-\beta\Delta G_{0\rightarrow 1}] = \frac{\mathfrak{C}}{N^{\text{sa}}} \sum_{o=1}^{N^{\text{eq}}} \sum_{n=1}^{N^{\text{diff}}} \exp[-\beta(E_n^{\text{micro}} - E_o^{\text{micro}})] \quad (3.1)$$

where  $N^{\text{sa}} = N^{\text{eq}}N^{\text{diff}}$  is the number of energy difference samples collected. Here,  $N^{\text{eq}}$  is the number of microstates  $o$  sampled from the equilibrium distribution of the initial macrostate and  $N^{\text{diff}}$  is the number of energy difference samples collected for each microstate  $o$ . To keep the notation concise, the microstates are simply numbered consecutively in order of their appearance in the simulation, where, in principle, a microstate can also occur multiple times. The constant  $\mathfrak{C}$  depends on the relative number of microstates accessible in the two macrostates and on the number of energy difference samples collected per microstate  $o$  (see Manuscript B for details). The symbols  $E_o^{\text{micro}}$  and  $E_n^{\text{micro}}$  denote the microstate energies of  $o$  and  $n$ , respectively.

The original FEP theory<sup>23</sup> and its recent extension<sup>37</sup> employ a fixed, bijective mapping in the calculation of energy difference samples between the end states. That is, each microstate  $o$  is mapped to exactly one microstate  $n$  from the adjacent macrostate. This approach has several disadvantages. First, in the case of discrete microstates, it is only possible to define such a bijective mapping if the number of microstates is the same in the initial and final macrostates 0 and 1. Second, there is no prescription on how the bijective mapping should be defined, i.e., on how to select the microstates to be paired. The exponential average in Eq. (3.1) is dominated by the energy difference

samples with the lowest values. Thus, it would be advantageous to define the mapping such that the microstates in all pairs have microstate energies that are as similar as possible. It is not difficult to define such a mapping if there is a limited number of microstates, because in this case, one can simply sort the microstates according to their energy (see example in Section 3.1 of Manuscript B). In such a case, however, the sought free energy difference is easily obtainable from an analytical calculation. In contrast, systems of practical interest mostly possess a huge number of microstates, which makes a sorting of all microstates and analytical calculations impossible. Thus, the definition of an appropriate bijective mapping could be cumbersome in the general case. The third disadvantage of using a bijective mapping between the initial and final macrostates, is that this approach does not make optimal use of the information obtainable from a simulation. The deterministic association between fixed pairs of microstates makes it impossible to collect more than one energy difference sample per microstate  $o$  sampled from the initial macrostate. Usually, the mapping will not be optimal and one will thus frequently obtain energy difference samples with high values that effectively do not contribute to the exponential average. Thus, it would be advantageous to collect multiple energy differences samples for each microstate  $o$  to increase the probability of obtaining a low energy difference sample.

The generalized FEP theory derived in Manuscript B uses a stochastic mapping instead of a fixed, deterministic mapping. That is, the selection of a microstate from the final macrostate to be associated with a given microstate sampled from the initial macrostate is based on a probability distribution. Consequently, there are multiple possible microstates  $n$  that can be associated with a given microstate  $o$ . Two such probability distributions are explored in Manuscript B. A random choice (according to a uniform distribution) of the microstate from the final macrostate is shown to be more efficient than a choice based on the equilibrium distribution of the final macrostate. The reason for the different performance of simulations based on the two mappings is discussed in detail in Manuscript B. Here, it is important to note

the advantages of using a stochastic mapping over the use of a deterministic, bijective mapping. First, one can consider transformations between macrostates with a differing number of microstates, as is often the case in practical free energy calculations. A simple example for such a case is the calculation of the protonation free energy for a carboxylic acid sidechain in a fixed conformation. While there is just one possible configuration for the deprotonated form of the site, we could place the proton on either of the two oxygen atoms of the carboxyl group for the protonated form of the site. Note, that these two proton positions are generally not equivalent in an anisotropic protein environment. Thus, there are twice as many microstates in the final macrostate than in the initial macrostate. The second advantage of using a stochastic mapping is, that the potentially cumbersome search for an appropriate bijective mapping is not required. Third, it is possible to make better use of the information obtainable from a simulation, because one can collect more than one energy difference sample for each microstate sampled from the initial macrostate. This possibility can be used to accelerate the convergence of a free energy estimate and can thus lead to more efficient simulations. The practical consequence is a possible decrease in the computation time needed to obtain an estimate of the sought free energy difference with a given statistical uncertainty. The magnitude of this decrease depends on the particular system under study but can be substantial, while the computational overhead due to the collection of multiple energy difference samples is limited. The optimal choice of the number of energy difference samples collected in each sampling step does also depend on the particular simulation in a non-trivial way.

The derivation of the generalized FEP theory also lead to interesting new insights into general statistical thermodynamics. I showed that, besides the Boltzmann-Gibbs entropy, there is a second thermodynamically meaningful entropy measure. I argued, that the Süßmann entropy<sup>38</sup> can be interpreted as measure of the extent of the thermally accessible portion of the configuration space that is actually visited by a system in equilibrium. This interpretation is in accord with ear-

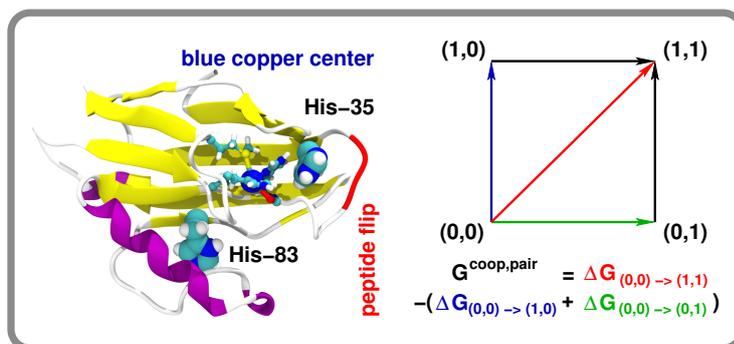
lier findings of Georg Süßmann obtained for quantum chemical systems.<sup>38,39</sup> Furthermore, I showed that the difference between the two entropy measures can be used to quantify the range of fluctuations of the microstate energy in equilibrium (see Section C of the Supporting Information of Manuscript B).

The novel FEP method was implemented in the software GMCT presented in Manuscript A and has proven to be useful in the simulations presented in Manuscripts C and D. The conceptual framework developed in this work provided the basis for the implementation of all free energy calculation methods in GMCT. I hope that I will have the opportunity to implement my FEP method also for the use with a molecular dynamics method in the future. Molecular dynamics allows to explicitly model the flexibility of ligand and receptor in binding free energy calculations. This possibility has made the method very popular in the calculation of ligand binding free energies.<sup>40-44</sup> Recent extensions of the method also allow to account for different binding forms of receptor and ligand sites.<sup>45-50</sup> The success of the method is, however, limited by the high computational burden imposed by the necessity to achieve statistical convergence of the simulations.<sup>51</sup> I expect that the application of my FEP theory can help to reduce this burden, because it enables one to use more of the information obtainable from a given simulation and offers more flexibility for the sampling of energy differences between adjacent states. In this context, it could also be interesting to develop a method to automatically determine the optimum number of energy difference samples collected in each sampling step for a given simulation. It could also be worthwhile to explore the use of alternative probability distributions for the stochastic mappings in order to optimize the efficiency of a free energy calculation.

**Author Contributions:** I conceptualized and derived the generalized free energy perturbation theory, conceived the example calculations including the derivation of the associated theory for the classical particle in the box models. G. Matthias Ullmann verified the formal correctness of the formalism and discussed its physical interpretation and significance for practical applications with me. The paper was written by

myself with support from G. Matthias Ullmann who suggested many improvements for the presentation of the formalism and the design of some figures. The publishing of novel theoretical developments is not easy, especially for a young scientist without a long publishing record like me. G. Matthias Ullmann was a constant source of encouragement and support during this process.

## Manuscript C: Coupling of Protonation, Reduction and Conformational Change in Azurin from *Pseudomonas aeruginosa* Investigated with Free Energy Measures of Cooperativity



**Synopsis:** In this work, I applied free energy measures of cooperativity to study the coupling of protonation, reduction and conformational change in azurin from *Pseudomonas aeruginosa* (PaAz). Such a coupling is prototypic for bioenergetic systems because it forms the thermodynamic basis of their energy transducing function. Proton-pumps often couple redox reactions of substrates and protein cofactors, or photochemical reactions of membrane bound chromophores to the translocation of protons across biological membranes. The mechanism of these systems often also involves conformational changes of the protein. The cooperativity free energy enables a thermodynamically meaningful quantification of coupling between two or more reactions or events.<sup>52</sup> This free energy measure of cooperativity has an intuitive interpretation as effective interaction energy between the molecules or molecule parts involved in the coupled events. Cooperativity free energy calculations are also attractive because they have experimental counterparts in double-mutant and triple-mutant cycle experiments.<sup>53,54</sup> This analogy could be exploited in the future in possible joint theoretical and experimental studies on the often still enigmatic details of the mechanism of bioenergetic systems.

PaAz is experimentally well characterized which provided opportunities to validate the results of my simulations by comparison to exper-

imental data. Therefore I chose PaAz as model system to demonstrate the usefulness of cooperativity free energies for studying the thermodynamic coupling of events in molecular systems. PaAz is a small electron transport protein with a blue copper center.<sup>55-57</sup> The exact physiological role of PaAz and the identity of its physiological electron transfer partners are not yet known.<sup>58</sup> It is well known from experiment that the reduction of the copper center is coupled to the protonation of at least two titratable residues (His-35 and His-83) and to the flip of the peptide bond between Pro-36 and Gly-37. PaAz is also interesting from a medical perspective, because it has been shown to be a potential therapeutic agent against cancer cells,<sup>59</sup> malaria pathogens and the human immunodeficiency virus.<sup>60</sup> The basis for this potential as therapeutic agent is the ability of PaAz to induce apoptosis in infected or tumorous mammalian cells by interacting with the mammalian tumor suppressor protein p53.<sup>59</sup>

In the construction of the computational model of PaAz, I made extensive use of the increased detail offered by GMCT and GCEM for the description of biomolecules (Manuscript A summarized above). I used a very detailed model with multiple hydrogen positions for all protonatable aminoacid sidechains and sidechain rotamers for flexible residues on the protein surface. The model was validated by comparison of computed  $pK_a$  values and reduction potentials to available experimental data. The computed data was in reasonable agreement with the experimental data. Here it should be noted that the available experimental  $pK_a$  values are afflicted with a high degree of uncertainty because of widely differing results obtained by different experimenters and/or from different experimental methods. The direct measurement of titration curves with nuclear magnetic resonance (NMR) spectroscopy is one of the most accurate experimental methods for obtaining protein  $pK_a$  values. Recently, however, it has been shown that the analysis of NMR measurements can be seriously complicated in case of multiple interacting protonatable sites that titrate in the same pH range.<sup>61</sup> In this case, the titration of the interacting residues can not always be distinguished reliably. The theoretical prediction of chem-

ical shifts has shown promising results.<sup>62-66</sup> In future work, it could thus be possible to attempt a direct comparison between NMR measured and simulated pH dependent chemical shift values. Such a comparison could provide a more realistic impression of the accuracy of the theoretical model and avoid misleading artifacts due to the interactions among titratable sites. The reasonable agreement of the computed with validated the further use of the constructed computational model of PaAz in the subsequent simulations.

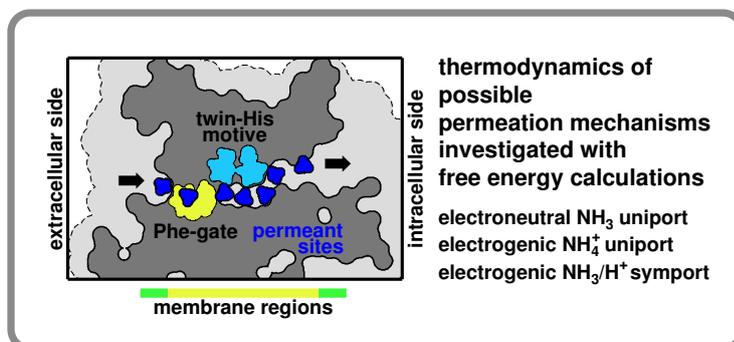
I studied the cooperativity between protonation, reduction and conformational change in PaAz in detail including the contributions of all protonatable protein sites to this phenomenon. The calculations showed the usefulness of the new features provided by GMCT as described in Manuscript A. Overall thermodynamic properties of the system were studied with the Wang-Landau MC method. The free energy perturbation theory developed in Manuscript B was used to study the system in detail from the perspective of individual sites. The calculations showed that His-35 is the protonatable site that makes the largest contribution to the cooperativity between protonation and reduction in PaAz in agreement with experimental findings. The coupling between the protonation of His-35 and the peptide flip was found to be so strong that the two happen effectively in a concerted manner. In effect, PaAz is essentially always found in one of two forms under equilibrium conditions termed high-pH form and low-pH form. I demonstrated, that the coupling between the peptide flip and the reduction of the copper center is mediated by coupling of both events to the protonation of His-35. That is, the coupling between protonation, reduction and conformational change in PaAz is based on hydrogen bonding between His-35 and the peptide flip region on the one hand and the electrostatic interaction of His-35 and the copper center on the other hand. This arrangement can be symbolized by the linear interaction scheme BCC - His-35 - PEP, where BCC stands for blue copper center and PEP for the peptide flip region and a hyphen symbolizes a strong coupling. The results also indicated that besides His-35, there are many more protonatable sites including His-83 that contribute significantly to the

cooperativity of protonation and reduction in PaAz.

An interesting effect of the transition between the high-pH form and the low-pH form is a marked alteration of the electrostatic surface potential of PaAz at the rim of a region termed hydrophobic patch. The hydrophobic patch is known to be the primary site of interaction between PaAz and other proteins in inter-protein electron transfer reactions.<sup>67-71</sup> This region has also been implicated in the interaction of PaAz with the tumor suppressor p53, which is crucial for the role of PaAz as therapeutic agent.<sup>72,73</sup> Consequently, the variability in the surface potential might affect the interaction of PaAz with its binding partners and could thus be relevant for the roles of PaAz as electron carrier and as therapeutic agent. In turn, interactions of PaAz with binding partners can affect the reduction potential of PaAz by shifting the equilibrium between its high-pH and low-pH forms. This shift is based on the above mentioned coupling between the protonation of His-35, the peptide flip and the reduction of the copper center. Finally, I speculated about the possible physiological significance of these effects. I proposed that the change of the reduction potential and the alteration of the surface potential might serve to regulate and balance the electron flow through different branches of the respiratory system of *P. aeruginosa* in response to the extent of substrate availability. The resulting redirection of the electron flow towards alternative electron sinks could contribute to the reported resistance of the bacterium against redox stress.<sup>58</sup>

**Author Contributions:** G. Matthias Ullmann initially proposed PaAz as interesting model system for my diploma thesis. I designed and performed the calculations presented herein, analyzed the results and wrote the paper. G. Matthias Ullmann commented on the manuscript and contributed ideas for the design of some figures.

## Manuscript D: Thermodynamics of Transport through the Ammonium Transporter Amt-1 Investigated with Free Energy Calculations



**Synopsis:** In this work, I applied free energy calculations to study the thermodynamics of transport through the ammonia/ammonium transporter Amt-1 from *Archaeoglobus fulgidus* (AfAmt-1). Ammonium is the most directly utilizable nitrogen source for plants and microbes.<sup>74-84</sup> AfAmt-1 and its homologues facilitate the transport of ammonia/ammonium across biological membranes in living beings from all domains of life. It is intensely debated how these proteins perform their function and whether ammonia or its protonated form ammonium is actually transported. To address these questions, I applied free energy calculations on the overall transport thermodynamics and on the thermodynamics of potential microscopic transport mechanisms.

Several aspects of this work are interesting in a broader context also, some of which I want to note here. The transport of protons, small ions and bigger molecules across biological membranes is of central importance for the function of bioenergetic systems and their role in living beings. The simulations presented in this work include the effect of electrochemical transmembrane potential differences, which is important for a realistic description of these transport processes. The formalism developed in this work for the treatment of ligands with multiple binding forms in binding free energy calculations is generally applicable. The formalism may thus also be useful outside the realm of continuum electrostatics models, for instance in binding free

energy calculations with the popular molecular dynamics methods.

For my calculations, I made extensive use of the novel software package GMCT and its companion software GCEM (Manuscript A summarized above). I used a very detailed model with multiple hydrogen positions for all protonatable aminoacid sidechains and rotamers for the permeants in a number of sites along the transmembrane pore of AfAmt-1. The parameters for the titratable aminoacid sidechains were taken from Manuscript C. Furthermore, I modeled the membrane environment of AfAmt-1 with a three-slab dielectric model to represent the hydrophobic core region and the more polar headgroup regions of the lipid membrane. The electric transmembrane potential was also included in the model. The free energy perturbation theory developed in Manuscript B was employed for the free energy calculations. The formalism developed in this work for the treatment of ligands with multiple protonation or other binding forms was subsequently developed further. This development resulted in a simple and general derivation of the binding theory used in this thesis in terms of chemical potentials which is presented in Chapter 2.

Under physiologically relevant conditions, plants and microbes need to acquire ammonium/ammonia at extracellular concentrations that are much lower than the intracellular concentrations.<sup>74,85-89</sup> Transport processes against concentration gradients are often driven by the proton-motive force (pmf) across biomembranes by coupling their transport to the transport of one or more protons. The pmf consists of a chemical component due to a proton concentration gradient across the membrane and an electric component due to a charge imbalance between the two membrane sides. The results of my calculations on the overall transport thermodynamics showed that only the electrogenic transport of ammonium is thermodynamically possible under physiological conditions. That is, because the import of the positively charged ammonium ion can be driven by the electric transmembrane potential while the import of the uncharged ammonia molecule is unaffected by the electric transmembrane potential. The import of the ammonia ion is actually even hindered under physiological conditions, because

ammonia is formed by deprotonation of ammonium on the intracellular side and reprotonated on the intracellular side. Thus, the import of ammonia is thermodynamically equivalent to an antiport of ammonium and a proton, where the proton would be transported against its concentration gradient.<sup>77</sup>

In the next stage, I studied the thermodynamics of three possible microscopic mechanism of AfAmt-1. The crystal structures of AfAmt-1 and its homologues revealed a transmembrane pore that is the most likely transport pathway for their substrates.<sup>90-95</sup> The transmembrane pore can be subdivided into a central hydrophobic pore lumen and two wider and less hydrophobic vestibule regions at the intracellular and extracellular ends of the pore.

The hydrophobicity of the pore lumen of AfAmt-1 and its homologues supported an earlier hypothesis that the transport mechanism is an ammonia uniport.<sup>84,90,96</sup> This hypothesis was based on the rationale that it would be thermodynamically prohibitive to transfer the charged ammonium ion from the polar aqueous solvent to the hydrophobic pore lumen. The simulation results indicated that the uniport of ammonia was found to be kinetically possible because of a low overall free energy barrier. A net transport in either direction does however require a driving force, which is not available for the ammonia import under physiological conditions as discussed above.

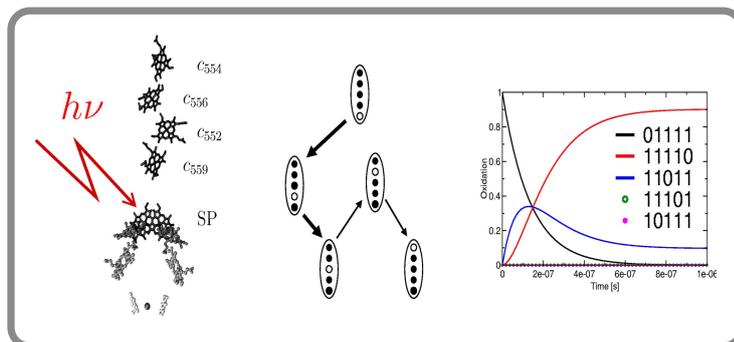
A net import of an ammonium ion can be realized microscopically as ammonium uniport or as a symport of ammonia and a proton. The symport would require that the proton and the ammonium molecule that form the ammonium ion are transported along separated routes for part of the transport pathway. Bernèche et al. proposed a plausible mechanistic scheme for the symport that involves a coplanar pair of histidine residues (twin-His motif) in the hydrophobic pore lumen as transient acceptor of the proton.<sup>78</sup> This proposal matched our ideas on the role of the twin-His motif in the conduction mechanism, which was based on an analysis of the calculated titration curves of the protein and the location of the twin-His motif in the pore lumen. The analysis of the titration curves showed that the twin-His residues

were the only titratable protein sidechains whose protonation probability showed a marked dependence on the presence or absence of ammonia/ammonium in the pore and on the electric transmembrane potential. The free energy calculations showed that it is thermodynamically not as prohibitive as initially thought to translocate the ammonium ion into the pore lumen. This is, because water molecules present in the pore lumen and the negative electrostatic potential of the protein partly compensate for the lost favorable interactions of the ammonium ion with the water molecules in the bulk solvent.

The free energy calculations showed that the ammonium uniport and the ammonia/proton symport are in principle thermodynamically viable. The overall free energy barriers of both mechanisms have similar magnitudes, which did not allow to prefer one of the mechanisms on the basis of my free energy calculations alone. The high genetic conservation of the twin-His motif argues, however, for the symport mechanism as actually operative mechanism of AfAmt-1. The high structural similarity of AfAmt-1 and its homologues makes it plausible that my findings are applicable to these proteins too. A very recently published work of Lamoureux et al. supports our findings based on simulations with a different theoretical method.<sup>97</sup>

**Author Contributions:** Susana L. A. Andrade and G. Matthias Ullmann proposed the problem. I designed and performed the calculations, analyzed the results and wrote the paper. Susana L. A. Andrade solved the X-ray crystal structure which formed the basis of my simulations, discussed the results with me from an experimental perspective and commented on the manuscript. G. Matthias Ullmann discussed the results with me, commented on the manuscript and contributed ideas for the design of some figures.

## Manuscript E: Simulation of the Electron Transfer between the Tetraheme Subunit and the Special Pair of the Photosynthetic Reaction Center Using a Microstate Description



**Synopsis:** This paper presents a first application of the microstate description to the simulation of the non-equilibrium dynamics of a molecular system. I contributed to this work in the beginning of my research for this thesis. We simulated the electron transfer kinetics underlying the re-reduction of the primary electron donor in the mechanism of the bacterial photosynthetic reaction center (bRC). The success of this work showed that a microstate description is well suited to simulate the dynamics of charge transfer reactions in biomolecular systems. Charge transfer reactions and their coupling are central elements in the function of bioenergetic systems. Thus, simulations within a microstate description have promising prospects for future simulations of these systems and may aid in the understanding of the often enigmatic details of their mechanism.

The bacterial photosynthetic reaction center is a multi-subunit integral membrane protein located in the inner membrane of *Blastochloris viridis*.<sup>98</sup> The bRC couples the oxidation of a small soluble cytochrome *c* or an iron-sulfur protein on the periplasmic side of the membrane to the reduction of a quinone at the cytoplasmic side.<sup>98</sup> The initial electron donor for this reaction is a chlorophyll dimer termed special pair (SP) on the periplasmic side of the membrane. This chromophore is excited by visible light, which initiates the electron transfer

along a chain of redox active cofactors to the quinone substrate on the intracellular side of the membrane. The oxidized SP is re-reduced by a small soluble electron carrier from the periplasmic side. The re-reduction of the SP in the *B. viridis* bRC is mediated by a permanently bound tetra-heme subunit. The heme farthest from the SPP is believed to accept the electron from the soluble periplasmic electron carrier. The four heme cofactors are arranged such that they form a linear electron transfer chain beginning at the heme farthest from the SP and ending at the heme in closest vicinity of the SP.

The values of the equilibrium reduction potentials of the four hemes are arranged along the electron transfer chain in a low-high-low-high pattern. This pattern initially caused confusion because the electron transfer cofactors along an electron transfer chain were widely believed to be always arranged such that the reduction potential is monotonically increasing. Intermediate endothermic electron transfer steps were widely regarded as impossible. This notion was supported by the failure of experiments to detect a transient reduction of the low potential hemes. The simulated re-reduction kinetics revealed an involvement of endothermic electron transfer steps. The transiently formed, thermodynamically unfavorable microstates are however quickly converted to thermodynamically more favorable microstates. The rapid depletion of the thermodynamically unfavorable microstates is the reason why these microstates are never accumulated to reach populations of more than a few percent. These low populations could explain why these experiments did not detect the transient reduction of the hemes with the low reduction potentials. The simulated kinetics showed very good agreement with experimental results which lends credibility to our results. One more lesson to be learned from our simulations is, that the microscopic kinetics underlying the experimentally observable average time evolution of the system can exhibit considerable complexity even if there are relatively few sites involved.

In this manuscript, the time evolution of the microstate populations is computed with an analytical method employing matrix algebra operations. This method has the disadvantage that its computational

tractability is limited to relatively small systems and that it does not account for stochastic effects. Stochastic effects can occur in case of low copy numbers of the simulated molecules.<sup>99,100</sup> Such low copy numbers are not uncommon in biology due to the small volumes of cells, cell compartments and even smaller volumes enclosed by substructures of cell organelles like the cristae of mitochondria (Figure 1.1). Subsequent work in our lab demonstrated the application of a dynamic MC method, which includes stochastic effects, to simulate the dynamics of charge transfer reactions.<sup>3</sup> In addition, that work showed that a microstate model is not only applicable electron transfer but also well suited to simulate the dynamics of proton transfer dynamics.

**Author Contributions:** Torsten Becker and G. Matthias Ullmann conceived and designed the calculations. G. Matthias Ullmann wrote the software for computing the electron transfer kinetics. Torsten Becker performed most of the calculations. I took part in the calculations by computing the electron transfer reorganization energies, and the molecular packing densities between the cofactors. Torsten Becker and G. Matthias Ullmann performed most of the data analysis and wrote most of the paper. I took part in the discussion of the the results and theoretical aspects of the electron transfer theory. I wrote the part of the Methods section on the calculation of electron transfer reorganization energies, helped to prepare Figure 1 for the introduction of the paper and took part in the discussion of the manuscript.

## Bibliography

- [1] Ferreira, A. M.; Bashford, D. *J. Am. Chem. Soc.* **2006**, *128*, 16778–16790. doi.
- [2] Becker, T.; Ullmann, R. T.; Ullmann, G. M. *J. Phys. Chem. B* **2007**, *111*, 2957–2968. doi.
- [3] Till, M. S.; Essigke, T.; Becker, T.; Ullmann, G. M. *J. Phys. Chem. B* **2008**, *112*, 13401–13410. doi.
- [4] Bombarda, E.; Ullmann, G. M. *Faraday Discuss.* **2011**, *148*, 173–193. doi.
- [5] Tanford, C.; Roxby, R. *Biochemistry* **1972**, *11*, 2192–2198. doi.
- [6] Gilson, M. K. *Proteins* **1993**, *15*, 266–282. doi.
- [7] Yang, A.-S.; Gunner, M. R.; Sampogna, R.; Honig, B. *Proteins* **1993**, *15*, 252–265. doi.
- [8] Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 9557–9561. doi.
- [9] Spassov, V. Z.; Bashford, D. *J. Comput. Chem.* **1999**, *20*, 1091–1111. doi.
- [10] Ullmann, G. M. *J. Phys. Chem. B* **2000**, *104*, 6293–6301. doi.
- [11] Calimet, N.; Ullmann, G. M. *J. Mol. Biol.* **2004**, *339*, 571–589. doi.
- [12] Bombarda, E.; Becker, T.; Ullmann, G. M. *J. Am. Chem. Soc.* **2006**, *128*, 12129–12139. doi.
- [13] Bashford, D. , “An object-oriented programming suite for electrostatic effects in biological molecules An experience report on the MEAD project”. In *Scientific computing in object-oriented parallel environments*; Ishikawa, Y., Oldehoeft, R., Reynders, J., Tholburn, M., Eds.; Springer : Berlin / Heidelberg, Gemany, 1997; Vol. 1343, pp 233–240. doi.
- [14] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H. *J. Chem. Phys.* **1953**, *21*, 1087–1092. doi.
- [15] Beroza, P.; Fredkin, D. R.; Okamura, M. Y.; Feher, G. *Biophys. J.* **1995**, *68*, 2233–2250. doi.

- [16] Antosiewicz, J.; Briggs, J. M.; Elcock, A. H.; K., G. M.; McCammon, J. A. *J. Comput. Chem.* **1996**, *17*, 1633–1644. doi.
- [17] Kieseritzky, G.; Knapp, E.-W. *Proteins* **2008**, *71*, 1335–1348. doi.
- [18] Song, Y.; Mao, J.; Gunner, M. R. *J. Comput. Chem.* **2009**, *30*, 2231–2247. doi.
- [19] Baptista, A. M.; Soares, C. M. *J. Phys. Chem. B* **2001**, *105*, 293–309. doi.
- [20] Couch, V.; Stuchebruckhov, A. *Proteins* **2011**, *79*, 3410–3419. doi.
- [21] Wang, F.; Landau, D. P. *Phys. Rev. Lett* **2001**, *86*, 2050–2053. doi.
- [22] Wang, F.; Landau, D. P. *Phys. Rev. E* **2001**, *64*, 056101. doi.
- [23] Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426. doi.
- [24] Ullmann, R. T.; Ullmann, G. M. *J. Phys. Chem. B* **2011**, *115*, 507–521. doi.
- [25] Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300–313. doi.
- [26] Jarzynski, C. *Phys. Rev. E* **1997**, *56*, 5018–5035. doi.
- [27] Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268. doi.
- [28] Homeyer, N.; Essigke, T.; Ullmann, G. M.; Sticht, H. *Biochemistry* **2007**, *46*, 12314–12326. doi.
- [29] Homeyer, N.; Essigke, T.; Meiselbach, H.; Ullmann, G.; Sticht, H. *J. Mol. Model.* **2007**, *13*, 431–444. doi.
- [30] Munusami, P. , “Electrostatic and quantum chemical investigation of the proton pumping mechanism of cytochrome c oxidase”, Ph.D. thesis, Universität Bayreuth, Germany, 2008. url.
- [31] Dumit, Verónica I. and Cortez, Néstor and Ullmann, G. Matthias, *J. Mol. Biol.* **2010**, *397*, 814–825. doi.
- [32] Dumit, Verónica I. and Cortez, Néstor and Ullmann, G. Matthias, *Proteins* **2011**, *79*, 2076–2085. doi.
- [33] Spiegelhauer, O.; Mende, S.; Dickert, F.; Knauer, S. H.; Ullmann, G. M.; Dobbek, H. *J. Mol. Biol.* **2010**, *398*, 66–82. doi.

- [34] Martins, B. M.; Blaser, M.; Feliks, M.; Ullmann, G. M.; Buckel, W.; Selmer, T. *J. Am. Chem. Soc.* **2011**, *133*, 14666–14674. doi.
- [35] Feliks, M.; Ullmann, G. M. *J. Phys. Chem. B* **2012**, *116*, 7076–7087. doi.
- [36] Müller, P. *Pure Appl. Chem.* **1994**, *66*, 1077–1184. doi.
- [37] Jarzynski, C. *Phys. Rev. E* **2002**, *65*, 046122. doi.
- [38] Süßmann, G. *Z. Naturforsch., A: Phys. Sci.* **1997**, *52a*, 49–52.
- [39] Schleich, W. P. In *Quantum Optics in Phase Space*; Wiley-VCH, Weinheim, 2005; Chapter Appendix C: Süßmann Measure, pp 611–613. doi.
- [40] Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492. doi.
- [41] Kollmann, P. *Chem. Rev.* **1993**, *93*, 2395–2417. doi.
- [42] Shirts, M. R.; Mobley, D. L.; Chodera, J. D. , “Alchemical free energy calculations: Ready for prime time?”. In *Annual Reports in Computational Chemistry*; Wheeler, R. A., Ed.; Elsevier B. V., Amsterdam, 2007; Vol. 3, Chapter 4, pp 41–59. doi.
- [43] Chipot, C.; Pohorille, A. *Free Energy Calculations*; Springer, Berlin and Heidelberg, 2007; Vol. 86. doi.
- [44] Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114*, 10235–10253. doi.
- [45] Bürgi, R.; Kollman, P. A.; van Gunsteren, W. F. *Proteins* **2002**, *47*, 469–480. doi.
- [46] Mongan, J.; Case, D. *Curr. Opin. Struct. Biol.* **2005**, *15*, 157–163. doi.
- [47] Chen, J.; Brooks III, C. L.; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140–148. doi.
- [48] Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmüller, H. *J. Chem. Theory Comp.* **2011**, *7*, 1962–1978. doi.
- [49] Wallace, J. A.; Shen, J. K. *J. Chem. Theory Comp.* **2011**, *7*, 2617–2629. doi.

- [50] Goh, G. B.; Knight, J. L.; Brooks, C. L. *J. Chem. Theory Comp.* **2012**, *8*, 36–46. doi.
- [51] Mobley, D. J. *Comput. Aid. Mol. Des.* **2012**, *26*, 93–95. doi.
- [52] Ben-Naim, A. *Cooperativity and Regulation in Biochemical Processes*; Kluwer Academic / Plenum Publishers, New York, 2001.
- [53] Carter, P. J.; Winter, G.; Wilkinson, A. J.; Fersht, A. R. *Cell* **1984**, *38*, 835–840. doi.
- [54] Horovitz, A. *Fold. Des.* **1996**, *1*, R121–R126. doi.
- [55] Horio, T. *J. Biochem.* **1958**, *45*, 195–205.
- [56] Horio, T. *J. Biochem.* **1958**, *45*, 267–279.
- [57] Adman, E. T. *Curr. Opin. Struct. Biol.* **1991**, *1*, 895–904. doi.
- [58] Vijgenboom, E.; Busch, J. E.; Canters, G. W. *Microbiol.* **1997**, *143*, 2853–2863. doi.
- [59] Yamada, T.; Hiraoka, Y.; Gupta, T. K. D.; Chakrabarty, A. M. *Cell Cycle* **2004**, *3*, 750–753. doi.
- [60] Chaudhari, A.; Fialho, A. M.; Ratner, D.; Gupta, P.; Hong, C. S.; Kahali, S.; Yamada, T.; Haldarf, K.; Murphy, S.; Cho, W.; Chauhan, V. S.; Gupta, T. K. D.; Chakrabarty, A. M. *Cell Cycle* **2006**, *5*, 1642–1648. doi.
- [61] Webb, H.; Tynan-Connolly, B. M.; Lee, G. M.; Farrell, D.; O'Meara, F.; Søndergaard, C. R.; Teilum, K.; Hewage, C.; McIntosh, L. P.; Nielsen, J. E. *Proteins* **2010**, *79*, 685–702. doi.
- [62] Oldfield, E. *Annual Review of Physical Chemistry* **2002**, *53*, 349–378. doi.
- [63] Mulder, F. A. A.; Filatov, M. *Chem. Soc. Rev.* **2010**, *39*, 578–590. doi.
- [64] Bühl, M.; van Mourik, T. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 634–647. doi.
- [65] Frank, A.; Müller, H. M.; Exner, T. E. *J. Chem. Theory Comp.* **2012**, *8*, 1480–1492. doi.

- [66] Frank, A.; Onila, I.; Müller, H. M.; Exner, T. E. *Proteins* **2011**, *79*, 2189–2202. doi.
- [67] van de Kamp, M.; Floris, R.; Hali, F. C.; Canters, G. W. *J. Am. Chem. Soc.* **1990**, *112*, 907–908. doi.
- [68] van de Kamp, M.; Floris, R.; Hali, F. C.; Canters, G. W. *Eur. J. Biochem.* **1990**, *194*, 109–118. doi.
- [69] van Amsterdam, I. M. C.; Ubbink, M.; Jeuken, L. J. C.; Verbeet, M. P.; Einsle, O.; Messerschmidt, A.; Canters, G. W. *Chem. Eur. J.* **2001**, *7*, 2398–2406. doi.
- [70] van Amsterdam, I. M. C.; Ubbink, M.; Jeuken, L. J. C.; Verbeet, M. P.; Einsle, O.; Messerschmidt, A.; Canters, G. W. *Nat. Struct. Biol.* **2002**, *9*, 48–52. doi.
- [71] van Amsterdam, I. M. C.; Ubbink, M.; Canters, G. W. *Inorg. Chimica Acta* **2002**, *331*, 296–392. doi.
- [72] Goto, M.; Yamada, T.; Kimbara, K.; Horner, J.; Newcomb, M.; Gupta, T. K. D.; Chakrabarty, A. M. *Molec. Microbiol.* **2003**, *47*, 549–559. doi.
- [73] Xu, C.; Yin, J.; Zhao, B. *Sci. China Life Sci.* **2010**, *53*, 1181–1188. doi.
- [74] Kleiner, D. *FEMS Microbiol. Lett.* **1985**, *32*, 87–100. doi.
- [75] Howitt, S. M.; Udvardi, M. K. *Biochim. Biophys. Acta - Biomembranes* **2000**, *1465*, 152–170. doi.
- [76] Wirén, N.; Merrick, M. , “Regulation and function of ammonium carriers in bacteria, fungi, and plants”. In *Molecular mechanisms controlling transmembrane transport*; Springer Berlin / Heidelberg, 2004; Vol. 9, pp 95–120. doi.
- [77] Andrade, S. L. A.; Einsle, O. *Mol. Membr. Biol.* **2007**, *24*, 357–365. doi.
- [78] Lamoureux, G.; Javelle, A.; Baday, S.; Wang, S.; Bernèche, S. *Transfus. Clin. Biol.* **2010**, *17*, 168–175. doi.
- [79] Khademi, S.; Stroud, R. M. *Physiol.* **2006**, *21*, 419–429. doi.

- [80] Khademi, S.; Stroud, R. M. , “Gas channels for ammonia”. In *Structural biology of membrane proteins*; The Royal Society of Chemistry, 2006; pp 212–234. doi.
- [81] Winkler, F. *Pflugers Arch. Eur. J. Phy.* **2006**, *451*, 701–707. doi.
- [82] Ludewig, U.; Neuhäuser, B.; Dynowski, M. *FEBS Lett.* **2007**, *581*, 2301–2308. doi.
- [83] Ludewig, U.; von Wirén, N.; Rentsch, D.; Frommer, W. *Genome Biol.* **2001**, *2*, reviews1010.1–reviews1010.5. doi.
- [84] Boron, W. F. *Exp. Physiol.* **2010**, *95*, 1107–1130. doi.
- [85] Wang, M. Y.; Glass, A. D. M.; Shaff, J. E.; Kochian, L. V. *Plant Physiol.* **1994**, *104*, 899–906. doi.
- [86] Britto, D. T.; Glass, A. D.; Kronzucker, H. J.; Siddiqi, M. Y. *Plant Physiol.* **2001**, *125*, 523–526. doi.
- [87] Schmidt, I.; Look, C.; Bock, E.; Jetten, M. S. M. *Microbiology* **2004**, *150*, 1405–1412. doi.
- [88] Søgaard, R.; Alsterfjord, M.; MacAulay, N.; Zeuthen, T. *Pflugers Arch. Eur. J. Phy.* **2009**, *458*, 733–743. doi.
- [89] Boogerd, F. C.; Ma, H.; Bruggeman, F. J.; van Heeswijk, W. C.; García-Contreras, R.; Molenaar, D.; Krab, K.; Westerhoff, H. V. *FEBS Lett.* **2011**, *585*, 23–28. doi.
- [90] Khademi, S.; O’Connell, J.; Remis, J.; Robles-Colmenares, Y.; Miercke, L. J. W.; Stroud, R. M. *Science* **2004**, *305*, 1587–1594. doi.
- [91] Zheng, L.; Kostrewa, D.; Bernèche, S.; Winkler, F. K.; Li, X.-D. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17090–17095. doi.
- [92] Andrade, S. L. A.; Dickmanns, A.; Ficner, R.; Einsle, O. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 14994–14999. doi.
- [93] Li, X.; Jayachandran, S.; Nguyen, H.-H. T.; Chan, M. K. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19279–19284. doi.

- [94] Lupo, D.; Li, X.-D.; Durand, A.; Tomizaki, T.; Cherif-Zahar, B.; Matassi, G.; Merrick, M.; Winkler, F. K. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19303–19308. doi.
- [95] Gruswitz, F.; Chaudhary, S.; Ho, J. D.; Schlessinger, A.; Pezeshki, B.; Ho, C.-M.; Sali, A.; Westhoff, C. M.; Stroud, R. M. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 9638–9643. doi.
- [96] Soupene, E.; He, L.; Yan, D.; Kustu, S. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 7030–7034.
- [97] Wang, S.; Orabi, E. A.; Baday, S.; Bernèche, S.; Lamoureux, G. *J. Am. Chem. Soc.* **2012**, *134*, 10419–10427. doi.
- [98] Deisenhofer, J.; Michel, H. *EMBO J.* **1989**, *8*, 2149–2170.
- [99] Gillespie, D. T. *J. Phys. Chem. B* **2009**, *113*, 1640–1644. doi.
- [100] van Kampen, N. G. *Stochastic Processes in Physics and Chemistry*, 3rd ed.; North-Holland Personal Library (Elsevier, Amsterdam), 2007; doi.



# Chapter 4

## Manuscript A:

GMCT : A Monte Carlo Simulation

Package for Macromolecular Receptors

R. Thomas Ullmann and G. Matthias Ullmann

J. Comput. Chem. 2012, 33, 887-900

doi: 10.1002/jcc.22919

# GMCT : A Monte Carlo Simulation Package for Macromolecular Receptors

R. Thomas Ullmann<sup>[a]\*</sup> and G. Matthias Ullmann<sup>[a]\*</sup>

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](http://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.<sup>[1–16]</sup> 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.<sup>[17,18]</sup> 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.<sup>[19–23]</sup> There are several programs for Metropolis Monte Carlo Simulations based on similar physical models: WhatIf<sup>[6,24]</sup> DOPS<sup>[8,22,25]</sup> MCCE<sup>[26,27]</sup> Karlsberg<sup>[28]</sup> MCTI<sup>[4]</sup> Petite<sup>[29]</sup> and Monte.<sup>[30]</sup> 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,<sup>[31,32]</sup> thermodynamic integration,<sup>[33]</sup> the non-equilibrium work method<sup>[34]</sup> and the Bennett acceptance ratio method<sup>[35]</sup> have been implemented. Further features that have not been previously available are the possibility to account for electrochemical potential gradients<sup>[36,37]</sup> 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.

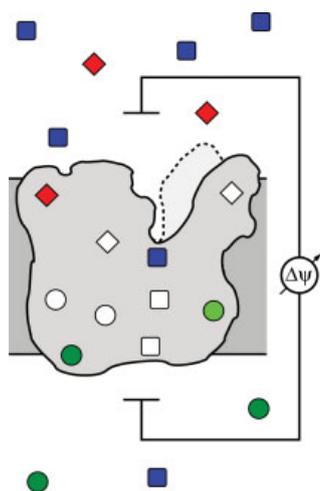
GMCT is open source software written in C/C++. GMCT was originally developed for Linux systems but should be compatible on any system with a standard C/C++ compiler. The programs are frugal in terms of hardware requirements but can of course profit from modern hardware. GMCT was optimized in many parts 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,  
BG, Bayreuth 95447, Germany  
E-mail: [thomas.ullmann@uni-bayreuth.de](mailto:thomas.ullmann@uni-bayreuth.de);  
[matthias.ullmann@unibayreuth.de](mailto:matthias.ullmann@unibayreuth.de)

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

© 2012 Wiley Periodicals, Inc.



**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](http://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 emphasis 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<sup>[12,13,15,16]</sup> and other groups.<sup>[1–11,14]</sup> 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,<sup>[14,15,23,38]</sup> while the surrounding ligand solution is described by electrochemical potentials in a mean field approach.<sup>[33,39]</sup> 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.<sup>[40]</sup> 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  $N^{\text{confs}}$ . 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

$$\mathbf{q} = \{c, \vec{s}\} \quad (1)$$

Here,  $c$  is the global conformation currently adopted by the background and the  $N^{\text{sites}}$  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.

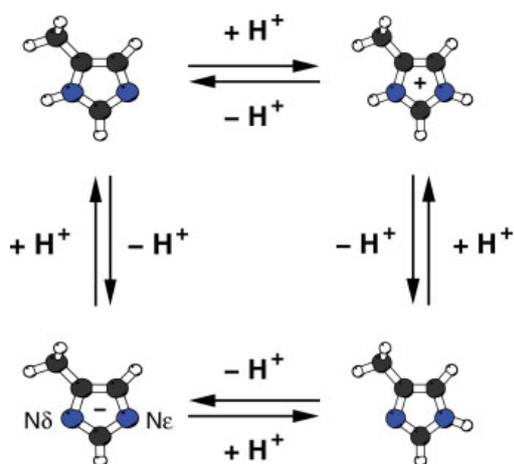
The partitioning of the protein into background and individual sites and the representation of the sites by a set of instances must be made such as to comply with the assumption of pairwise interactions. That is, the partitioning of the receptor must be chosen such that any change in a receptor constituent 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.<sup>[2,6–9,12,23,24,27–29,41]</sup> 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  $\delta$  nitrogen atom and one proton can be bound at the  $\epsilon$  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](http://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<sup>[5]</sup>

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

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

$$\Delta E_i^{\text{int}} = E_i^{\text{intr}, r} - E_i^{\text{int}, m} \quad (3)$$

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  $n$  is a function of its system configuration  $\mathbf{q}_n$ <sup>[5,11,16]</sup>

$$E_n^{\text{micro}} = E_c^{\text{conf}} + \sum_{i=1}^{N_{\text{sites}}} \left( E_i^{\text{intr}}(c, \mathbf{s}_i) - \sum_m v_m(c, \mathbf{s}_i) \bar{\mu}_m \right) + \sum_{i=1}^{N_{\text{sites}}} \sum_{j=1}^{j < i} W(c, \mathbf{s}_i, \mathbf{s}_j) \quad (4)$$

where the state vector elements  $\mathbf{s}_i, \mathbf{s}_j$  are the instances occupied by site  $i$  and site  $j$ , respectively. The first sum runs over all  $N_{\text{sites}}$  sites of the system. The second sums run over all pairs of sites in their currently occupied instances.  $W(c, \mathbf{s}_i, \mathbf{s}_j)$  is the interaction energy of instance  $\mathbf{s}_i$  of site  $i$  with instance  $\mathbf{s}_j$  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  $E_c^{\text{conf}}$  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  $\mathbf{s}_i$ . The intrinsic energy of the added site  $E_i^{\text{intr}}(c, \mathbf{s}_i)$  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  $v_m(c, \mathbf{s}_i)$ , which is specific to the global conformation of the receptor  $c$ , and to the instance  $\mathbf{s}_i$  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.<sup>[36,37]</sup> 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.

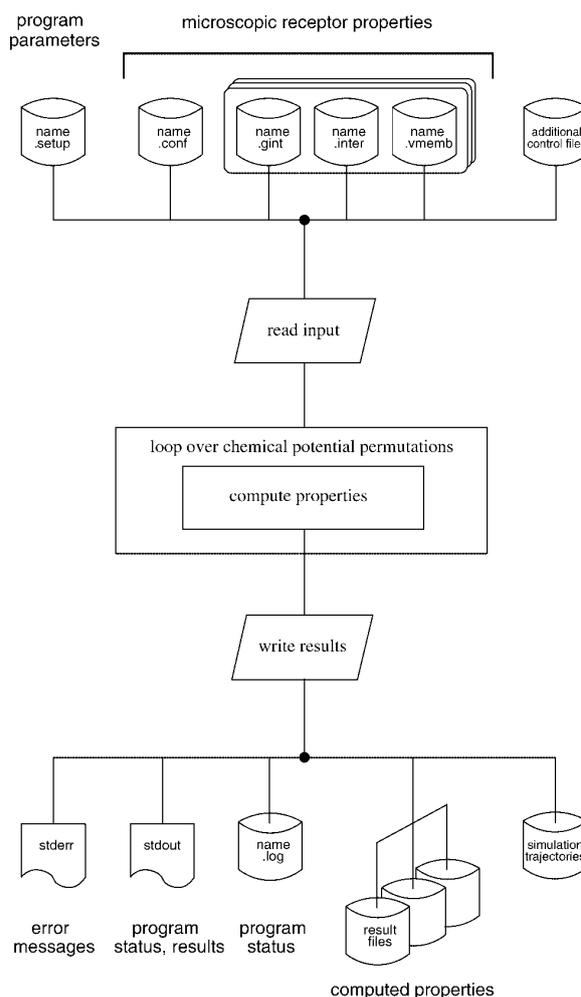


Figure 3. Overview flowchart of the programs of the GMCT suite.

The properties considered for each global conformation are the conformational energy not including the transmembrane potential dependent contributions<sup>[39]</sup> 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,<sup>[39]</sup> 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](http://www.bisb.uni-bayreuth.de). This program uses an extended version of the MEAD library<sup>[23,43]</sup> to compute the microscopic receptor properties from a continuum electrostatics/molecular mechanics model and will be subject of a separate publication.<sup>[44]</sup> 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 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.<sup>[45–48]</sup> Due to their stochastic nature, MC methods require a reliable source of random numbers.<sup>[49]</sup> The default random number generator used in GMCT is the Mersenne Twister,<sup>[50]</sup> 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.

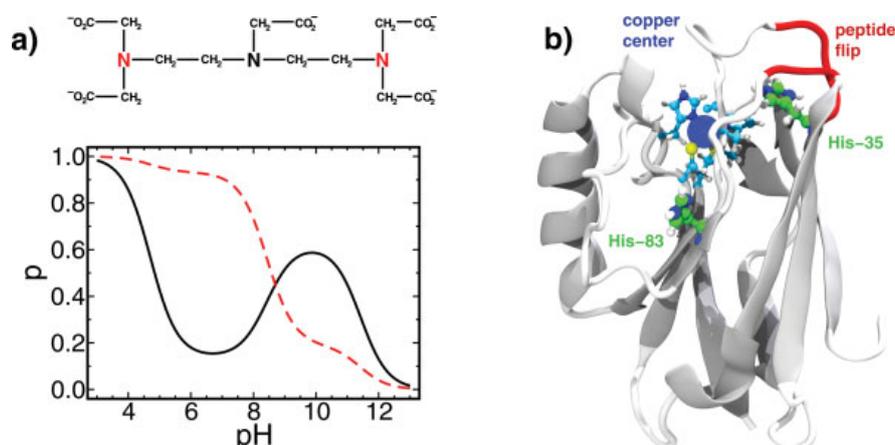
The next section introduces two example systems that will be used in the following sections to demonstrate possible applications for these methods and practical issues in the use of GMCT. Currently, GMCT features two basic MC simulation techniques that have utility in different applications. The sections “Metropolis Monte Carlo” and “Wang-Landau Monte Carlo” introduce 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-triamine-pentaacetate (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.<sup>[51]</sup> Experimental protonation probabilities in dependence on the pH value are available from nuclear magnetic resonance measurements in this range of pH values.<sup>[51,52]</sup> 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.<sup>[53]</sup> 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  $E_i^{\text{int,m}} = -\beta^{-1} \ln 10 pK_{a,i}^{\text{int}}$ . 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<sup>[55]</sup> and Tachyon.<sup>[56]</sup>

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.<sup>[38]</sup> 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.<sup>[13]</sup> 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<sup>[57]</sup> 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.<sup>[4]</sup> The method allows the treatment of such large systems as the bacterial photosynthetic reaction center, cytochrome *c* oxidase and cytochrome *bc*<sub>1</sub>.<sup>[4, 58–64]</sup>

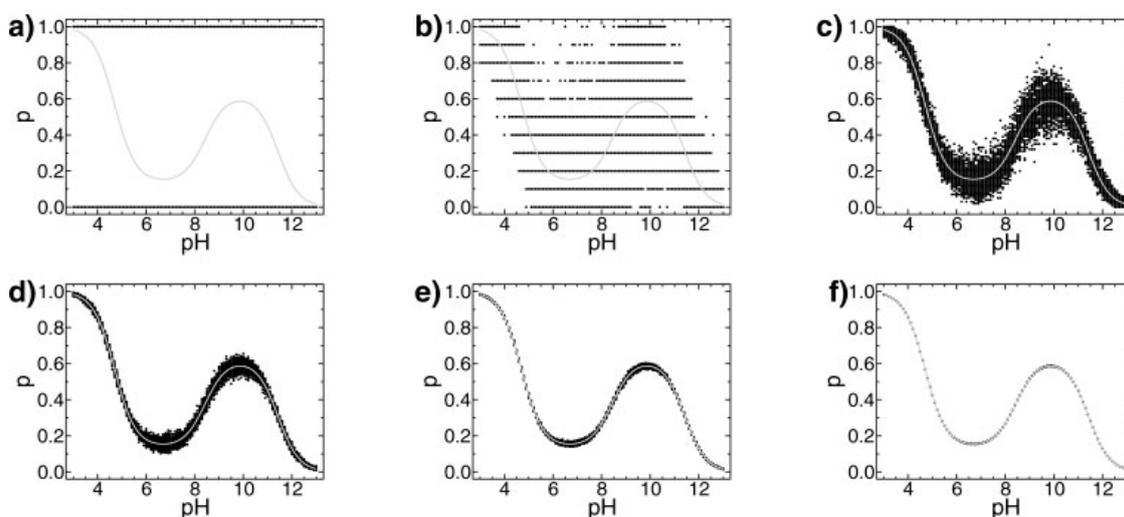
**General principle and possible applications.** A Metropolis MC<sup>[57]</sup> simulation generates a sequence of microstates starting from a randomly selected microstate. The respective next microstate of the sequence is obtained by attempting a randomly selected MC move from a set of possible moves. The MC move can involve a change in one or more variables of the configuration  $\mathbf{q}$ . 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

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

where  $N^{\text{sa}}$  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  $\beta^{-1}$ . 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,<sup>[4,47]</sup> 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.<sup>[4]</sup> 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.<sup>[19,21,65,66]</sup> GMCT was used to study the effect of serine and histidine phosphorylation on the protonation and interaction properties of the protein HPr.<sup>[67,68]</sup> In two further publications, protonation probability calculations with GMCT were employed to study the role of protonatable amino acid sidechains in the enzymatic mechanism of ferredoxin reductases.<sup>[69,70]</sup> 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<sup>[71]</sup> and 4-hydroxyphenylacetate decarboxylase.<sup>[72]</sup> Further analysis of computed titration curves and their correlation with experimentally measured titration curves can be used to identify functionally important titratable groups<sup>[38,53,74–76]</sup> or to analyze the effect of mutations on the titration behavior of a protein.<sup>[20,73,74]</sup> The analysis of computed titration curves can also be used to understand cooperative binding like the allosteric binding of oxygen and protons to hemoglobin,<sup>[77]</sup> the coupling of protonation and reduction in electron transfer proteins and oxidoreductases.<sup>[13,54,64,78,79]</sup> 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.<sup>[36,37,39]</sup> 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<sup>[80,81]</sup> 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<sup>[80,81]</sup> 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  $E_i^{\text{micro}}$  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  $i$ .

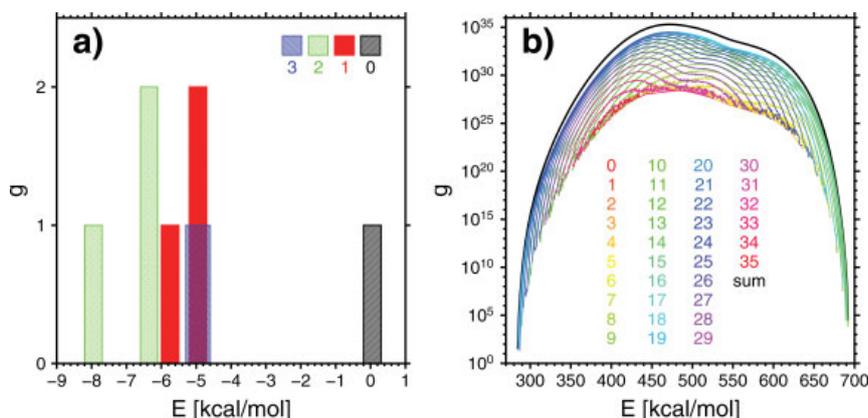
The distribution of the microstates on the microstate energy surface for the small triprotic acid DTPA is depicted in Figure 6a. With increasing system size, the number of discrete energy values can become very high. In effect,  $g$  can appear as a continuous function of the microstate energy already 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  $h_a^{\text{subst}}$  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  $\alpha$  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  $\tau^{\text{flat}}$  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  $\tau^{\text{flat}}$  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 energy for 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](http://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

$$\mathcal{Z} = \sum_i^{N_{\text{bins}}} g_i \exp[-\beta E_i] \quad (6)$$

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

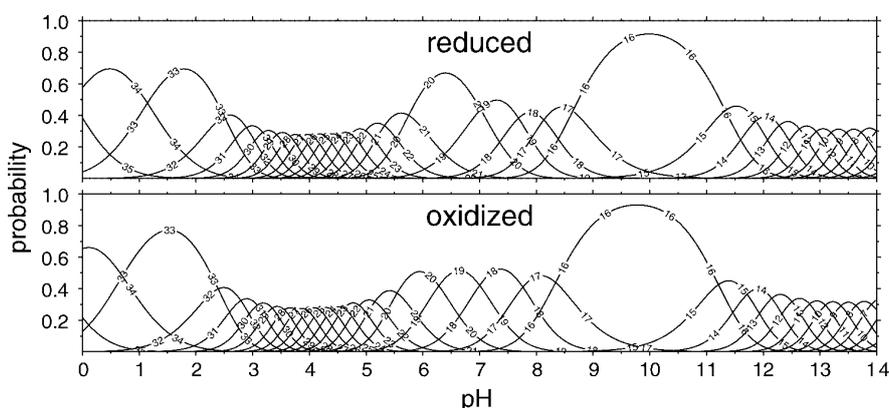
Free energy is the central quantity of thermodynamics and thus of outstanding importance in physics, chemistry, and biology. Free energy differences provide the driving force for any physical process. Free energy calculations with GMCT can, e.g., be used to study the thermodynamics of binding processes 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.<sup>[82–87]</sup> 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<sup>[58]</sup> is conceptual 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

$$\Delta G_{0 \rightarrow 1} = -\frac{1}{N^{\text{sa}}} \sum_s^{N^{\text{sa}}} B_s \quad (7)$$

computed from a number  $N^{\text{sa}}$  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  $0 \leq \lambda \leq 1$ .<sup>[87,88]</sup> These intermediate states are chimeras of the two end states, where the relative weight of the end states is determined by  $\lambda$  (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),<sup>[31,32]</sup> the non-equilibrium work (NEW)<sup>[34]</sup> method or thermodynamic integration (TI)<sup>[33,82,84]</sup> (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<sup>[33]</sup> is a very common and popular free energy calculation method.<sup>[89-91]</sup> GMCT uses the discrete version of TI instead of slow growth TI to ensure that the systems is always in equilibrium.<sup>[84]</sup> That is, the integral over the average derivative of the state energy with respect to  $\lambda$  is computed by evaluating the integrand at a number of discrete values of the coupling parameter  $\lambda$  and applying a numeric integration technique.<sup>[89,90]</sup> The integrand is evaluated at the two end states and a number of  $N^{\text{int}}$  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  $N^{\text{int}}$  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.<sup>[89,90]</sup>

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<sup>[89,92,93]</sup> and efficient free energy simulation method with a long tradition.<sup>[31,82,85-87,94]</sup> The implementation of the FEP method in GMCT is based on the initial work of Zwanzig<sup>[31]</sup> and our recent generalization of the FEP theory.<sup>[32]</sup> 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).<sup>[32]</sup> 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<sup>[34]</sup> 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<sup>[35]</sup> can be used in free energy calculations with the FEP or NEW methods to compute a maximum likelihood estimate of the free energy difference.<sup>[35,95,96]</sup> 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.<sup>[35,89,92,93,96–98]</sup>

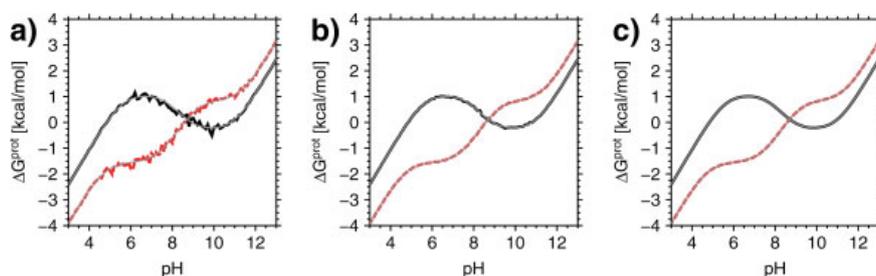
**Applications of free energy calculations.** The range of applications for free energy calculations with GMCT is broad. One can study isolated events like conformational changes, binding reactions and charge transfer reactions, as well as concerted events. Derived properties that can be calculated from the computed free energy differences include binding constants,  $pK_a$  values,<sup>[38,54,76]</sup> and reduction potentials.<sup>[13,54]</sup> 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

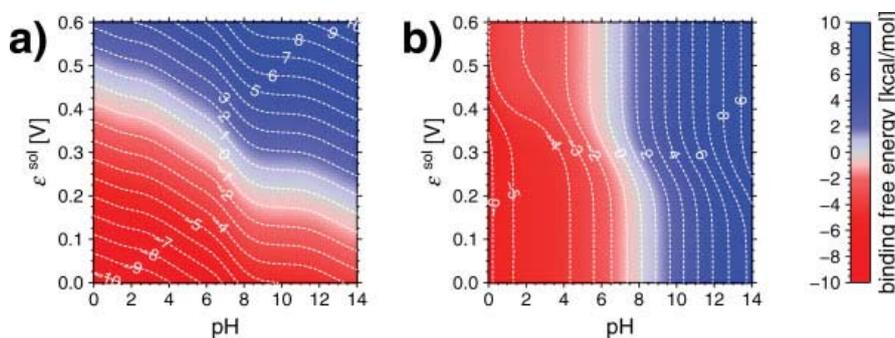
cards in the specification of conformation and site and instance names. One can also specify fixed values or value ranges for the number of ligands bound to a site. For convenience, there is the possibility to let the program automatically determine all stepwise binding reactions with the corresponding 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  $\epsilon^{\text{sol}}$ . 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.<sup>[54]</sup> 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 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<sup>[36,38]</sup>

$$C_{a,b}(\tau = 0) = p(a)p(b) - p(a,b) \quad (8)$$

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  $G^{\text{coop}(n)}$ <sup>[99]</sup> 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

$$G^{\text{coop}(n)} = G_{(0,0,\dots) \rightarrow (1,1,\dots)} - (G_{(0,0,\dots) \rightarrow (1,0,\dots)} + G_{(0,0,\dots) \rightarrow (0,1,\dots)} + \dots) \quad (9)$$

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.<sup>[100,101]</sup>

**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.<sup>[54,99]</sup> 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  $\text{p}K_{1/2}$  values can be derived from computed titration curves. Binding free energies can be expressed in terms of thermodynamically defined reduction potentials and  $\text{p}K_a$  values.<sup>[13,38,76]</sup>

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.<sup>[19,64,102]</sup>

GMCT can also be helpful in setting up and complementing molecular dynamics (MD) simulations.<sup>[67,68]</sup> 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.<sup>[103–106]</sup> 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).<sup>[19,21–23]</sup> GMCT can be obtained free of charge under the terms of the GNU Affero General Public License from [www.bisb.uni-bayreuth.de](http://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. Kligen, T. Becker, E. Bombarda, *Photosynth. Res.* **2008**, *97*, 33.
- [20] A. R. Kligen, 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. Pilat, 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. Pilat, 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. Kligen, P. Sebban, G. M. Ullmann, G. Fritzsche, *J. Mol. Biol.* **2007**, *371*, 396.
- [64] A. R. Kligen, 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



# Chapter 5

## **Manuscript B:**

A Generalized Free Energy Perturbation  
Theory Accounting for End States with  
Differing Configuration Space Volume

R. Thomas Ullmann and G. Matthias Ullmann  
J. Phys. Chem. B 2011, 115, 507-521  
doi: 10.1021/jp1093838

## A Generalized Free Energy Perturbation Theory Accounting for End States with Differing Configuration Space Volume

R. Thomas Ullmann\* and G. Matthias Ullmann

Structural Biology/Bioinformatics, University of Bayreuth, Universitätsstrasse 30, BGI,  
95447 Bayreuth, Germany

Received: September 30, 2010; Revised Manuscript Received: November 23, 2010

We present a generalized free energy perturbation theory that is inspired by Monte Carlo techniques and based on a microstate description of a transformation between two states of a physical system. It is shown that the present free energy perturbation theory stated by the Zwanzig equation follows as a special case of our theory. Our method uses a stochastic mapping of the end states that associates a given microstate from one ensemble with a microstate from the adjacent ensemble according to a probability distribution. In contrast, previous free energy perturbation methods use a static, deterministic mapping that associates fixed pairs of microstates from the two ensembles. The advantages of our approach are that end states of differing configuration space volume can be treated easily also in the case of discrete configuration spaces and that the method does not require the potentially cumbersome search for an optimal deterministic mapping. The application of our theory is illustrated by some example problems. We discuss practical applications for which our findings could be relevant and point out perspectives for further development of the free energy perturbation theory.

### 1. Introduction

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 and ultimately of life. Free energy calculations have contributed to our understanding of the reactivity and function of complex systems encountered in physics, chemistry, and biology and especially in the interesting problems emerging at their interface. These calculations supplement and guide experiment by modeling details not easily accessible to experiment. Continuous improvements have been achieved over the years both in theory and in practical application of free energy calculation methods in such diverse fields as biochemistry, nanotechnology, materials research, pharmacy, fluid phase theory, bioenergetics, and biotechnology.<sup>1–8</sup>

A problem not worked out so far occurs in free energy perturbation (FEP) when dealing with transformations between two systems for which no unequivocal one-to-one mapping of initial and final microstates exists. The purpose of the present paper is to derive a universally applicable free energy perturbation theory based on a microscopic description of the system.

**Motivating Problem.** Let us consider a simple system consisting of three microstates with equal energy  $E = 0$ . The Helmholtz free energy difference between an ensemble consisting only of the first microstate and an ensemble consisting of the remaining microstates two and three is  $-\beta^{-1} \ln 2$ , where  $\beta^{-1} = k_B T$ , as can easily be seen from elementary statistical mechanics. Analogous problems from our own work occur, for example, in the calculation of  $pK_a$  values and protonation free energies for titratable protein residues or protein cofactors.<sup>9</sup> Prototypical realizations of the example system are carboxylic acids, where the first microstate corresponds to the deprotonated state while the remaining two microstates correspond to the two

possible positions of the proton at either of the two oxygen atoms of the carboxy group in the protonated state. Such problems can in general not be solved by simple symmetry corrections, because the microstates might not be equivalent from the onset or because they become nonequivalent due to differing interactions in an anisotropic environment. Equivalent protonation forms of amino acid residues, for example, lose their equivalency in the anisotropic protein environment.

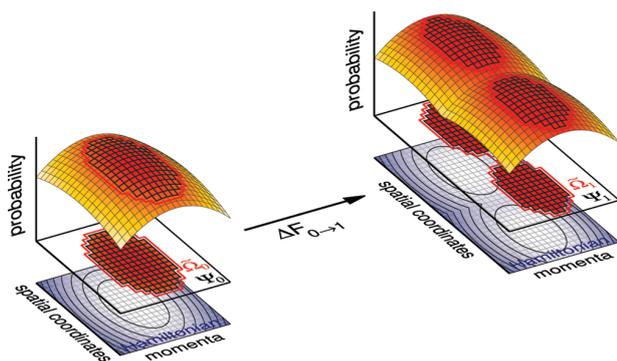
Suppose we want to calculate this free energy difference from a FEP simulation. Traditional FEP is based on the following equation derived by Zwanzig in 1954<sup>10</sup>

$$\exp(-\beta\Delta F_{0\rightarrow 1}) = \int \rho_0 \exp[-\beta(E_1(\mathbf{q}_1) - E_0(\mathbf{q}_0))] d\mathbf{q}_0$$

with  $\mathbf{q}_1 = \mathcal{M}(\mathbf{q}_0)$ ,  $\mathcal{M}: \mathbf{q}_0 \rightarrow \mathbf{q}_1$  bijective and  $|d\mathbf{q}_1| = |d\mathbf{q}_0|$  (1)

This equation tells us to sample configurations  $\mathbf{q}_0$  of the initial ensemble according to its equilibrium probability distribution  $\rho_0$  and to collect state energy difference samples  $E_1(\mathbf{q}_1) - E_0(\mathbf{q}_0)$ . Each of these energy differences describes a forced transition from a sampled configuration  $d\mathbf{q}_0$  to one from the final ensemble  $d\mathbf{q}_1$ . This transition or FEP move can be considered as perturbation of the system.<sup>11</sup> In choosing the term FEP move we seek to carry on and extend an existing interpretation of the perturbation as a special kind of Monte Carlo trial move<sup>11</sup> or step.<sup>12</sup> As we will show in the Theory section, the Zwanzig equation implicitly assumes the existence of an unequivocal one-to-one mapping,  $\mathcal{M}$ , of each configuration of the initial ensemble to exactly one configuration of the final ensemble and vice versa. More precisely, the mapping is bijective and preserves the volume of the differential volume elements mapped onto each other. Due to this one-to-one mapping, the configurations of each pair  $\mathbf{q}_0, \mathbf{q}_1$  can be perceived as one configuration  $\mathbf{q}$  subject to different energy functions  $E_0, E_1$  in the initial and the final

\* To whom correspondence should be addressed. E-mail: thomas.ullmann@uni-bayreuth.de.



**Figure 1.** Schematic view of the problem to solve. Sought is the free energy change  $\Delta F_{0 \rightarrow 1}$  for the transformation of a system from state 0 to state 1 (end states). The plots for the two end states show the occupation probability distributions and the underlying energy surfaces in phase space. The transformation describes the traveling of the system between two regions of phase space with differing extent and energy landscape. The phase space is spanned by the spatial coordinates and the momenta of the particles, depicted here exemplarily for a single particle in one spatial dimension. The contour plots at the bottom show projections of the energy surfaces into the phase space plane (constant contour interval, darker shade means higher energy). Overlaid are plots of the probability surfaces and their projections into the phase space plane. In the projections, only the significantly populated regions of phase space are shown. Each square on the surfaces represents an infinitesimal volume element in phase space. The integral over all volume elements of the phase space regardless of their occupation probability gives the *total phase space volume*  $\Psi$ . The phase space elements with a significant occupation probability are indicated by their dark shade and bold frame. These elements constitute the thermally accessible regions of phase space enclosed in the red contours. The extent of these regions in phase space is termed *effective phase space volume* and denoted by the symbol  $\Omega$ . The example for the plot is taken from the classical particle in the box models with one or two harmonic energy wells presented in the example section.

state. Within this picture, the transition between the initial and final state can be interpreted as a perturbation or switching of the energy function.<sup>10,11</sup> Accordingly, eq 1 is written in most cases without indices attached to the configuration  $\mathbf{q}$  or with the index 0 for both end states

$$\exp(-\beta\Delta F_{0 \rightarrow 1}) = \int \rho_0 \exp[-\beta(E_1(\mathbf{q}) - E_0(\mathbf{q}))] d\mathbf{q} \quad (2)$$

This perception does, however, not necessarily correspond to the real physical situation modeled in a free energy simulation. In fact, the configurations  $\mathbf{q}_0, \mathbf{q}_1$  and the systems described by them can be quite different.<sup>12–14</sup> The required one-to-one mapping is impossible to define in our example system, since the number of possible configurations differs between initial and final ensemble. An apparent solution seems to be found in a random selection of one of the two possible final configurations. It is easily seen, however, that this approach will also not lead to the correct result, because the calculated energy difference for each of the two possible FEP moves is zero. Thus, already with our simple model system we were not able to accomplish the calculation of the free energy difference in a FEP simulation with the Zwanzig equation.

**Generalized Statement of the Problem and Our Proposed Solution.** A generalized formulation of the problem which this article attempts to solve is schematically depicted in Figure 1. The aim is the calculation of the free energy difference between two states of a system corresponding to two regions of phase

space whose extent can differ. In most cases, the initial and final states of a transformation will be defined on the basis of the system configuration, while the range of the momenta is not restricted. For a classical system all momentum-dependent terms cancel in this case from free energy differences and can thus be neglected in their calculation (see Supporting Information). Consequently, the calculation can be carried out entirely in the configuration space of the system which is spanned by the coordinate axes of all particles residing in the system.

We will derive a novel formalism for FEP calculations which can be applied to arbitrary transformations in which the initial and the final states may possess a differing configuration space volume. The formalism is applicable to systems defined on discrete or continuous configuration spaces. The Zwanzig equation<sup>10</sup> will be recovered as special case of our formalism. Our formalism describes the transformation by a statistical mechanical average over energy differences for microscopic transitions between the two ensembles of microstates. Inspired by Monte Carlo methods, these transitions can be interpreted as special kind of Monte Carlo trial moves termed FEP move here.

The principal question to be answered by our derivations is, what is to be done if there are multiple possible mappings of the end states onto each other and accordingly multiple possible FEP moves? The derivations will result in three different simulation schemes. The first simulation scheme uses energy differences from all possible FEP moves and leads directly to the desired transformation free energy. The two remaining simulation schemes use single FEP moves chosen according to probability distributions. That is, these simulation schemes use a stochastic mapping of the end states onto each other instead of a static, deterministic mapping used by previous FEP methods. The single-move simulation schemes need to be augmented by a correction term that accounts for the relative configuration space volume of the end states. As an interesting side product, our theory offers insights into general statistical thermodynamics regarding the relation of configurational entropy and volume of the configuration space regions effectively explored by a system.

Recently, Jarzynski presented an alternative generalization of the Zwanzig equation named targeted FEP.<sup>14</sup> Targeted FEP uses a deterministic mapping to optimize the overlap<sup>11,15</sup> between the end states resulting in a more efficient simulation (see section 3.2 for an example). This mapping is bijective but not necessarily volume preserving. That is, the coordinates of one end state can be mapped to the coordinates of the other end state by an arbitrary coordinate transformation. Thus, targeted FEP is also applicable if the configuration space volume of the end states differs. Targeted FEP has, however, two drawbacks. First, a suitable mapping has to be found which may be cumbersome in the general case. Second and more important for our own applications, the method is not generally applicable to systems defined on a discrete configuration space. This is because a bijective mapping can only be defined if the number of discrete microstates is the same in both ensembles.

In the next section, we will formally derive our novel FEP formalism. The subsequent section illustrates the theoretical findings with some example problems. The example section is followed by a discussion of applications for which we believe our findings could be of importance and possible future prospects of the FEP theory opened by our theoretical findings. We close with a summary of the main findings of this paper and the conclusions thereof.

## 2. Theory

**2.1. Definitions.** Before we begin our derivation, it is necessary to define some terms and quantities to avoid ambiguities and to provide a formal basis for the derivation of our FEP theory. Mathematically, our physical system is described by a set of variables defining the momenta  $\mathbf{p}$  (not to be confused with the occupation probability  $p$ ) and a second set of variables  $\mathbf{q}$  defining the spatial coordinates. A given set of particular values for the coordinates  $\mathbf{q}$  defines a configuration of the system. The configuration can also include further parameters, as for example partial charges. A microstate of the system is defined by particular values for all variables describing the system, that is by a particular configuration of the system and particular momenta of all particles of the system. The microstate is fully described by a particular configuration of the system if the momenta are not considered explicitly.

The free energy difference  $\Delta F_{0 \rightarrow 1}$  refers to a transformation between two states 0 and 1. The initial and final states of a transformation are often termed end states in free energy calculations. It is important to notice that in general the initial and final states are ensembles of microstates of the system. Sometimes the term macrostate is used to indicate the ensemble nature of these states. The transformation can be partitioned into multiple stages, where initial and final state of a stage can also correspond to a fictitious (but mathematically well-defined) intermediate state constituting a chimera of the end states.<sup>16</sup> These intermediate states can be used to improve the efficiency of a free energy calculation, as briefly sketched in section B1 of the Supporting Information. The ensembles which constitute initial and final state encompass subsets of the entire phase or configuration space  $\Gamma$ , which may or may not overlap. In our case, initial and final state will be defined on the basis of the configuration. We denote sets of configurations constituting the initial ensemble and final ensemble as configuration subspaces  $\Gamma_0$  and  $\Gamma_1$ , respectively. The relation between the total configuration space  $\Gamma$  and the configuration subspaces encompassed by the initial and final states  $\Gamma_0, \Gamma_1$  can be formally defined as

$$\Gamma \supseteq \Gamma_0 \cup \Gamma_1 \quad (3)$$

indicating that the union of the initial configuration subspace and the final configuration subspace is a subset of, or equal to, the total configuration space of the system. The initial and the final configuration subspaces may or may not overlap.

When taking the integral of unity over all configurations constituting a configuration space or configuration subspace, we obtain its total configuration space volume<sup>17</sup>

$$\Psi = \int_{\Gamma} 1 \, d\mathbf{q} \quad (4)$$

The symbol  $\Psi$  is chosen to distinguish the total configuration space volume from a spatial volume. Analogously, the total phase space volume would be obtained by integrating over the momentum variables also. Here, total means that  $\Psi$  corresponds to the entire volume of configuration or phase space theoretically accessible to the system within the respective configuration space regardless of the energetics and actual probability distribution. The total phase or configuration space volume  $\Psi$  is not to be confused with the corresponding effective phase or configuration space volume actually populated by the system denoted by  $\tilde{\Omega}$ . The difference between  $\Psi$  and  $\tilde{\Omega}$  is depicted schematically in Figure 1. Note that the actual effective phase

space volume does not necessarily correspond to a region of phase space with sharp boundaries, since the occupation probability density is nonzero throughout the whole phase space on which the model is defined wherever the energy function has a finite value. A more theoretical definition of the effective phase space volume is given below. The  $DN$ -dimensional continuous phase or configuration space volume of a system of  $N$  particles in a  $D$ -dimensional spatial region in a Cartesian coordinate system has units of action  $[(J \, s)^{DN}]$ , or units of length  $[m^{DN}]$  (in SI units). See the Supporting Information for an example calculation of the phase and configuration space volumes of a system formulated in continuous, Cartesian coordinates using SI units.

Without loss of generality for the formalism, we will in the following use a description of the system based on its configuration. For a system formulated in a full phase space description, the microstate energy  $E$  becomes the Hamiltonian explicitly accounting for the momenta of the system particles and the integrations must be performed over the momenta also (see the Supporting Information for an explicit example calculation). For classical systems, however, the momentum-dependent part of the integral constituting the partition function can be separated and cancels in free energy differences (see for example refs 17 and 18 and the Supporting Information).

**2.2. Derivation of the Multimove FEP Equation.** We start our derivation with the expression of the Helmholtz free energy difference in terms of the configurational partition functions  $Q_0$  and  $Q_1$  of initial ensemble and final ensemble, respectively

$$\begin{aligned} \Delta F_{0 \rightarrow 1} &= -\beta^{-1} \ln \frac{Q_1}{Q_0} \\ &= -\beta^{-1} \ln \frac{\int_{\Gamma_1} \exp[-\beta E_1] \, d\mathbf{q}_1}{\int_{\Gamma_0} \exp[-\beta E_0] \, d\mathbf{q}_0} \end{aligned} \quad (5)$$

where the integrations are performed over all  $d\mathbf{q}_0$  and  $d\mathbf{q}_1$  configurations of the respective configuration subspaces. All the integrations in the following parts of this article are to be understood as total integrals over the whole domain on which the integration variables are defined unless otherwise stated (explicit example calculations can be found in the Supporting Information).

The microstate of the system is described by a set of variables. On the basis of these variables, we partition the system in two subsystems: the reactive subsystem and the environment subsystem. The variables with a different range of possible values in the initial configuration subspace and the final configuration subspace define the reactive subsystem, whereas the variables with the same range of possible values in both subspaces define the environment subsystem. This partitioning is, however, just a mathematical device to make the calculations more efficient. The environment subsystem can freely respond to changes in the reactive subsystem. Accordingly, we can express the total configuration  $\mathbf{q}$  of the system as combination of the configurations of these subsystems

$$\mathbf{q} = \{\mathbf{q}^e, \mathbf{q}^r\} \quad (6)$$

where the superscripts e and r indicate environment and reactive subsystem, respectively. The total configuration space volume is equal to the product of the total configuration space volumes of the environment subsystem and the reactive subsystem

$$\Psi = \Psi^e \Psi^r \quad (7)$$

The microstate energy  $E$  is a function of the configuration of the environment and of the reactive subsystem. The total state energy can thus be written as sum of terms of the energy function depending on the configuration of both subsystems  $E^r(\mathbf{q}^e, \mathbf{q}^r)$  and terms depending only on that of the environment subsystem  $E^e(\mathbf{q}^e)$

$$E_i = E^e(\mathbf{q}^e) + E_i^r(\mathbf{q}^e, \mathbf{q}^r) \quad (8)$$

where the subscript  $i = \{0,1\}$  indicates the initial or final state, respectively. The term  $E^e(\mathbf{q}^e)$  carries no subscript because it is identical for the initial and final state. In the following, the explicit statement of the dependencies in the parentheses will be omitted to keep the notation concise.

Partitioning the energy function and the configuration in the numerator of eq 5 according to this concept transforms eq 5 into

$$\Delta F_{0 \rightarrow 1} = -\beta^{-1} \ln \frac{\int \exp[-\beta E^e] \exp[-\beta E_1^r] d\mathbf{q}^e d\mathbf{q}_1^r}{\int \exp[-\beta E_0] d\mathbf{q}_0} \quad (9)$$

where the integrations are again performed over all  $d\mathbf{q}_0 = d\mathbf{q}^e d\mathbf{q}_0^r$  and  $d\mathbf{q}_1 = d\mathbf{q}^e d\mathbf{q}_1^r$  configurations of the initial and final configuration subspaces, expressed now in terms of the configurations of the environment and the reactive subsystem. The next step constitutes the major difference from the derivation of Zwanzig.<sup>10</sup> We account for multiple possible transitions of a microstate from the initial ensemble to a microstate from the final ensemble by multiplying the integrand in the numerator of eq 9 by unity expressed as

$$1 = \frac{1}{\Psi_0^r} \int \exp[-\beta(E^e - E^e)] \exp[-\beta(E_0^r - E_0^r)] d\mathbf{q}_0^r \quad (10)$$

and obtain

$$\Delta F_{0 \rightarrow 1} = -\beta^{-1} \ln \left[ \frac{\int \exp[-\beta(E_1^r - E_0^r + E^e + E_0^e)] d\mathbf{q}^e d\mathbf{q}_0^r d\mathbf{q}_1^r}{\Psi_0^r \int \exp[-\beta E_0] d\mathbf{q}_0} \right] \quad (11)$$

Before we proceed with our derivation, it seems worthwhile commenting on why the right-hand side of eq 10 is equal to 1. The product of the exponential energy terms in eq 10 is equal to 1 for each configuration  $\mathbf{q}_0^e$  of the reactive subsystem of the initial state. When taking the integral of unity over all configurations of the reactive subsystem, we obtain the total configuration space volume  $\Psi_0^r = \int 1 d\mathbf{q}_0^r$  theoretically accessible to the reactive subsystem within the configuration subspace of the initial state. To obtain unity on the left-hand side of eq 10, we thus have to divide the right-hand side by  $\Psi_0^r$ . In the case of a discrete configuration space, this volume becomes simply the total number of possible configurations of the reactive subsystem within the initial configuration subspace.

In eq 11 the canonical occupation probability density of a configuration in the initial configuration subspace given by

$$\rho_0 = \frac{\exp[-\beta(E^e + E_0^r)]}{\int \exp[-\beta E_0] d\mathbf{q}_0} \quad (12)$$

can be recognized. We thus obtain from eq 11

$$\Delta F_{0 \rightarrow 1} = -\beta^{-1} \ln \left[ \frac{1}{\Psi_0^r} \int \rho_0 \exp[-\beta(E_1^r - E_0^r)] d\mathbf{q}^e d\mathbf{q}_0^r d\mathbf{q}_1^r \right] \quad (13)$$

Equation 13 shows the benefit of partitioning the configuration into reactive and environment subsystem for a simulation. During the calculation of the energy difference, only energy terms depending on the configuration of the reactive subsystem need to be computed while terms depending solely on the configuration of the environment subsystem cancel from the difference. By symmetry, the corresponding equation for the reverse transformation is simply obtained by exchanging source and target indices. Equation 13 can be written in a more compact form as

$$\exp(-\beta \Delta F_{0 \rightarrow 1}) = \frac{1}{\Psi_0^r} \langle \int \exp[-\beta(E_1^r - E_0^r)] d\mathbf{q}_1^r \rangle_0 \quad (14)$$

where  $\langle \dots \rangle_0$  denotes the ensemble average over the initial state.

A special case occurs if the reactive subsystem has just one possible configuration in each end state. In this case, the variables constituting  $\mathbf{q}_0^r$  and  $\mathbf{q}_1^r$  have only one possible value specific to either end state; i.e., they become constants. The dimensionality of the configuration subspaces is then reduced relative to that of the entire configuration space, because the constants  $\mathbf{q}^r$  are omitted from the integration in eq 4. Consequently, the total configuration space volume of the reactive subsystem  $\Psi^r$  is unity for both end states. The total volumes of initial and final configuration subspaces  $\Psi_0 = \Psi^e \Psi_0^r$  and  $\Psi_1 = \Psi^e \Psi_1^r$  are equal in this special case. Furthermore, the mapping of initial to final configurations  $\mathcal{M}: \mathbf{q}_0 \rightarrow \mathbf{q}_1$  is, in this special case, unequivocally determined by changing the value of  $\mathbf{q}^r$  to the only possible value in the adjacent end state while keeping the configuration of the environment subsystem  $\mathbf{q}^e$

$$\mathcal{M}: \mathbf{q}^e \mathbf{q}_0^r \rightarrow \mathbf{q}^e \mathbf{q}_1^r |_{\mathbf{q}_0^e, \mathbf{q}_1^e \text{ const}} \quad (15)$$

This mapping is bijective and volume preserving; i.e., there is a one-to-one correspondence of each differential volume element  $d\mathbf{q}_0$  to exactly one differential volume element  $d\mathbf{q}_1$ . The size of the differential volume elements  $d\mathbf{q}_0$  and  $d\mathbf{q}_1$  and the total volume of the configuration subspaces mapped onto each other is equal. The derived one-to-one mapping can be perceived as defining a unique FEP move which is defined by changing the configuration of the reactive subsystem  $\mathbf{q}^r$  from the only set of values allowed in the initial ensemble to the respective only set of values allowed in the final ensemble. This FEP move can be described as a switching of the energy function,<sup>11</sup> where the microstate energies  $E_i$  specific to the end states  $i = \{0,1\}$  would be conceived as two different energy functions whose range of validity is defined by the configuration of the reactive

subsystem (see setup A of Bennett's model system in section 3.2 for an example). Consequently, the familiar FEP equation as given in ref 10 follows as a special case of eq 13 in the case of just one possible configuration of the reactive subsystem

$$\exp(-\beta\Delta F_{0\rightarrow 1}) = \int \rho_0 \exp[-\beta(E_1^r(\mathbf{q}^e, \mathbf{q}_1^r) - E_0^r(\mathbf{q}^e, \mathbf{q}_0^r))] d\mathbf{q}^e|_{\mathbf{q}_0^e, \mathbf{q}_1^e \text{ const}} \quad (16)$$

where the configuration  $\mathbf{q}^e$  is often written as  $\mathbf{q}$  or  $\mathbf{q}_0$  as in eq 2, while  $\mathbf{q}^r$  is perceived as a parameter that defines the initial or final state and the corresponding energy function. A volume preserving, bijective (one-to-one) mapping  $\mathcal{M}: \mathbf{q}_0 \rightarrow \mathbf{q}_1$  of each volume element in the initial configuration subspace to a volume element in the final configuration subspace can also be defined if the total configuration space volume (see Figure 1) of the reactive subsystem is equal for the initial and the final state. However, the Zwanzig equation makes no statement on how to define this mapping—it does not need to correspond to a mapping in real physical space.<sup>12–14</sup> Such a defined mapping leads, for example, from setup B to setup A of Bennett's model transformation presented in section 3.2. These considerations make it clear that the Zwanzig equation<sup>10</sup> implicitly assumes a system in which an unequivocal, volume preserving, bijective mapping of initial and final configuration subspaces exists. This prerequisite limits the direct applicability of the Zwanzig equation to transformations between end states of equal total configuration space volume.

### 2.3. Simulations Conducted with Single FEP Moves.

Equation 13 has the disadvantage that at each configuration sampled from the equilibrium distribution of the initial state one would have to evaluate the integral  $\int \exp[-\beta(E_1 - E_0)] d\mathbf{q}_1^r$ , which might be difficult or impossible especially in the case of a continuous configuration space. In the case of a discrete configuration space, the evaluation of this integral simplifies to collecting as many state energy samples  $E_1 - E_0$  as there are configurations of the reactive subsystem in the final state. To circumvent this difficulty, we can simplify eq 13 to use only one perturbation energy sample per data collection step and correct afterward for the deviation from the true free energy difference. Such a form of eq 13 is more practical to use, but requires a decision on how to choose the FEP move to be executed and an appropriate correction. Two possibilities to choose among the FEP moves are explored here. The choice can be made randomly as pursued in the next section or based on the equilibrium distribution of the final state as explored in the subsequent section.

**2.3.1. Perform Random Single FEP Moves.** To obtain a simpler expression as eq 13, we randomly select one configuration of the reactive subsystem  $\mathbf{q}_1^r$  at the final state instead of realizing all possible switching moves

$$\Delta \tilde{F}_{0\rightarrow 1}^{\text{rand}} = -\beta^{-1} \ln \left[ \frac{1}{\Psi_1^r} \int \rho_0 \exp[-\beta(E_1^r - E_0^r)] d\mathbf{q}^e d\mathbf{q}_0^r d\mathbf{q}_1^r \right] \quad (17)$$

This equation corresponds to weighting all switching moves by the same uniform probability  $1/\Psi_1^r$ . In doing so, we do not grant the system time to relax toward equilibrium. Hence, this process can be described as an infinitely fast switching from initial to final state. Note that eq 17 reduces to the Zwanzig equation (eq 16) if the total configuration space volume of the reactive

subsystem is unity in the initial and final states. The difference between the true free energy difference eq 5 and eq 17 is

$$\begin{aligned} \beta\Delta\Delta F_{0\rightarrow 1}^{\text{rand}} &= \beta(\Delta F_{0\rightarrow 1} - \Delta \tilde{F}_{0\rightarrow 1}^{\text{rand}}) \\ &= -\ln \frac{Q_1}{Q_0} + \ln \int \frac{\rho_0}{\Psi_1^r} \exp[-\beta(E_1^r - E_0^r)] d\mathbf{q}^e d\mathbf{q}_1^r d\mathbf{q}_0^r \end{aligned} \quad (18)$$

using eqs 8 and 12 we obtain

$$\beta\Delta\Delta F_{0\rightarrow 1}^{\text{rand}} = -\ln \frac{Q_1}{Q_0} + \ln \int \frac{\exp[-\beta E_0]}{Q_0} \frac{1}{\Psi_1^r} \times \exp[-\beta(E_1 - E_0)] d\mathbf{q}^e d\mathbf{q}_0^r d\mathbf{q}_1^r \quad (19)$$

$$\beta\Delta\Delta F_{0\rightarrow 1}^{\text{rand}} = -\ln Q_1 + \ln \int \frac{1}{\Psi_1^r} \exp[-\beta E_1] d\mathbf{q}^e d\mathbf{q}_0^r d\mathbf{q}_1^r \quad (20)$$

Since there are no terms depending on  $\mathbf{q}_0^r$  remaining, integrating over all  $d\mathbf{q}_0^r$  gives the total configuration space volume of the reactive subsystem in the initial state  $\Psi_0^r$

$$\beta\Delta\Delta F_{0\rightarrow 1}^{\text{rand}} = -\ln Q_1 + \ln \int \exp[-\beta E_1] d\mathbf{q}^e d\mathbf{q}_1^r + \ln \frac{\Psi_0^r}{\Psi_1^r} \quad (21)$$

Finally, we obtain the deviation of our free energy estimate from the true free energy difference as

$$\Delta\Delta F_{0\rightarrow 1}^{\text{rand}} = -\beta^{-1} \ln \frac{\Psi_1^r}{\Psi_0^r} \quad (22)$$

This correction term can often easily be calculated analytically especially in the case of a discrete configuration space. Thus, we can use eqs 17 and 22 to obtain the free energy difference from

$$\Delta F_{0\rightarrow 1} = \Delta \tilde{F}_{0\rightarrow 1}^{\text{rand}} + \Delta\Delta F_{0\rightarrow 1}^{\text{rand}} \quad (23)$$

**2.3.2. Perform Single FEP Moves to an Equilibrated System.** In this section, the single FEP move is performed from the configuration sampled from the equilibrium distribution of the initial state to a configuration sampled from the equilibrium distribution of the final state. Thus, each energy difference in the exponential average of this simulation scheme corresponds to an infinitely slow transition between the corresponding initial and final configurations. That is, the system has enough time to relax to equilibrium during the transition. The free energy estimate obtained from this simulation scheme is described by

$$\Delta \tilde{F}_{0\rightarrow 1}^{\text{equi}} = -\beta^{-1} \ln \int \rho_0 \rho_1 \exp[-\beta(E_1 - E_0)] d\mathbf{q}_0 d\mathbf{q}_1 \quad (24)$$

The integral is taken over all possible pairs of microstates from the initial and final ensemble, where in contrast to the random single move defined by eq 17, the configurations from both

ensembles are weighted by their equilibrium occupation probability. In principle, one could simulate actual trajectories for very slow transitions between initial and final state, but this is not required. In contrast to the Jarzynski equation,<sup>19</sup> eq 24 averages over state energy differences between the initial and final microstate of each transition instead of the amounts of work done along actual trajectories. Since the energy differences in eq 24 make no reference to actual trajectories connecting the microstates  $\mathbf{q}_0$  and  $\mathbf{q}_1$ , there are different equivalent ways to implement the simulation scheme in practice. One could randomly select a microstate from the final ensemble as described in the previous section and equilibrate the system afterward. Alternatively, one could run independent equilibrium simulations of both end states in parallel and take microstate energy differences in certain intervals chosen such as to ensure uncorrelated samples.

The difference between the true free energy difference eq 5 and eq 24 is

$$\begin{aligned}\beta\Delta\Delta F_{0\rightarrow 1}^{\text{equi}} &= \beta(\Delta F_{0\rightarrow 1} - \Delta\tilde{F}_{0\rightarrow 1}^{\text{equi}}) \\ &= -\ln\frac{Q_1}{Q_0} + \ln\int\rho_0\rho_1\exp[-\beta(E_1 - E_0)]d\mathbf{q}_0d\mathbf{q}_1\end{aligned}\quad (25)$$

and using eqs 8 and 12 we obtain

$$\beta\Delta\Delta F_{0\rightarrow 1}^{\text{equi}} = -\ln\frac{Q_1}{Q_0} + \ln\int\frac{\exp[-\beta E_0]}{Q_0}\rho_1\times\exp[-\beta(E_1 - E_0)]d\mathbf{q}_0d\mathbf{q}_1\quad (26)$$

which can be further simplified to

$$\beta\Delta\Delta F_{0\rightarrow 1}^{\text{equi}} = -\ln Q_1 + \ln\int\rho_1\exp[-\beta E_1]d\mathbf{q}_0d\mathbf{q}_1\quad (27)$$

Again, there is no term depending on the initial configuration  $\mathbf{q}_0$  left, so that integrating over all  $d\mathbf{q}_0$  gives the total configuration space volume of the whole system in the initial state  $\Psi_0$ . Writing  $\int\rho_1\exp[-\beta E_1]d\mathbf{q}_1$  as ensemble average  $(\exp[-\beta E_1])_1$ , we obtain

$$\beta\Delta\Delta F_{0\rightarrow 1}^{\text{equi}} = -\ln Q_1 + \ln\langle\exp[-\beta E_1]\rangle_1 + \ln\Psi_0\quad (28)$$

At first sight, this expression looks very discouraging, since we want to use the FEP simulation exactly because the partition function is difficult, if not impossible, to calculate for a large system. But when looking at the equation in more detail, we see something familiar. The first two terms on the right-hand side are very similar to the definition of the total configurational entropy of the final ensemble  $-\beta TS_1 = -\ln Q_1 - (\beta E_1)_1$ , with the average state energy exchanged by the exponential average of the state energy. Thus, we can define a modified configurational entropy of the final ensemble as

$$-\beta_1 T\tilde{S}_1^{\text{def}} = -\ln Q_1 - (-\ln\langle\exp[-\beta E_1]\rangle_1)$$

A possible interpretation of this quantity is examined in more detail in the following two sections. The last term in eq 28 is

the free energy contribution of the ideal, maximum reachable configurational entropy of the system in the initial state. This maximum configurational entropy would be reached if all microstates had equal energy (see section 2.4). Using the definition of the modified entropy  $\tilde{S}$ , eq 28 can be written as

$$\beta\Delta\Delta F_{0\rightarrow 1}^{\text{equi}} = -\beta T\tilde{S}_1 + \ln\Psi_0\quad (29)$$

We define the effective configuration space volume

$$\tilde{\Omega}^{\text{def}} = \exp[\beta T\tilde{S}_1]$$

and rewrite eq 29 as

$$\beta\Delta\Delta F_{0\rightarrow 1}^{\text{equi}} = -\ln\frac{\tilde{\Omega}_1}{\Psi_0}\quad (30)$$

Using eq 30, the free energy difference  $\Delta\tilde{F}_{0\rightarrow 1}$  is given by

$$\Delta F_{0\rightarrow 1} = \Delta\tilde{F}_{0\rightarrow 1}^{\text{equi}} + \Delta\Delta F_{0\rightarrow 1}^{\text{equi}}\quad (31)$$

The correction term eq 30 involves the effective configuration space volume of the final state  $\tilde{\Omega}_1$ . The interpretation of this quantity as volume of the thermally accessible portions of the total configuration space volume of the final state is addressed in detail in the following sections. In the remainder of this section, we will be concerned with the derivation of an expression for the calculation of this quantity.

A possibility for the calculation of  $\tilde{\Omega}$  follows from the expressions for the free energy difference  $\Delta F_{0\rightarrow 1}$  derived for our two single-move simulation schemes. Combining the right-hand sides of eqs 31 and 23 and solving for  $\tilde{\Omega}$  gives

$$\frac{\tilde{\Omega}_1}{\Psi_1} = \exp[-\beta(\Delta\tilde{F}_{0\rightarrow 1}^{\text{rand}} - \Delta\tilde{F}_{0\rightarrow 1}^{\text{equi}})]\quad (32)$$

Direct use of this equation results in a large statistical uncertainty of the obtained estimate of  $\tilde{\Omega}_1$ . A better expression can be found by noticing that the left-hand side is independent of the initial state 0 used in computing the averages on the right-hand side. The statistical uncertainty of free energy estimates computed with free energy perturbation theory is known to be minimized by maximizing the similarity or overlap of the probability distributions of the initial and final state.<sup>11,15,20</sup> Thus, one would expect to maximize the efficiency of the calculation by using the same state as initial and final state in the calculation of the averages  $\tilde{F}_{0\rightarrow 1}^{\text{rand}}$  and  $\tilde{F}_{0\rightarrow 1}^{\text{equi}}$

$$\frac{\tilde{\Omega}_1}{\Psi_1} = \exp[-\beta(\Delta\tilde{F}_{1\rightarrow 1}^{\text{rand}} - \Delta\tilde{F}_{1\rightarrow 1}^{\text{equi}})]\quad (33)$$

where  $\Delta\tilde{F}_{1\rightarrow 1}^{\text{rand}}$  is equal to 0. Expanding  $\Delta\tilde{F}_{1\rightarrow 1}^{\text{equi}}$  according to eq 24 and using  $\exp[\ln x] = x$  gives

$$\frac{\tilde{\Omega}_1}{\Psi_1} = \frac{1}{\int\rho'_1\rho_1\exp[-\beta(E'_1 - E_1)]d\mathbf{q}'_1d\mathbf{q}_1}\quad (34)$$

The integration on the right-hand side can be separated into two independent integrals

$$\frac{\tilde{\Omega}_1}{\Psi_1} = \frac{1}{\left(\int \rho'_1 \exp[-\beta E'_1] d\mathbf{q}'_1\right) \left(\int \rho_1 \exp[\beta E_1] d\mathbf{q}_1\right)} \quad (35)$$

This equation results in a reduced statistical uncertainty of the calculated  $\tilde{\Omega}_1$  as expected. Equation 35 formally allows one to calculate the effective configuration space volume from a set of  $N$  uncorrelated samples drawn from an equilibrium simulation as

$$\tilde{\Omega}_1 = \frac{\Psi_1 N^2}{\sum_i \exp[-\beta E_i] \sum_j \exp[\beta E_j]} \quad (36)$$

The estimate of  $\tilde{\Omega}_1$  obtained from such a simulation still converges slowly to the correct absolute value but does so significantly faster than an estimate based on eq 32. Besides the more favorable intrinsic convergence properties, also the computation of  $\Delta\tilde{F}_{0 \rightarrow 1}^{\text{rand}}$  is not required when using eq 36, resulting in an extra saving of computation time.

**2.4. Configurational Entropy and the Effective Configuration Space Volume.** In this section, we want to highlight one of the numerous possible interpretations of the quantity entropy, which will enable us to derive an interpretation of the correction terms derived in sections 2.3.1 and 2.3.2. The meaning of the correction terms is examined in detail in section 2.5.

The configurational part of the Boltzmann entropy is calculated from

$$-TS = \beta^{-1} \int \rho \ln \rho d\mathbf{q} \quad (37)$$

The meaning of the quantity entropy can be interpreted in various different ways. Two of them could be useful here. One follows from comparison with the ideal, maximum achievable configurational entropy of the system, which is also encountered in the microcanonical ensemble

$$-TS = -\beta^{-1} \ln \Omega \quad (38)$$

where  $\Omega$  is the partition function of a microcanonical ensemble.  $\Omega$  is often interpreted as the volume of the thermally accessible configuration space, hence also termed effective configuration space volume.

The second possible interpretation of the entropy follows from rearranging eq 37

$$\begin{aligned} -TS &= \beta^{-1} \int \rho \ln \rho d\mathbf{q} \\ &= \beta^{-1} \int \rho \ln \frac{\exp[-\beta E]}{Q} d\mathbf{q} \\ &= \beta^{-1} (-\ln Q + \int \rho \ln \exp[-\beta E] d\mathbf{q}) \\ &= \beta^{-1} (-\ln Q - \int \rho[\beta E] d\mathbf{q}) \end{aligned} \quad (39)$$

Writing this as

$$-TS = -\beta^{-1} \ln Q - \langle E \rangle \quad (40)$$

we can interpret the entropy as the mean deviation of the state energy from the absolute free energy. In section 2.3.2, a quantity similar to the configurational entropy emerged from our derivation, in which the arithmetic ensemble average of the state energy is substituted with the exponential ensemble average

$$-T\tilde{S} = -\beta^{-1} \ln Q + \beta^{-1} \ln \langle \exp[-\beta E] \rangle \quad (41)$$

This modified entropy can be rearranged to be written in a form similar to the Boltzmann entropy in eq 37

$$\begin{aligned} -T\tilde{S} &= -\beta^{-1} \ln Q + \beta^{-1} \ln \langle \exp[-\beta E] \rangle \\ &= -\beta^{-1} \ln Q + \beta^{-1} \ln \int \rho \exp[-\beta E] d\mathbf{q} \\ &= -\beta^{-1} \ln \int \rho \frac{\exp[-\beta E]}{Q} d\mathbf{q} \end{aligned} \quad (42)$$

from which we finally obtain

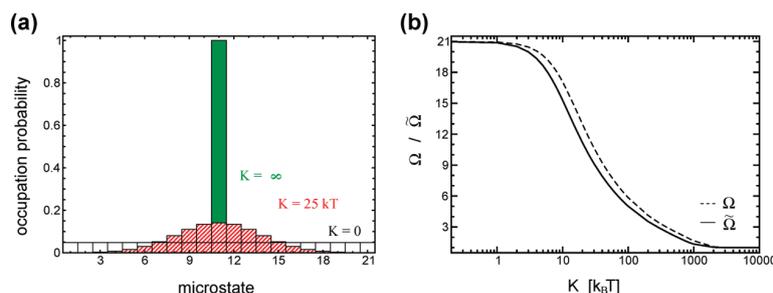
$$-T\tilde{S} = \beta^{-1} \ln \int \rho^2 d\mathbf{q} \quad (43)$$

Interestingly, the form of this result shows analogy to the corresponding equation for the Boltzmann entropy eq 37. A practical advantage of this entropy formulation over eq 37 is that it does not suffer numerical problems in the case of zero probabilities. By analogy to eq 38, we can write now a measure of the effective configuration space volume

$$\tilde{\Omega} = \exp[\beta T\tilde{S}] = \frac{1}{\int \rho^2 d\mathbf{q}} \quad (44)$$

This result has been obtained before by Sussmann in a quantum mechanical context as a measure of uncertainty about the position of a particle in phase space.<sup>21</sup> This interpretation is absolutely analogous to that as effective configuration or phase space volume in the present context. The difference is that a quantum mechanical system can often even be described by a blend of different microstates at a single point in time, while the uncertainty about the position of the system in phase space in the classical case arises from occupation of different microstates at different points in time. A similar interpretation of entropy, as measure of the incompleteness of knowledge about the current microstate of the system, was already provided by the seminal works of Edwin Jaynes, which established a formulation of statistical mechanics in terms of information theory.<sup>22</sup> See ref 28 for a review on the relation of entropy and information in physics. We note that the functional form of the Sussmann entropy eq 43 is formally identical to that of the information theoretic Rényi entropy of order two.<sup>24</sup>

Regarding the relation of Boltzmann entropy and Sussmann entropy, we can see from comparison of eqs 40 and 41 that the two entropy measures are equal if the exponential average of the state energy equals its arithmetic average. This condition is met in the microcanonical ensemble or in a degenerate system in which all configurations have the same state energy. From the considerations in this section, it is not yet entirely clear to us if there is more physical meaning hidden within the difference of Boltzmann and Sussmann entropy than the obvious math-



**Figure 2.** Occupation probability and effective (thermally accessible) configuration space volume available to a classical particle in a one-dimensional box with a discretized harmonic potential. See text for a detailed description of the system. (a) The occupation probability distribution of the model system is plotted for its 21 microstates using three different harmonic potentials of different strength determined by the parameter  $K$ . (b) The effective configuration space volume  $\tilde{\Omega} = \exp[\beta TS]$  is plotted as a function of  $K$  (solid curve). For comparison, the plot contains also  $\Omega = \exp[\beta TS]$  (dashed curve).

emational differences tell. Since this question is outside the immediate scope of this paper, we will not pursue it further here. The interested reader is referred to section C of the Supporting Information for a more elaborate analysis of this question. Nevertheless, we can deduce a possible interpretation of the Sussmann entropy as effective configuration space volume occupied by a system in a specific macrostate, i.e., an ensemble of microstates using eq 44. This interpretation is fortified in the next section by an analogy argument regarding the meaning of the derived correction terms eqs 22 and 30. The meaning of the effective configuration space volume and its dependence on the shape of the state energy landscape is illustrated by a simple example system at the beginning of the section Examples and by a slightly more complicated example in the Supporting Information.

**2.5. Meaning of the Correction Terms for Simulations Conducted with Single FEP Moves.** For the random single-move scheme outlined in section 2.3.1, the correction term eq 22 describes the relative total volume of the initial and final configuration subspace. The interpretation of the correction term for the single FEP move scheme with selection of the target state from the equilibrium distribution eq 30 is analogous. We restate eqs 22 and 30 here to emphasize the analogy. The correction term for the equilibrated single-move scheme is

$$\Delta\Delta F_{0 \rightarrow 1}^{\text{equi}} = -\beta^{-1} \ln \frac{\tilde{\Omega}_1}{\Psi_0} \quad (45)$$

The configurational entropy can be interpreted as measure for the effective volume of the configuration space (see section 2.4). In the case of a discrete configuration space,  $\tilde{\Omega}$  is the effective number of configurations thermally accessible to the system. The correction term for the random single-move scheme reads

$$\Delta\Delta F_{0 \rightarrow 1}^{\text{rand}} = -\beta^{-1} \ln \frac{\Psi_1}{\Psi_0} \quad (46)$$

where we multiplied the numerator and the denominator of the fraction in eq 22 with  $\Psi^c$  and substituted the products of the total configuration space volumes of the two subsystem with the total configuration space volume of the whole system (see eq 7). Comparing eqs 45 and 46, one can easily spot the analogy between  $\tilde{\Omega}_1$  and  $\Psi_1$ . Since we collected only one state energy difference per sampling step corresponding to one transition instead of integrating over all possible FEP moves as prescribed by eq 13, we have to correct for the actual number of possible

transitions. The ratio of the numbers of possible forward and backward transitions between initial and final state is given by the relative configuration space volumes appearing in the correction terms. The numbers of possible transitions is determined by the numbers of configurations in each configuration subspace which are accessible as initial or final configuration of the perturbation. The accessibility of a configuration as final point of the perturbation is determined by the probability of selecting a configuration in the respective single-move FEP scheme. For the random single-move scheme (eq 23), there is a uniform probability for all configurations of the final configuration subspace that is given by  $\Psi_1^{-1}$  (or by  $(\Psi_1^c)^{-1}$  if the variables are subdivided into environment and reactive subsystem). For the equilibrated single-move FEP scheme (eq 31) the probability to select a configuration as the final point of the perturbation is given by its equilibrium probability.

### 3. Examples

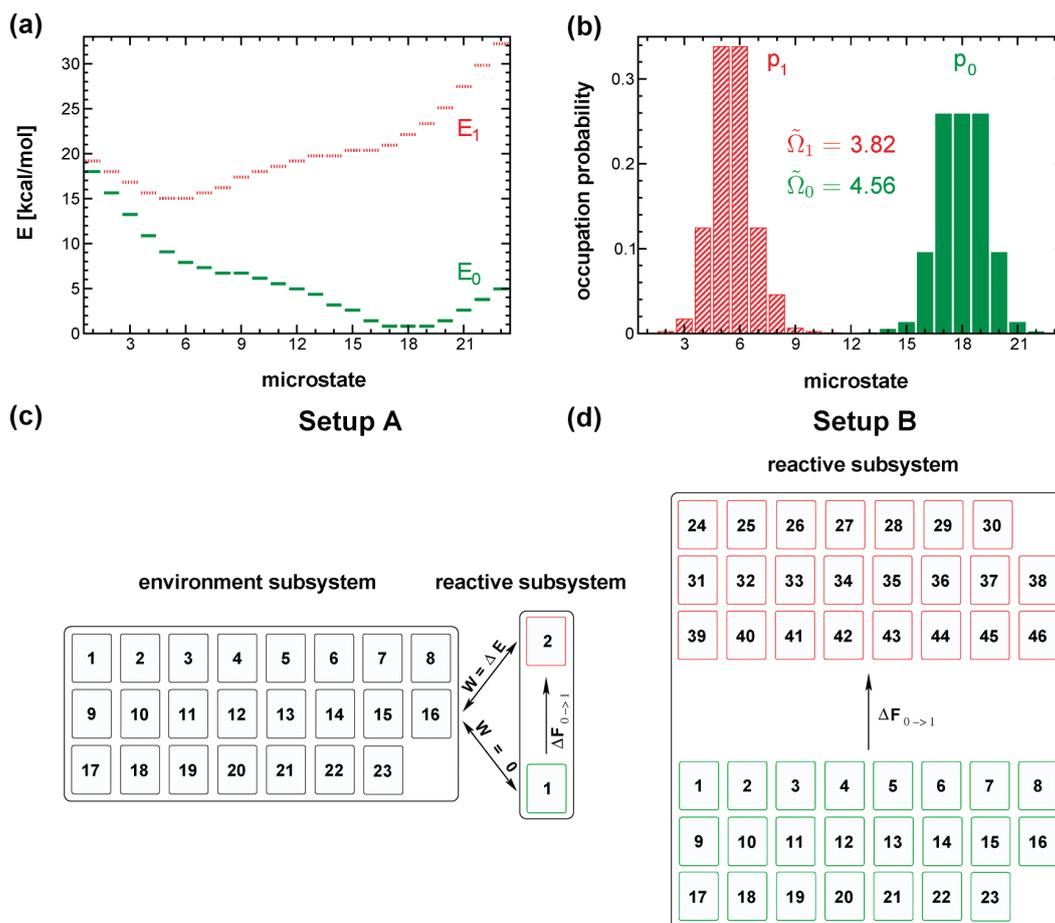
This section has the purpose to illustrate the theory derived in the previous section and its application with the aid of some model systems. These model systems are simple enough to be analytically explored on a piece of paper but sufficiently complex to underline the importance of our theoretical findings.

**3.1. Effective Configuration Space Volume Occupied by a Classical Particle in a Box with a Discretized Harmonic Potential.** Consider a model system consisting of 21 discrete microstates in a one-dimensional box evenly distributed along the coordinate  $0 \leq \mathbf{q} \leq 1$ . The energy of the microstate  $i$  is given by the harmonic potential

$$E_i = K(0.5 - \mathbf{q}_i)^2 \quad (47)$$

The potential is symmetric around the center of the length interval at  $\mathbf{q} = 0.5$  and its strength is defined by the constant  $K$ . The energy of the microstates is evaluated at the discrete points  $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{21}\} = \{0, 1/20, \dots, 1\}$ .

The population probability of the microstates and the effective configuration space volume  $\tilde{\Omega}$  are plotted in Figure 2. For comparison, the plot also contains the analogous quantity  $\Omega$  derived from the Boltzmann entropy for this model system. Figure 2b shows the effect of  $K$  on the effective configuration space volume  $\tilde{\Omega}$ . If the energy landscape is very flat ( $K \rightarrow 0$ ), all microstates are thermally accessible and so  $\tilde{\Omega} = 21$ . If on the other hand the harmonic potential is very steep, the particle is trapped at the center of the box and will essentially never populate a state other than the central one, which is reflected by the convergence of  $\tilde{\Omega}$  to 1 for  $K \rightarrow \infty$ . In the extreme limits,



**Figure 3.** Bennett's model system.<sup>11</sup> The system can be set up in two different, yet fully equivalent, ways. The setups exemplify the consequences of a reactive subsystem with just one or multiple possible configurations in the initial and final state for the outcome of FEP simulations performed in different ways. See the text for a detailed description of the system. (a) The green solid and the red dashed profiles show the energies of the 23 microstates of the initial state and the final state, respectively. (b) The green filled columns show the occupation probabilities of the 23 microstates of the initial state while the red hatched columns show those of the final state. (c) Setup A: An unequivocal one-to-one mapping of the initial and final configurations exists, making the application of the Zwanzig equation possible. (d) Setup B: There are multiple possibilities to define a one-to-one mapping of the initial and final configurations. Mapping the configurations in the order in which they are numbered in the figure leads to setup A. In (c) and (d) the configurations are symbolized by the rectangles with the numbering inscribed.

the meaning of  $\tilde{\Omega}$  is thus easily understandable. In the intermediate regime, however, fractional values of  $\tilde{\Omega}$  can occur. These fractional values of  $\tilde{\Omega}$  in case of discrete microstates can be understood considering eq 44. One realizes that  $\tilde{\Omega} = 1/(\int p^2 d\mathbf{q})$  is the inverse integral of the squared occupation probability carried out over the whole configuration space. The integration can be visualized as summation of the squares of the microstate occupation probabilities  $p$  depicted for our example system by the bars in Figure 2a. The interpretation of the integral  $\int p^2 d\mathbf{q}$  is straightforward—it is the mean probability to find the system in the same microstate when comparing two uncorrelated samples picked from the ensemble.

**3.2. Bennett's Model Transformation.** The second example that we want to present here is a model transformation taken from an article of Charles Bennett.<sup>11</sup> In this article, Bennett introduced his acceptance ratio method which allows to extract the most probable value of the free energy difference from a given set of FEP samples. This method can be used to minimize the statistical error of the free energy estimate obtained from a FEP simulation.<sup>11,25</sup> An outline of the method in the context of our formalism is given in section B2 of the Supporting Information. We will not make use of the method in this paper, but we wish to use the model transformation proposed by

Bennett as an example for the calculation of the transformation free energy from the different simulation schemes derived in the theory section.

The system consists of an ensemble of 23 microstates in the initial state and an ensemble of 23 microstates in the final state. The microstate energy surfaces of the initial and final state are plotted in Figure 3 along with the resulting occupation probability distributions. This example system has the nice property that it can be formulated in two equivalent ways: one in which the reactive subsystem has just one possible configuration in the initial and final states (setup A) and one in which the reactive subsystem has multiple possible configurations in both states (setup B). In setup A shown in Figure 3c, the system consists of an environment subsystem with 23 possible configurations and a reactive subsystem with 2 possible configurations where occupation of configuration 1 defines the initial state and occupation of configuration 2 defines the final state. In setup A, the energy functions of the initial and final state are given by

$$E_0 = E_i^{\text{intr}}; \quad E_1 = E_i^{\text{intr}} + W_i \quad (48)$$

Each configuration  $i$  of the environment subsystem contributes a specific energy  $E_i^{\text{env}}$  to the total energy of the system, which is equal in the initial and final state. This energy contribution is equal to the energy of the corresponding microstate from the initial ensemble as depicted in Figure 3a. The switching of the energy function from state 0 to state 1 is realized in setup A by an interaction of configuration 2 of the reactive subsystem with the environment subsystem. The reactive subsystem does not interact with the environment subsystem if it adopts configuration 1. The magnitude of the interaction energy  $W_i$  is specific to the configuration  $i$  of the environment subsystem and equal to the respective microstate energy difference between the initial and final states given for each of the 23 microstates in the appendix of ref 11 (see Figure 3a). In setup B, the energy of the system is given by the energy of each of the 23 configurations in initial or final ensemble given in the appendix of ref 11 (see Figure 3a).

The two possible setups of the system are ideally suited to illustrate the mapping between the configuration subspaces of the end states. In setup A, there is an unequivocal one-to-one mapping of each configuration from the initial state to exactly one configuration of the final state and vice versa. The mapping  $\mathcal{M}: \mathbf{q}^e \mathbf{q}_0^i \rightarrow \mathbf{q}^e \mathbf{q}_1^i$  is unequivocal because there is only one configuration of the reactive subsystem  $\mathbf{q}^r$  for each of the end states. Changing the configuration of the reactive subsystem constitutes the unique switching move between the initial and final configurations derived in section 2.2 as prerequisite for the applicability of the Zwanzig equation, eq 1. Picking up the original description of FEP by Zwanzig,<sup>10</sup> the configuration of the reactive subsystem would be interpreted as control parameter which perturbs the energy function, while the configuration of the environment subsystem would be perceived as the configuration  $\mathbf{q}$  appearing in eq 1. The switching move can then be interpreted as changing the energy function, while keeping the configuration of the system as described by Bennett.<sup>11</sup> In contrast to the situation in setup A, there is no unequivocal one-to-one mapping of the initial and final configurations in setup B, because there are multiple configurations of the reactive subsystem at each end state, leading to multiple possible FEP moves between the initial and final state. However, since the number of configurations (the total configuration space volume) is the same in the initial and final state, it is possible to define a one-to-one mapping of the initial and final configurations by forming pairs of configurations from the initial and final state. One of the possible one-to-one mappings can be chosen to make the application of the Zwanzig equation possible.<sup>13</sup> In our example, the configurations of setup B are simply paired in the order in which they are listed (1 is mapped to 24, 2 to 25, ...), leading back to setup A, which is the original arrangement given by Bennett.<sup>11,26</sup>

Table 1 shows analytical results of FEP calculations performed according to the different schemes developed in section 2 and compares them to the exact free energy difference obtained from eq 5. The multimove FEP formula given by eq 13 demands collecting state energy difference samples for all possible FEP moves and leads to the correct results with both setups. Simulations performed with single FEP moves without the appropriate correction term do not always lead to correct results. A simulation performed with random single FEP moves leads to the correct results without the correction term for both, setup A and B. Why this is the case becomes clear upon examination of the correction term corresponding to this simulation scheme given by eq 22. The correction term adopts a value of zero for the two setups, because the number of theoretically possible

**TABLE 1: Results from Perfect FEP Simulations at Setup A and Setup B of Bennett’s Model System<sup>a</sup>**

		free energy difference (kcal/mol)			
		forward	−reverse	mean	difference
Setup A					
$F_{0 \rightarrow 1}$	exact/multimove	14.37791	14.37791	14.37791	0
$\tilde{F}_{0 \rightarrow 1}^{\text{and}}$	uncorrected	14.37791	14.37791	14.37791	0
	deviation	0	0	0	
	corrected	14.37791	14.37791	14.37791	0
	deviation	0	0	0	
$\tilde{F}_{0 \rightarrow 1}^{\text{qui}}$	uncorrected	15.17150	13.47887	14.32519	1.69262
	deviation	0.79359	−0.89904	−0.05272	
	corrected	14.37791	14.37791	14.37791	0
	deviation	0	0	0	
Setup B					
$\tilde{F}_{0 \rightarrow 1}^{\text{and}}$	uncorrected	14.37791	14.37791	14.37791	0
	deviation	0	0	0	
	corrected	14.37791	14.37791	14.37791	0
	deviation	0	0	0	
$\tilde{F}_{0 \rightarrow 1}^{\text{qui}}$	uncorrected	14.92590	13.72447	14.32519	1.201425
	deviation	0.65343	−0.54799	−0.05272	
	corrected	14.37791	14.37791	14.37791	0
	deviation	0	0	0	

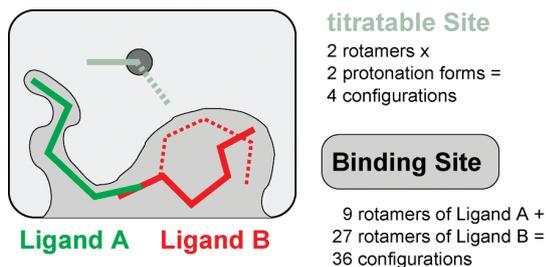
<sup>a</sup> The data is calculated analytically using eqs 17 and 23 for the uncorrected and corrected  $\tilde{F}_{0 \rightarrow 1}^{\text{and}}$  and eqs 24 and 31 for the uncorrected and corrected  $\tilde{F}_{0 \rightarrow 1}^{\text{qui}}$ . The results are checked to be exactly reproducible in actual simulations within statistical error. The calculated values are compared to the exact statistical mechanics result  $F_{0 \rightarrow 1}$  calculated from eq 5. The exact values are identical to those calculated from the multimove FEP formula eq 13.

configurations of the reactive subsystem  $\Psi^r$  is equal in the initial and final states for both setups. In case of a simulation with single FEP moves selected according to the equilibrium probability distribution, the situation is different. Deviations arise for both setups without the correction term given by eq 30. It can be observed that the correction term for simulations conducted in forward direction  $0 \rightarrow 1$  differs from that for simulations conducted in the reverse direction  $1 \rightarrow 0$ . The reason for this asymmetry is the different effective configuration space volume populated by the system in the initial and final state as can be seen from eq 30 and Figure 3b. One can see that the deviations from the exact result obtained with this simulation scheme can be fully corrected for by the appropriate correction term eq 30.

### 3.3. Relative Binding Free Energy from Alchemical FEP.

In this model example, we go one step further and examine a system with a differing number of reactive subsystem configurations at the two end states. The example is adopted from an application of FEP simulations frequently encountered in computational drug design pioneered by Tembe and McCammon.<sup>27</sup>

The aim is to calculate the relative binding free energy of two ligands to a receptor. For simplicity, we assume that the chemical potential of both ligands in solution is equal, so that the task reduces to the calculation of the free energy change upon exchanging ligand A with ligand B in the binding site. The model system setup for the relative ligand binding free energy example, as depicted in Figure 4, consists of a receptor with the ligand binding site and a pH-titratable group. The titratable group can adopt two proton binding forms (bound or unbound, one of which is charged) and two rotamers each oriented toward one of the two niches of the ligand binding site. The ligands A and B bind exclusively to the two distinct but overlapping niches of the ligand binding site (i.e., only one



**Figure 4.** Model system setup for the relative ligand binding free energy example consists of a receptor with the ligand binding site and a pH-titratable group. The transformation is defined by the exchange of ligand A with ligand B in the binding site. See text for detailed explanation.

of them at a time) and interact favorably with the charged form of the titratable site. The flexibility of the ligands is treated by considering three rotamers of each dihedral angle. Since ligand A possesses two rotatable bonds and ligand B possesses three rotatable bonds, this discretization results in nine rotameric forms of ligand A and 27 rotameric forms of ligand B. Ligand B can be accommodated by the binding site in all of its rotameric forms, while ligand A can only be adopted in one of its rotameric forms by its binding niche, but is bound more tightly and interacts also more favorably with the charged form of the titratable site. The state energy of this model is given by

$$E(i, j) = E_i^{\text{intr}} + E_j^{\text{intr}} + W_{ij} \quad (49)$$

with the energy contributions

$$E_i^{\text{intr}} = \begin{cases} -3.0 \text{ kcal/mol} & \text{if } i = 1 \\ 0.0 \text{ kcal/mol} & \text{if } 2 \leq i \leq 9 \\ -2.0 \text{ kcal/mol} & \text{if } 10 \leq i \leq 36 \end{cases}$$

$$E_j^{\text{intr}} = \begin{cases} 0.0 \text{ kcal/mol} & \text{if } j = \{1, 2\} \\ 2.0 \text{ kcal/mol} & \text{if } j = \{3, 4\} \end{cases}$$

$$W_{ij} = \begin{cases} -4.3 \text{ kcal/mol} & \text{if } j = 3; i = 1 \\ -4.0 \text{ kcal/mol} & \text{if } j = 4; 10 \leq i \leq 36 \\ 0.0 \text{ kcal/mol} & \text{otherwise} \end{cases} \quad (50)$$

where  $i$  is the configuration of ligand A or B in the ligand binding site and  $j$  is the configuration occupied by the titratable site. The first  $i = \{1, \dots, 9\}$  configurations of the binding site are the rotameric forms of ligand A, while the remaining  $i = \{10, \dots, 36\}$  configurations of the binding site correspond to the rotameric forms of ligand B. The titratable site is oriented toward the binding niche of ligand A in the configurations  $j = \{1, 3\}$  and oriented toward the binding niche of ligand B when adopting the configurations  $j = \{2, 4\}$ .

The results of FEP simulations conducted according to our three different simulation schemes are given in Table 2. It can be seen that all simulation schemes yield the correct result. FEP simulations employing a random single-move scheme (eq 17) or a single-move scheme according to the target ensemble equilibrium distribution (eq 24) without the appropriate correction term lead to severe deviations from the true relative binding free energy. Clearly, the observed deviations in the uncorrected FEP simulations conducted with single moves originate from the failure to account properly for the relative statistical weight of the ensembles. The appropriate correction terms derived in section 2 fully correct for the deviations. The

**TABLE 2: Results from Perfect FEP Simulations at the Relative Ligand Binding Free Energy Example Calculated Analytically Using Eqs 17 and 23 for the Uncorrected and Corrected  $\tilde{F}_{0 \rightarrow 1}^{\text{rand}}$  and Eqs 24 and 31 for the Uncorrected and Corrected  $\tilde{F}_{0 \rightarrow 1}^{\text{equi}}$ <sup>a</sup>**

		free energy difference (kcal/mol)			
		forward	−reverse	mean	difference
$F_{0 \rightarrow 1}$	exact/multimove	−0.67	−0.67	−0.67	0
$\tilde{F}_{0 \rightarrow 1}^{\text{rand}}$	uncorrected	−0.02	−0.02	−0.02	0
	deviation	0.65	0.65	0.65	
	corrected	−0.67	−0.67	−0.67	0
	deviation	0	0	0	
$\tilde{F}_{0 \rightarrow 1}^{\text{equi}}$	uncorrected	0.06	1.24	0.65	−1.17
	deviation	0.73	1.90	1.32	
	corrected	−0.67	−0.67	−0.67	0
	deviation	0	0	0	

<sup>a</sup> The results are checked to be exactly reproducible in actual simulations within statistical error. The calculated values are compared to the exact statistical mechanics result  $F_{0 \rightarrow 1}$  calculated from eq 5. The exact values are identical to those calculated from the multimove FEP formula eq 13.

importance of such contributions to the free energy of binding has already been well recognized in the literature<sup>8,28,29</sup> and is further underlined by our example. We show that application of FEP with a single move per data collection point chosen randomly or according to target state equilibrium distribution without proper correction terms can lead to quantitatively and even qualitatively wrong results.

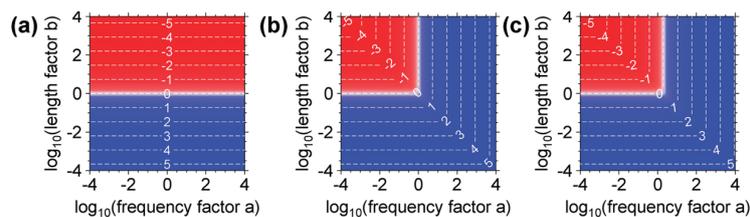
**3.4. Relative Free Energy of Two Classical Particle in the Box Models with a Single-Well or a Double-Well Harmonic Potential in Continuous Coordinates.** This model example demonstrates the application of our formalism to a system formulated in nondimensional, continuous coordinates. The model system consists of a particle of mass  $m$  in a one-dimensional box of length  $L$ . The spatial coordinate  $\mathbf{q}$  is defined on the interval  $-L/2 \leq \mathbf{q} \leq L/2$  and constitutes the reactive subsystem. The momentum  $\mathbf{p}$  must not be confused with the occupation probability  $p$ . The range of values available to the momentum variable is not restricted at either end state and does consequently not differ among them. Thus, the momentum constitutes the environment subsystem. The potential energy is given by a single-well or double-well harmonic potential denoted by  $sw$  or  $dw$ , respectively. The Hamiltonian of the system is given by

$$H^{\text{sw}} = \frac{\mathbf{p}^2}{2m} + \frac{m\omega^2 \mathbf{q}^2}{2}$$

$$H^{\text{dw}} = \frac{\mathbf{p}^2}{2m} + \begin{cases} \frac{m\omega^2}{2} \left( \mathbf{q} + \frac{L}{4} \right)^2 & \text{if } \mathbf{q} \leq 0 \\ \frac{m\omega^2}{2} \left( \mathbf{q} - \frac{L}{4} \right)^2 & \text{if } \mathbf{q} > 0 \end{cases} \quad (51)$$

where the kinetic energy is given by the first terms on the right-hand sides, whereas the potential energy is given by the second terms on the right-hand sides. The strength of the harmonic potential is determined by the angular frequency  $\omega$  of the oscillator. The single-well model is equivalent to a (confined) classical harmonic oscillator.

In a classical formulation, as adopted here, the Hamiltonian can be separated into the momentum-dependent and configuration-dependent terms which are mutually independent. This property, together with the unrestricted range of values available to the two end states, has the consequence that all momentum-



**Figure 5.** Correction terms for the single-move FEP schemes as function of the frequency scaling factor  $a$  and the box length scaling factor  $b$ . Positive and negative values are indicated by blue and red color, respectively. The contour values are given in kcal/mol. The particle mass is set to  $m = 2.0 \times 10^{-26}$  kg. The box length of the end states 0 and 1 are  $L_0 = 1.6 \times 10^{-9}$  m and  $L_1 = bL_0$ , where  $b > 0$  is a constant factor. The corresponding angular frequencies are  $\omega_0 = 10^{-12}$  s $^{-1}$  and  $\omega_1 = a\omega_0$  where  $a > 0$  is a constant factor. (a)  $\Delta\Delta F^{\text{rand}}$ : given otherwise equal parameters, the correction term is identical for transformations leading from single-well to single-well or from single-well to double-well harmonic potential. (b)  $\Delta\Delta F^{\text{equi}}$  for transformations of a single-well to a single-well harmonic potential. (c)  $\Delta\Delta F^{\text{equi}}$  for transformations involving a transition from a single-well to a double-well harmonic potential.

dependent terms cancel from free energy differences. Thus, the kinetic energy terms can be omitted from the calculation.

A detailed derivation of the theory underlying this example, plots of the potential energy and the probability distributions in configuration or phase space are given in the Supporting Information. Additionally, the Supporting Information contains example calculations and plots of the effective phase and configuration space volume as function of the system parameters.

Figure 5 shows the dependency of correction terms for the single-move FEP simulation schemes defined by eqs 45 and 46 for transformations which can involve a change in box length, angular frequency, and potential type (single well or double well). It can be seen that the correction term for the random single-move scheme depends only on differences in the box length between the end states, but not on changes in the angular frequency or in the harmonic potential type of the final state. This is because  $\Delta\Delta F^{\text{rand}}$  depends only on the ratio of the total configuration space volumes of the end states, which is given by the box length. In contrast, the correction terms for the equilibrated single-move scheme depend also on the harmonic potential type and on the angular frequency of the final state, because these correction terms depend on the ratio of the effective configuration space volume of the final state and the total configuration space volume of the initial state. While the total configuration space volume is given by the box length, the effective configuration space volume is a function of all system parameters (see Table 1 of the Supporting Information).

Comparison of Figure 5, b and c, shows that the correction terms for the transformations of a system with a single-well to another system with a single-well potential are very similar to those of the corresponding transformations which substitute the single-well with a double-well potential. This can be explained by the logarithmic dependence of  $\Delta\Delta F^{\text{equi}}$  on the effective configuration space volume of the final state, considering that the effective configuration space volume of the double-well model differs at most by a factor 2 from that of the corresponding system with a single-well potential (see Figure 2 of the Supporting Information).

Furthermore, it can be seen that the correction terms can reach considerable values, especially if the box length or the frequency differs widely between the end states. Especially for the equilibrated single-move scheme, the correction terms are often comparable in magnitude to the total transformation free energies themselves (plots of the transformation free energies can be found in the Supporting Information). The Supporting Information also contains further example calculations and plots showing the free energy changes and correction terms for two more example transformations involving a change in box length and particle number or in angular frequency and particle number.

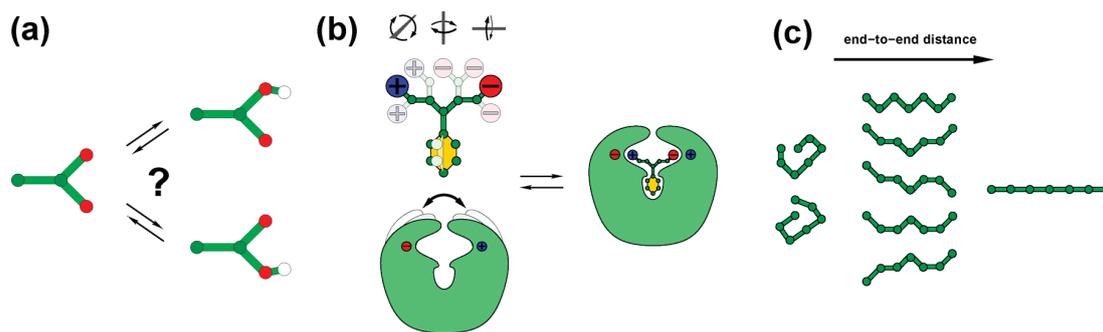
#### 4. Discussion

In this section we will discuss applications for which the use of our FEP theory might be advantageous, advantages and disadvantages of the different simulation schemes, and possible perspectives for further developments of the FEP theory.

**4.1. Applications.** Figure 6 shows a schematic depiction of applications for which our formalism could be advantageous. The formalism is especially well suited for studying transformations with end states of differing configuration space volume. The use of our simulation schemes might, however, also be favorable for transformations with end states of equal configuration space volume that are treatable with traditional FEP according to the Zwanzig equation.

Figure 6a shows the protonation of a carboxylic acid residue with multiple tautomers of the protonated form. This problem was already used as motivating example in the introduction of this article. The problem is a simple representative of a whole class of free energy calculations on systems that are defined on discrete configuration spaces. The number of configurations for either of the end states can be much larger, as for example if rotameric forms of protein side chains or multiple redox and protonation forms of protein cofactors are considered. Our novel formalism has already been fruitfully applied in such free energy calculations on different systems. One application was the calculation of free energy changes in coupled protonation and reduction processes in the small electron carrier protein azurin from *Pseudomonas aeruginosa*. The results of these structure-based calculations are in very good agreement with experimental results<sup>30,31</sup> and provide a microscopic picture of the underlying thermodynamics. Another application was the calculation of free energy differences for small molecule binding processes to study the permeation of ammonium and ammonia through the membrane transporter protein Amt-1 from *Archaeoglobus fulgidus*. The calculations were based on the experimental crystal structure of the transporter determined by Andrade and co-workers.<sup>32</sup> The results of our calculations were compared to experimental results of the same group and aid in their interpretation. A particularly interesting part of both applications is the calculation of free energy measures of cooperativity.<sup>33</sup> These works are in preparation for publication and will be presented elsewhere. The FEP simulation schemes of the present article are implemented in our Monte Carlo simulation suite GMCT that will be made publicly available as open-source software.<sup>9</sup>

Figure 6b shows the binding of a flexible ligand to a flexible receptor as representative for the calculation of absolute binding free energies. Such calculations are often done with molecular mechanics models to account for the flexibility of the binding



**Figure 6.** Three example applications of free energy simulations for which it could be crucial to consider the principles presented in the present article. (a) Calculation of protonation free energies with multiple tautomers. One can recognize the similarity to the motivating example from section 1. Similar situations arise in the treatment of other amino acids (e.g., histidine) and many important biological cofactors like flavines and quinones.<sup>9</sup> (b) General binding free energy calculations with flexible ligands and/or flexible receptors. (c) Calculation of (un)folding free energies for flexible polymers like proteins or nucleic acids.

partners.<sup>8,34</sup> The necessity to consider the change in the total configuration space volume upon binding was recognized earlier and investigated in depth by Luo and Sharp<sup>29</sup> and Hermans and Wang.<sup>35</sup> In their formalism,  $\Psi$  occurs as a spatial volume, relating the volume of the simulated system in the bound state to some reference volume in the unbound state. The reference volume corresponds to some reference concentration of the unbound ligand. Our formalism automatically includes this concept and generalizes it to include all degrees of freedom of the system. The change in the total configuration space volume can be conceptually described by three components. The first component is caused by the confinement of the ligand to the vicinity of the binding site and restricts the translational degrees of freedom. The second component can arise from a restriction of the rotational degrees of freedom of the ligand relative to the receptor that is compatible with the binding mode. The third component arises from the restriction of internal degrees of freedom in ligand and receptor due to geometric constraints that are imposed by the necessity to form specific contacts of receptor moieties with ligand moieties.

Figure 6c shows the stretching or force induced unfolding of a macromolecule. The end states of transformations in which the polymer changes its length will often have a different total configuration space volume.<sup>36</sup> There will be a smaller number of possible configurations in the folded state which is characterized by a compact shape (small length or end-to-end distance) and many defined intramolecular contacts, which further restricts the available configuration space. The complementary extreme is constituted by the fully elongated polymer, which has a very small total configuration space volume. In contrast to the folded and fully elongated states, there is a much larger number of configurations compatible with an intermediate elongation. Additionally, there can be restrictions of the configuration space imposed by the need to form the native intramolecular contacts which stabilize the folded state. A simplified model of the latter example is constituted by the freely jointed chain model of a macromolecular polymer examined in detail in ref 37.

**4.2. Efficiency Comparison of the Simulation Schemes and Possible Perspectives for Further Developments.** All the problems presented in the example section are solved analytically, that is, with perfect sampling giving exact results. The efficiency of a given simulation scheme is, however, a very important topic in actual free energy calculations from simulation data and requires careful consideration. The choice of the simulation scheme has practical consequences regarding the statistical uncertainty of a free energy estimate obtained from a given number of energy difference samples. The statistical

uncertainty of a single estimate determines the number of individual free energy estimates or the length of a simulation needed to reach a given target value of the standard deviation for the overall free energy estimate.

The equilibrated single-move scheme (eq 31) leads to much larger statistical uncertainties than the other simulation schemes. This is mainly due to the high statistical uncertainty of the correction term given by eq 36 which converges slowly and often in a very unsteady fashion. Because of these properties, the equilibrated single-move scheme is in the present form of little value for practical simulations. The true value of this simulation scheme might be less immediate as we will discuss in the end of this section. The choice between the two remaining simulation schemes has two aspects. On the one hand, the statistical uncertainty is smaller when collecting energy difference samples for all possible FEP moves in a simulation according to eq 13, because always the whole available information is used including the FEP moves with lowest energy difference that dominate the exponential average. On the other hand, the computational effort per sampling step is higher with the multimove scheme than with the random single-move scheme (eq 23) due to the higher number of energy difference samples that need to be calculated in each collection step. It is difficult to decide a priori which simulation scheme to choose, because the efficiency of the simulations schemes depends on the actual probability distributions of initial and final states; i.e., the relative efficiency varies with system, energy function, and transformation. For systems defined on a discrete configuration space, one is in principle free to choose one of the simulation schemes. However, for systems with a continuous configuration space it might be difficult to evaluate the integral over all possible FEP moves as demanded by the multimove simulation scheme eq 13.

Staging<sup>16</sup> and the Bennett acceptance ratio method<sup>11</sup> can be used to increase the efficiency of both simulation schemes as with traditional FEP.<sup>7,8,25</sup> The use of these two methods with our FEP simulation schemes is outlined in section B of the Supporting Information.

Although the equilibrated single-move scheme is of limited value for practical simulations because of its low efficiency, it is interesting in its own right from a theoretical point of view. In addition, comparison to the random single-move scheme might hint at possible perspectives for further developments of the FEP theory. The two single-move schemes use stochastic mappings that employ different probability distributions. It is tempting to ask whether other probability distributions could

be used to obtain FEP simulation schemes with optimized efficiency. It seems worthwhile to pursue this idea in future work.

## 5. Summary

We have presented a generalized FEP theory that extends the traditional FEP theory by some novel concepts and possibilities. Our formalism is inspired by Monte Carlo methods and previous interpretations of FEP in their terminology.<sup>11,12</sup> The novel formalism is applicable to a general class of transformations. Transformations in which the end states have a differing configuration space volume can be considered for systems defined on continuous or discrete configuration spaces.

The main results of this paper are three simulation schemes for free energy calculations with FEP. All simulation schemes partition the variables that define the position of the system in configuration space into two types to make the energy calculations more efficient. The first variable type has differing value ranges at the two end states (reactive subsystem) whereas the second variable type does not have these differing value ranges (environment subsystem). Each simulation scheme starts from microstates sampled from the equilibrium distribution of the initial state and collects energy differences for transitions or FEP moves to microstates of the final state. A move between the end states amounts to altering one or more variables of the reactive subsystem from values permitted in the initial state to values permitted in the final state. The domain of possible moves is defined by the ranges of values permissible for the variables forming the reactive subsystem of the final state. The way to select from the possible moves differs among the simulation schemes. The multimove FEP simulation scheme uses energy differences for all possible transitions and requires no correction. The two remaining simulation schemes use single FEP moves like traditional FEP based on the Zwanzig equation. The novelty in comparison with previous FEP methods<sup>10,14</sup> is that these simulation schemes are based on a stochastic instead of a deterministic and bijective mapping of the two end states. Traditional FEP maps each differential configuration space element of the initial state to exactly one configuration space element of the final state and vice versa. Our new approach selects the configuration space element of the final state to be mapped to a given configuration space element of the initial state based on a probability distribution. The first single-move scheme randomly selects the microstates from the final state for the FEP move. The free energy estimate obtained from this simulation scheme must be corrected for the relative total configuration space volume of the end states. The second single-move scheme selects the microstates according to the equilibrium distribution of the final state. The free energy estimate obtained from this simulation scheme must be corrected for the ratio of the effective configuration space volume of the final state and the total configuration space volume of the initial state.

We have discussed the relevance of our method to applications of general interest in the field of free energy calculations on biomolecular systems. Our theoretical findings are in line with established knowledge and provide a general, formal integration of this knowledge into FEP theory. Furthermore, we pointed out advantages of our simulation schemes in practical free energy calculations. Our discussion specifically addressed the calculation of binding free energies and free energies of polymer folding.

The novel formalism provides a generally applicable theoretical framework for FEP that promises to have potential for further developments and applications.

**Acknowledgment.** We thank the members of our group for helpful discussions on earlier versions of this paper. R.T.U. thanks the state of Bavaria for a doctoral fellowship according to the Bayerische Eliteförderungsgesetz. This work was supported by the Deutsche Forschungsgemeinschaft through SFB 840 research project B2.

**Supporting Information Available:** Detailed derivations and additional plots for the classical particle in the box models, a brief outline of how to use staging and the Bennett acceptance ratio method in the framework of our formalism, and a discussion of the relation between Boltzmann and Sussmann entropy. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492.
- (2) Kollmann, P. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (3) Rickman, J. M.; LeSar, R. *Annu. Rev. Mater. Res.* **2002**, *32*, 195–117.
- (4) Kofke, D. A. *Fluid Phase Equilib.* **2005**, *228–229*, 41–48.
- (5) Oostenbrink, C.; van Gunsteren, W. F. *Chem. Phys.* **2006**, *323*, 102–108.
- (6) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. In *Alchemical free energy calculations: Ready for prime time?*; Elsevier: Amsterdam, 2007; Vol. 3, Chapter 4, pp 41–59.
- (7) Chipot, C.; Pohorille, A. *Free Energy Calculations*; Springer: Berlin, 2007; Vol. 86.
- (8) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114*, 10235–10253.
- (9) Ullmann, R. T.; Essigke, T.; Ullmann, G. M. “GMCT Monte Carlo simulation software for macromolecular receptors with multiple ligands, membrane potential and conformational flexibility”, to be submitted for publication.
- (10) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (11) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (12) Voter, A. F. *J. Chem. Phys.* **1985**, *82*, 1890–1899.
- (13) Severance, D. L.; Essex, J. W.; Jorgensen, W. J. *Comput. Chem.* **1995**, *16*, 311–327.
- (14) Jarzynski, C. *Phys. Rev. E* **2002**, *65*, 046122.
- (15) (a) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 054103. (b) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *123*, 084109.
- (16) Valleau, J. P.; Card, D. N. *J. Chem. Phys.* **1972**, *57*, 5457–5462.
- (17) Hill, T. L. *Statistical Mechanics Principles and Selected Applications*; Dover Publ., Inc.: New York, 1987.
- (18) Zhou, H.-X.; Gilson, M. K. *Chem. Rev.* **2009**, *109*, 4092–4107.
- (19) (a) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693. (b) Jarzynski, C. *Phys. Rev. E* **1997**, *56*, 5018–5035.
- (20) Ferguson, D. M. *J. Chem. Phys.* **1993**, *99*, 10086–10087.
- (21) Süssmann, G. Z. *Naturforsch. A* **1997**, *52a*, 49–52.
- (22) (a) Jaynes, E. T. *Phys. Rev.* **1957**, *106*, 620–630. (b) Jaynes, E. T. *Phys. Rev.* **1957**, *108*, 171–190. (c) Jaynes, E. T. In *Information Theory and Statistical Mechanics*; W. A. Benjamin: New York, 1963; Vol. 3, pp 181–218.
- (23) Maruyama, K.; Nori, F.; Vedral, V. *Rev. Mod. Phys.* **2009**, *81*, 1–23.
- (24) Rényi, A. On measures of information and entropy. *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability 1960*, Vol. 1: Contributions to the Theory of Statistics, 1961; pp 547–561.
- (25) (a) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. *Phys. Rev. Lett.* **2003**, *91*, 140601. (b) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740–5761. (c) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.
- (26) From Figure 3b it can be seen that the probability distributions of the two ensembles have very little overlap. In terms of efficiency of a FEP simulation with the Zwanzig equation, however, it would have been much better to reverse the ordering (1 is mapped to 46, 2 to 45, . . .). This mapping would result in a much better overlap of the probability distributions, and consequently in a much smaller variance of the free energy estimate obtained from such a simulation.<sup>11,14,15,20</sup>
- (27) Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8*, 281–283.
- (28) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.
- (29) Luo, H.; Sharp, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10399–10404.

- (30) Nar, H.; Messerschmidt, A.; Huber, R.; van de Kamp, M.; Canters, G. W. *J. Mol. Biol.* **1991**, *221*, 765–772.
- (31) Jeuken, L. J. C.; Wisson, L.-J.; Armstrong, F. A. *Inorg. Chim. Acta* **2002**, *331*, 216–223.
- (32) Andrade, S. L. A.; Dickmanns, A.; Ficner, R.; Einsle, O. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14994–14999.
- (33) Ben-Naim, A. *Cooperativity and Regulation in Biochemical Processes*; Kluwer Academic/Plenum Publishers: New York, 2001.

- (34) Deng, Y.; Roux, B. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.
- (35) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707–2714.
- (36) Chan, H. S.; Dill, K. A. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 447–90.
- (37) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover Publ. Inc.: New York, 1960.
- JP1093838

## Chapter 6

### Manuscript C:

Coupling of Protonation, Reduction, and  
Conformational Change in Azurin from  
*Pseudomonas aeruginosa* Investigated  
with Free Energy Measures of  
Cooperativity

R. Thomas Ullmann and G. Matthias Ullmann  
J. Phys. Chem. B 2011, 115, 10346-10359  
doi: 10.1021/jp204644h

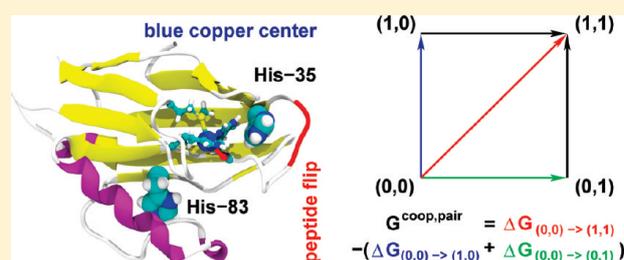
# Coupling of Protonation, Reduction, and Conformational Change in azurin from *Pseudomonas aeruginosa* Investigated with Free Energy Measures of Cooperativity

R. Thomas Ullmann\* and G. Matthias Ullmann\*

Structural Biology/Bioinformatics, University of Bayreuth, Universitätsstrasse 30, BGI, 95447 Bayreuth, Germany

**S** Supporting Information

**ABSTRACT:** We used free energy calculations within a continuum electrostatics model to analyze the coupling of protonation, reduction, and conformational change in azurin from *Pseudomonas aeruginosa* (PaAz). PaAz was characterized extensively with a variety of experimental methods. Experimentally determined  $pK_a$  values and pH-dependent reduction potentials are used to validate our computational model. It is well-known from experiment that the reduction of the copper center is coupled to the protonation of at least two titratable residues (His-35 and His-83) and to the flip of the peptide bond between Pro-36 and Gly-37. Free energy measures of cooperativity are used for a detailed analysis of the coupling between protonation, reduction, and conformational change in PaAz. The reduction of the copper center, the protonation of His-35, and peptide flip are shown to be cooperative. Our results show that cooperativity free energies are useful in detecting and quantifying thermodynamic coupling between events in biomolecular systems. The protonation of His-35 and the peptide flip are found to be so tightly coupled that these events happen effectively concerted. This concerted change results in a marked alteration of the electrostatic surface potential of azurin that might affect the interaction of azurin with its binding partners.



## 1. INTRODUCTION

Azurin from *Pseudomonas aeruginosa* (*P. aeruginosa*, PaAz) is a small electron-transport protein with a blue copper center.<sup>1,2</sup> The structure of PaAz is shown in Figure 1. The exact physiological role of PaAz and the identity of its electron-transfer partners are not yet known.<sup>3</sup> PaAz has been shown to be able to exchange electrons with cytochrome  $c_{551}$ , cytochrome  $cd_1$  nitrite reductase, and the cytochrome  $bc_1$  complex of *P. aeruginosa*.<sup>3,4</sup> Experiments with knockout mutants could not yet identify, however, an essential role for PaAz, at least under the conditions studied.<sup>3</sup> Electron transfer with all known possible redox partners can also be fulfilled by other electron carriers such as cytochrome  $c_{551}$ .<sup>3,4</sup> The understanding of the respiratory system of *P. aeruginosa* is also of medical interest, because this bacterium occurs as opportunistic pathogen in human diseases such as cystic fibrosis.<sup>5,6</sup> Such understanding could aid in the search for potent drugs against *P. aeruginosa*. The search for such drugs is especially important, because strains of the bacterium appeared among the multidrug-resistant pathogens that have become a major problem in medicine.<sup>5,6</sup>

In the past years, PaAz has found renewed interest as a potential therapeutic agent against cancer cells,<sup>7</sup> malaria pathogens, and HIV.<sup>8</sup> The basis for such applications is the ability of PaAz to induce apoptosis in mammalian cells.<sup>7</sup> This ability is independent of its redox activity<sup>9</sup> and depends on the interaction of PaAz with the mammalian tumor suppressor p53.<sup>7,10</sup> *P.*

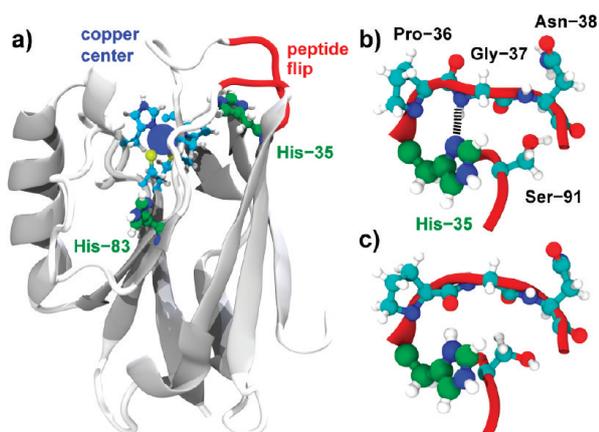
*aeruginosa* has been found to secrete PaAz and cytochrome  $c_{551}$  into the surrounding medium.<sup>7</sup> From there, these bacterial redox proteins are taken up by mammalian cells (preferentially cancer cells)<sup>11</sup> and induce apoptosis.<sup>7</sup> The antimalaria and anti-HIV activities of PaAz seem to be based on a prevention of pathogen entry into the mammalian cell.<sup>8</sup> For these reasons, it has been proposed that the cytotoxic capacity of PaAz and its antimalaria and anti-HIV activities may have evolved as a defense against predators or concurrents for the same nutrients in an infected host.<sup>7,8</sup>

PaAz was characterized extensively with a variety of experimental methods. Experimentally determined  $pK_a$  values and pH-dependent reduction potentials are compared to our results to validate our computational model of this system. It is well-known from experiment that the reduction of the copper center is coupled to the protonation of at least two titratable residues (His-35 and His-83) and to the flip of the peptide bond between Pro-36 and Gly-37.<sup>12–16</sup> Azurin from *Alcaligenes denitrificans* (*A. denitrificans*) has been used as a model system in an early continuum electrostatic study on the effect of metal oxidation on the  $pK_a$  values of protein residues.<sup>17</sup> We use azurin for a case study to show that free energy measures of cooperativity are useful for analyzing the coupling between events in biomolecular

Received: May 18, 2011

Revised: July 14, 2011

Published: July 20, 2011



**Figure 1.** Structure of azurin from *Pseudomonas aeruginosa* (PaAz). The coupling of reduction and protonation<sup>20</sup> in PaAz involves the flip of the peptide bond plane between Pro-36 and Gly-37 that occurs between the pH values 5.5 and 9.0.<sup>12</sup> This peptide flip is tightly coupled to the protonation of His-35.<sup>16</sup> The protonation of the residues His-35 and His-83 is believed to cause the pH-dependence of the reduction potential of the copper center.<sup>13–16,21</sup> (a) Overall structure: the locations of the copper center, His-35, His-83, and the peptide flip region are indicated. (b) Structure of the peptide flip region at high pH values (PDB 5AZU):<sup>12</sup> His-35 is deprotonated. (c) Structure of the peptide flip region at low pH values (PDB 4AZU):<sup>12</sup> His-35 is protonated. The conformationally variable residues included in modeling the peptide flip region are Pro-36, Gly-37, Asn-38, and Ser-91 from the adjacent loop (indicated in panel b) in line with findings from NMR experiments.<sup>22</sup> The figure was prepared with VMD<sup>23</sup> and Tachyon.<sup>24</sup>

systems. Cooperativity free energies can be used to detect and quantify this thermodynamic coupling.<sup>18,19</sup> The coupling of protonation, reduction, and conformational change in azurin is analyzed in detail. Implications of our findings for the physiological function of PaAz and its application as therapeutic agent are discussed.

The parts of this paper are organized as follows. The next section specifies the computational methods employed in our calculations. Our results are presented and discussed in section 3. We close with a summary of our findings and the conclusions thereof.

## 2. METHODS

The work flow of our computational methods begins with structure preparation with the program CHARMM. In this stage, hydrogen atoms are added to the crystal structure and their positions are energy minimized. Rotamer positions for hydroxyl hydrogens and alternative positions of protons on carboxyl groups are also added. Atomic radii are assigned to each atom.

The next stage is performed by an in-house program<sup>25</sup> based on our modified version of the MEAD library.<sup>26,27</sup> In this stage, side chain rotamer positions, and different binding and charge forms are generated. Microscopic intrinsic energies and interaction energies between the sites are computed on the basis of a continuum electrostatics/molecular mechanics model.<sup>19</sup>

The last stage constitutes the productive phase. Equilibrium properties of PaAz are computed from Monte Carlo (MC) simulation data.<sup>19</sup> The MC simulations are based on a microstate description of the system.<sup>19,28,29</sup> This description is a refined version of earlier models of protonation and reduction equilibria

in proteins and other macromolecules that were developed and evolved in our and other groups.<sup>27,30–33</sup> Details of the underlying theory and its computational implementation are given in ref 19. The microscopic intrinsic energies and interaction energies calculated in the previous stage can be used to compute the microstate energy of our model system during the simulations. A microstate is defined by a particular instance occupied by each site. The term instance describes the combination of a particular set of atomic partial charges, a particular number of bound ligands of each ligand type (here electrons and protons), and a particular set of atomic coordinates. The energy of a microstate is given by<sup>19</sup>

$$E_n^{\text{micro}} = \sum_{i=1}^{N_{\text{sites}}} (E_{i,k}^{\text{int}} - \sum_m v_{i,k,m} \bar{\mu}_m) + \sum_{i=1}^{N_{\text{sites}}} \sum_{j=1}^{i-1} W_{i,k,j,l} \quad (1)$$

where  $N_{\text{sites}}$  is the number of sites and  $\mathcal{L}$  is the number of ligand types.  $E_{i,k}^{\text{int}}$  is the intrinsic energy of site  $i$  in instance  $k$ . The stoichiometric coefficient  $v_{i,k,m}$  indicates the number of ligands of type  $m$  bound to instance  $k$  of site  $i$  and  $\bar{\mu}_m$  is the electrochemical potential of this ligand in the surrounding solution.  $W_{i,k,j,l}$  is the interaction energy of sites  $i$  and  $j$  in their respective instances  $k$  and  $l$ .

**Calculation of Parameters.** Partial charges for the T1 copper centers of azurins were obtained from density functional theory (DFT) calculations at a model system. The model system consisted of all side chain atoms of the coordinating residues His-46, Cys-112, His-117, and Met-121. For Gly-45, the coordinating carbonyl group was included. In addition, the backbone atoms connecting the coordinating residues His-46 and Gly-45 were included. The coordinates were taken from the PDB-entry 4AZU. Free valences were capped with hydrogen atoms. The DFT calculations were performed with the ADF program<sup>34</sup> (functionals VWN<sup>35</sup> and OPBE<sup>36</sup> with a TZ2P basis set). The atomic partial charges for the copper center were obtained with an in-house-modified version of the multipole-derived charge analysis method.<sup>37</sup> Our modifications allow one to include constraints in the charge fitting procedure, which were used to set the charge of link atoms to a value of zero.<sup>25</sup> The atoms included in the model of the copper site and their atomic partial charges are listed in Table S3 of the Supporting Information.

Four alternative hydrogen positions are constructed for protonated carboxylic acids representing syn and anti configurations for both protonatable oxygen atoms. The energy difference between syn and anti configurations was set to 2 kcal/mol on the basis of the upper bound for this value obtained from gas-phase DFT calculations with different functionals and basis sets. The geometric parameters for adding the dissociable hydrogen was adopted from the ideal gas phase geometry obtained for propionic acid with the PBE and VWN functionals. The DFT calculations were performed with the ADF program<sup>34</sup> (functionals VWN<sup>35</sup> and BP86<sup>38,39</sup> with a QZ4P basis set).

Atomic partial charges for forms of protonatable amino acids that were not available from the CHARMM force field were derived from DFT calculations. The DFT calculations were performed with the ADF program<sup>34</sup> (functionals VWN<sup>35</sup> and BP86<sup>38,39</sup> with a QZ4P basis set). Atomic partial charges for deprotonated arginine and lysine were fitted with the multipole-derived charge analysis.<sup>37</sup> The side chain analogues *n*-propylguanidine and *n*-butylamine were used for arginine and lysine, respectively. The intrinsic energies for the model compounds of deprotonated lysine were computed from the macroscopic

$pK_a$  value and a symmetry correction to account for the presence of three isoenergetic deprotonated forms.<sup>19,40</sup> The intrinsic energies for the model compounds of deprotonated arginine could not be simply computed from the macroscopic  $pK_a$  value of the model compounds, since the different tautomeric forms are not equivalent. The relative intrinsic energies of the tautomers were computed from their relative free energies in aqueous solution. The free energies of the model compounds were computed from DFT augmented with the electrostatic solvation. The DFT energy consisted of the single-point DFT energy and the vibrational energy in the gas phase. The electrostatic solvation energies were computed with the MEAD package.<sup>26,27</sup> The intrinsic energy difference relative to the protonated form was computed using the intrinsic energies of the deprotonated forms and the macroscopic  $pK_a$  value of arginine. The intrinsic model energies and atomic partial charges are listed in the Supporting Information.

**Structure Preparation and Continuum Electrostatics Calculations.** The structures of PaAz in the high-pH form (PDB code 5AZU chain A) and in the low-pH form (PDB code 4AZU chain A) were used as the basis for our calculations.<sup>12</sup> Hydrogen atoms were added with HBUILD<sup>41</sup> in CHARMM<sup>42</sup> and subsequently minimized using the CHARMM force field.<sup>43</sup> The peptide flip region (see Figure 1) was modeled as one site consisting of Pro-36, Gly-37, Asn-38, and Ser-91. These residues were found to be the only residues that undergo a significant conformational change between the high-pH and the low-pH crystal structures.<sup>12</sup> This finding is in line with findings from NMR experiments, where the other parts of the protein were found to be very rigid.<sup>22</sup> Also molecular dynamics (MD) simulations in which all parts of the protein except the Cu site (see above) were free to move confirmed this finding (data not shown).

Four conformers of the peptide flip region were generated from 1 ns MD simulations of PaAz with CHARMM in explicit solvent. The simulations were performed for the high-pH and low-pH forms with the copper center oxidized and reduced. All protonatable residues except His-35 and His-83 were set to standard protonation states. His-83 was fully protonated. His-35 was fully protonated for the low-pH form and singly protonated at NE2 for the high-pH form (see Figure 1). The protein was embedded in the cubic box of TIP3P water molecules with a length of 90 Å. Explicit potassium and chloride ions were added at a concentration of  $\approx 0.15$  mol/L. The initial coordinates of the water box and the ions were generated with VMD.<sup>23</sup> The simulation was run at standard temperature (298.15 K) and pressure (1 atm). The particle-mesh Ewald method was used for the electrostatics part of the energy function. During the simulation all parts of the protein apart from the peptide flip region, including the Cu site, were fixed to allow a reinsertion of the resulting conformers into the same structure. The conformational energy of the peptide flip region was estimated from the average dihedral energy plus the internal nonbonded interactions during a 1 ns MD simulation of PaAz with CHARMM<sup>42</sup> using the CHARMM force field.<sup>43</sup> The coordinates of the conformers were taken from the energy-minimized, average structure of each simulation. During the energy minimization, the coordinates of all other parts of the protein remained fixed and the dihedral angles of the peptide flip region were restrained with a harmonic potential using a force constant of 20.0 kcal/(mol rad<sup>2</sup>).

The intrinsic energies and interaction energies were computed with a continuum electrostatics/molecular mechanics model

within our in-house version<sup>25</sup> of the MEAD package.<sup>26,27</sup> The underlying theory is detailed in ref 19. All aspartate, histidine, glutamate, lysine, arginine, and tyrosine residues were considered as protonatable sites except for the copper ligands His-46 and His-117. The copper center was considered a redox-active site with an oxidized form and a reduced form. Side chain rotamers were added for flexible amino acid residues on the protein surface. The residues considered were Glu-2, Asp-6, Gln-8, Asn-18, Asp-23, Lys-24, Ser-25, Lys-27, Ser-34, Leu-39, Lys-41, Asp-55, Asp-62, Met-64, Asp-69, Lys-70, Asp-71, Lys-74, Asp-76, Asp-77, Arg-79, Lys-85, Glu-91, Lys-92, Asp-93, Asp-98, Lys-101, Lys-103, Glu-104, Glu-106, Gln-107, Lys-122, and Lys-128. Side chain rotamers were generated using the side chain dihedral angles from the backbone-dependent Squirrel rotamer library.<sup>44</sup> Conformational energies of side chain rotamers were computed using the CHARMM22 force field,<sup>43</sup> where we used dihedral energy terms, internal electrostatic energies within the site, and Lennard-Jones interaction energies within the site and between the site and the background (i.e., parts of the protein not belonging to any site). Lennard-Jones interactions were also added to the site–site interaction energies. The effect of the Lennard-Jones potential contribution to the site–site interaction energies is mainly the adding of a penalty for clashes between rotamers of flexible sites and other parts of the protein. The calculation of the electrostatic contributions to the intrinsic energies and interaction energies is described below.

We used a detailed charge model with explicit hydrogen positions for all protonatable sites. Atomic partial charges for standard forms of amino acid residues were taken from the CHARMM22 parameter set.<sup>43</sup> Atomic partial charges of nonstandard forms of amino acid residues and the copper center were obtained from DFT calculations (see above). Model compounds of protonatable amino acids include the entire residue plus the directly neighboring CHARMM charge groups belonging to the backbone of the preceding and succeeding amino acids to ensure charge neutrality and to mimic an *N*-formyl, *N*-methylamide blocked amino acid compounds.<sup>45</sup> The model compound for the copper center includes the copper ion, the backbone carbonyl group of Gly-45, and all side chain atoms of the ligating amino acids His-46, Cys-112, His-117, and Met-121. The intrinsic energies of the model compounds for protonatable amino acid residues in aqueous solution were calculated from  $pK_a$  values of appropriate model compounds taken from the literature<sup>32,46</sup> (see the Supporting Information). The intrinsic energies for the model compound of the copper center were fitted by matching the calculated midpoint potential of azurin from *A. denitrificans* to the experimental value of 278 mV.<sup>15</sup> The fitting was carried out on a static structure without side chain rotamers. The obtained intrinsic energy for the model compound in water is 13.3 kcal/mol for the oxidized form relative to the reduced form.

MEAD uses a finite-difference method on cubic grids to solve the linearized Poisson–Boltzmann equation. The dielectric constant of the protein was set to 4, and that of the solvent was set to 80.<sup>27</sup> The temperature was set to 298.15 K. The dielectric boundary between solute and solvent was calculated using a water probe sphere of 1.4 Å radius and the atomic radii (1.0 Å for H, 1.55 Å for N, 1.7 Å for C, 1.5 Å for O, 1.8 Å for S, and 1.3 Å for Cu). The ionic strength was set to 0.15 M. The thickness of the ion exclusion layer was set to 2.0 Å. Electrostatic potentials were computed using the focusing technique<sup>47</sup> with three cubic grids. The grids had grid spacings of 2.0, 0.5, and 0.2 Å, respectively.

The outer grid had a grid length of 101 points and was centered on the geometric center of the protein. The middle grid had a grid length of 201 points and was centered on the geometric center of the site. The inner grid had a grid length that was adjusted for each instance of each site separately to fit the dimensions of the site plus 15 Å in each direction and was centered on the geometric center of the site. The same grids were used for the model compound and the site in the protein.

In continuum electrostatics calculations with side chain rotamers, the dielectric boundary of the protein has to be approximated by assuming the presence of all rotameric forms of all sites other than the immediate site of interest at the same time.<sup>48</sup> As a consequence of the increased low-dielectric volume, electrostatic interaction energies are overestimated. To reduce the associated error, the number of side chain rotamers is reduced. First, all rotamers with large unfavorable Lennard-Jones interactions due to clashes with the background were removed. After a first calculation of the continuum electrostatics contributions to the intrinsic energies, the number of rotamers was further reduced with a modified version of the Goldstein criterion.<sup>49</sup> The modified Goldstein criterion uses an energy cutoff  $C$  to retain low-energy rotamers that could be thermally accessible. Each rotamer of site  $i$  can occur in multiple instances  $k$ . All rotamers of site  $i$  for which all instances  $k$  belonging to the rotamer fulfill

$$E_{i,k}^{\text{int}} - E_{i,k'}^{\text{int}} + \sum_{j \neq i} \min_l [W_{i,k,j,l} - W_{i,k',j,l}] > C \quad (2)$$

are removed. Here, the sum runs over the most favorable interaction energies  $W_{i,k,j,l}$  of site  $i$  in instance  $k$  with all other sites  $j$  in instance  $l$ . The reference instance  $k'$  of site  $i$  is determined via

$$k' = \arg \min_k [E_{i,k}^{\text{int}} + \sum_{j \neq i} \min_l [W_{i,k,j,l}]] \quad (3)$$

Multiple refinement cycles of the continuum electrostatics calculations were performed with an increasingly tight cutoff criterion. The final cutoff was set to  $C = 6.0$  kcal/mol.

**Monte Carlo Simulations.** All MC simulations were carried out with our program suite GMCT.<sup>19</sup> The temperature was set to 298.15 K. The interaction energy cutoff for pair and triplet moves were set to 0.5 kcal/mol and 1.0 kcal/mol, respectively.

Macroscopic binding properties were calculated with the Wang–Landau MC method.<sup>50</sup> A constant energy bin width of 0.1 kcal/mol was used for the energy histograms and the number of microstates. The initial modification factor  $f$  for the number of microstates was set to 2.0. The modification factor was decreased every 20 000 MC scans if the histogram flatness was improved or stayed equal relative to the last check. The modification factor was decreased according to the recursion formula  $f_{i+1} = f_i^{1/1.005}$  until it fell below a minimum value of 1.000 0005. The histogram flatness criterion was set to  $10^{-6}$ , to ensure that the simulation continues until the minimum modification factor is reached. This setting ensures good statistical convergence of the histograms for the macroscopic binding states. The obtained flatness of the overall histogram was always better than 0.025. Macroscopic properties were computed in terms of the intrinsic energies of the binding macrostates (see section A.1 of the Supporting Information).

Probabilities of binding states and conformations for individual sites were computed with the Metropolis MC method.<sup>51</sup> We used 500 MC scans for the equilibration and  $2 \times 10^5$  MC scans

for the production run. The statistical error of all probabilities was below  $10^{-3}$ .

Free energy calculations were performed with the free energy perturbation method<sup>52</sup> in our recently presented generalization.<sup>53</sup> We used staging<sup>54</sup> with two alchemical intermediate states evenly distributed along the transformation coordinate. The Bennett acceptance ratio method was used to minimize the statistical error of the free energy estimates.<sup>55</sup> Each free energy calculation consisted of multiple short simulations. The number of simulations was increased until the statistical error of the free energy estimate was smaller than 0.005 kcal/mol. Each separate simulation consisted of 500 MC scans for equilibration and 2000 to 10 000 MC scans for production. During production, microstate energy difference samples were collected after each MC scan according to the multimove simulation scheme.<sup>19,53</sup>

Events in a molecule can be described by reactions of sites or groups of sites in the molecule. The cooperativity between reactions can be measured with the cooperativity free energy  $G^{\text{coop.}(n)}$ , where the order  $n$  in the superscript denotes the number of reactions between which the cooperativity is measured. A negative cooperativity free energy indicates that the reactions favor each other; i.e., that they are cooperative. A positive cooperativity free energy indicates that the reactions disfavor each other, i.e., that they are anticooperative. If the cooperativity free energy is equal to zero, the reactions do not influence each other. Cooperativity free energies<sup>18,19</sup> can be computed according to the thermodynamic schemes in Figure 2. The cooperativity free energies between a pair and a triplet of reactions are given by

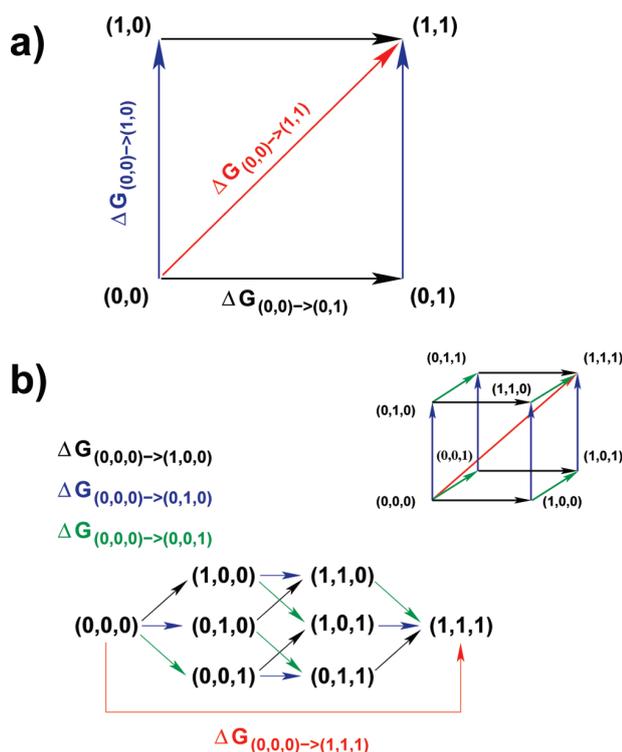
$$G^{\text{coop.}(2)} = G_{(0,0) \rightarrow (1,1)} - (G_{(0,0) \rightarrow (1,0)} + G_{(0,0) \rightarrow (0,1)}) \quad (4)$$

$$G^{\text{coop.}(3)} = G_{(0,0,0) \rightarrow (1,1,1)} - (G_{(0,0,0) \rightarrow (1,0,0)} + G_{(0,0,0) \rightarrow (0,1,0)} + G_{(0,0,0) \rightarrow (0,0,1)}) \quad (5)$$

The reaction free energies occurring in eqs 4 and 5 are defined and indicated in the thermodynamic schemes in Figure 2. Thereby, the direction in which the reaction proceeds is important and determined by defining the initial and final states of the reaction accordingly. A useful property of the pair cooperativity free energy is that it changes its sign but preserves its absolute value, if the direction of one of the reactions is reversed. We wanted to determine to what extent the cooperativity free energy is influenced by sites that are not directly involved in the separate reactions defining the cooperativity free energy. We quantified the influence of these sites with the difference between the cooperativity free energy calculated for the whole system and the cooperativity free energy computed in the absence of other sites. The absence of these sites was mimicked by setting their interaction energies with the sites immediately involved to zero.

### 3. RESULTS AND DISCUSSION

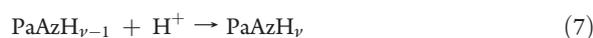
**Coupling of Overall Protonation and Reduction.** We began the characterization of PaAz by studying the coupling of its protonation and reduction from an overall, i.e., macroscopic, viewpoint. Here, macroscopic means that the total numbers of protons and electrons bound to PaAz are considered without regard to individual binding sites. Parts a and b of Figure 3 show the effect of reduction on the macroscopic protonation probabilities of PaAz as a function of the pH value. It can be seen that



**Figure 2.** Thermodynamic schemes for the calculation of the free energy of cooperativity between (a) a pair and (b) a triplet of reactions. The states and free energy differences that occur in the calculation are indicated. The states of the system are denoted by a pair or a triplet of numbers in braces. The first number stands for the first reaction, the second number for the second reaction, and the third number for the third reaction (in case of the triplet cooperativity). The number 0 indicates the initial state of a reaction, and the number 1 indicates the final state of a reaction.

the greatest effect of reduction occurs in the intermediate range of values between pH 6.0 and 8.0. The average number of protons taken up by PaAz upon reduction is plotted in Figure 3c in dependence on the pH value. It can be seen that the average number of protons taken up is smaller than 1 for the whole range of pH values. The proton uptake reaches a maximum in the intermediate range of pH values in agreement with panels a and b of Figure 3 and experimental findings.<sup>56–58</sup>

The coupling between the protonation and the reduction reactions of PaAz can be quantified with cooperativity free energies. These cooperativity free energies can be interpreted as effective interaction energies between the electron taken up in the reduction reaction and the protons bound in the protonation reaction. Figure 3d shows the cooperativity free energy between the reduction reaction and stepwise protonation reactions of PaAz. The two reactions can be formally written as



where  $\nu$  indicates the number of protons bound at the final state of the  $\nu$ th protonation step. From Figure 3d, it can be seen that the cooperativity between the stepwise protonation reactions is highest for the first protonation step. The 19th and 20th

protonation steps are the only other protonation steps with cooperativity free energies that are significantly greater than the thermal energy. The high magnitude of the cooperativity free energy for these protonation steps is mainly based on the involvement of the two protonation steps of His-35 which is in close distance to the copper center. The total effective interaction energy between the electron and the bound protons in reduced PaAz can be quantified with the cooperativity free energy between the reduction reaction and the overall protonation reaction. The reaction equation for the reduction reaction is given by eq 6. The reaction equation for binding  $\nu$  protons is given by

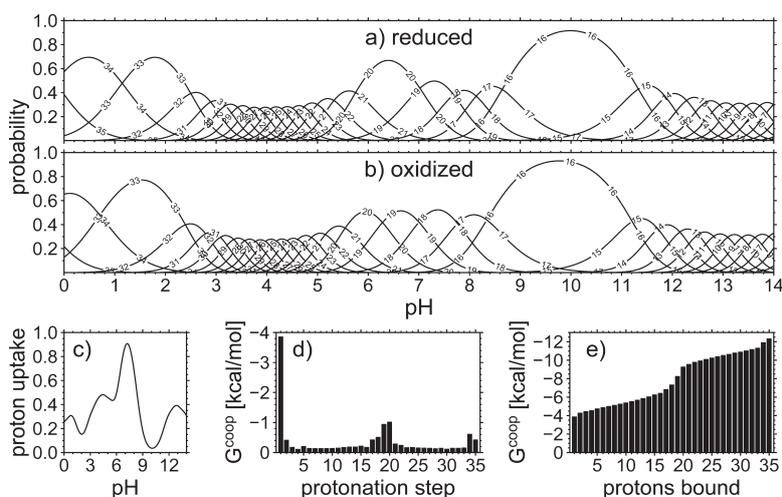


The total effective interaction energy between the electron and the protons in reduced PaAz is plotted in Figure 3e as a function of the total number of protons bound. In agreement with Figure 3d, it can be seen that about one-third of the interaction is already gained in the first protonation step. Furthermore, it can be observed that the rise in the effective interaction per protonation step is almost constant for the remaining steps except for the 15th to 20th and the last two protonation steps.

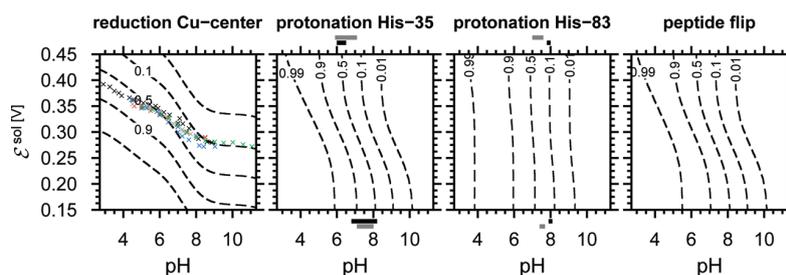
**Titration Behavior of Individual Sites.** Following the investigation of the macroscopic binding behavior of PaAz, we studied its behavior from a microscopic viewpoint by analyzing the titration behavior of individual sites. We started with a validation of our computational method by comparing our computed titration curves to available experimental data. Then we performed a detailed analysis of the binding equilibria and conformational equilibria of individual sites using free energy calculations. We concentrate the discussion on the copper center, the peptide flip region, His-35, and His-83, because these sites are involved in the coupling of protonation reduction and conformational change.<sup>12–16</sup> His-35 and His-83 are the only protonatable sites that interact strongly with the copper center and titrate in the physiologically relevant neutral range of pH values.

We correlated the computed titration behavior of important sites with experimentally determined midpoint reduction potentials ( $\varepsilon_{1/2}$ ) and  $\text{pK}_a$  values ( $\text{pK}_{1/2}$ ). The comparison of experimental and theoretical data is summarized in Figure 4. The calculated midpoint reduction potentials of the copper center agree well with the experimental data except for very alkaline pH values where the deviation is larger. Interestingly, more recent voltammetry experiments on PaAz immobilized on electrode surfaces showed a larger decrease in reduction potential at alkaline pH values in agreement with our results.<sup>21,58</sup> We did not use these data for comparison because it is not possible to quantify the influence of the electrode surface on the midpoint potentials.

The overall dependence of the reduction on the pH value of the solution is somewhat overestimated in our calculation, as judged from the steeper slope of the calculated curve of midpoint potentials relative to the experimental data. The overestimation of couplings by continuum electrostatic models originates in most cases from larger conformational arrangements that are not explicitly represented in the model.<sup>27,67</sup> The overestimated coupling can most likely be explained by the absence of explicit backbone flexibility in our model. Such flexibility has been reported for the loop regions of PaAz on the basis of NMR experiments.<sup>16,22</sup> An additional factor within the continuum electrostatic approach that can contribute to the overestimation



**Figure 3.** Macroscopic binding properties of PaAz. Top: Macroscopic protonation probabilities for (a) reduced and (b) oxidized PaAz as a function of the pH value. The numbers inscribed in the curves indicate the number of protons bound by the respective protonation macrostate. Bottom: (c) Average number of protons taken up upon reduction as a function of the pH value. (d) Macroscopic cooperativity free energy between protonation and reduction for each macroscopic protonation step (gain of effective interaction energy between the protons and the electron in the macroscopic binding reaction). (e) Total macroscopic cooperativity free energy as a function of the number of bound protons (effective interaction energy between the protons and the electron).

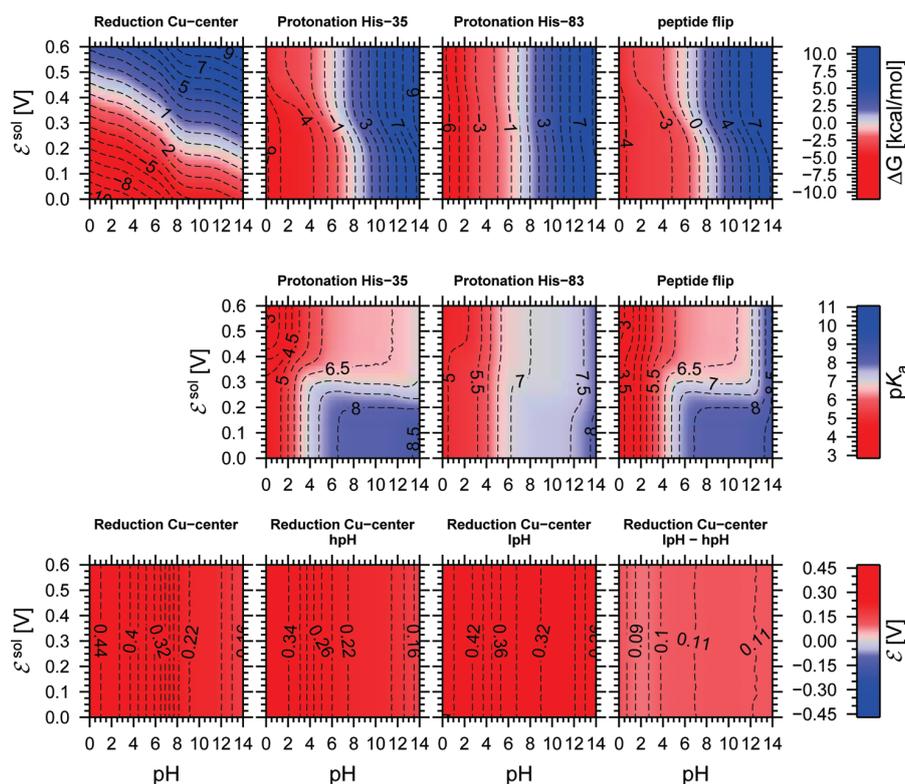


**Figure 4.** Validation of our model by comparison of computed titration curves with experimental midpoint reduction potentials  $\epsilon_{1/2}$  and  $pK_{1/2}$  values. The reduction probability of the copper center, the protonation probabilities of His-35 and His-83, and the occupation probability of the low-pH form of the peptide flip region are plotted as functions of the pH value and reduction potential of the solution  $\epsilon^{\text{sol}}$ . The probabilities are indicated by labeled isocontours. Experimental  $\epsilon_{1/2}$  and  $pK_{1/2}$  values should ideally coincide with the regions where the calculated reduction and protonation probabilities adopt a value of 0.5. Experimental midpoint reduction potentials are plotted in the leftmost panel as colored crosses. The corresponding data are taken from the following references: black, 59; gray, 15; red, 60; green, 14; dark blue, 61 and 62; light blue, 63. The ranges of experimentally determined  $pK_{1/2}$  values from direct measurements with NMR are shown as black bars.<sup>14,16,64–66</sup> The ranges of  $pK_{1/2}$  values that are indirectly determined from fits to kinetic and/or voltammetry experiments are shown as gray bars.<sup>20,21,56,58</sup> The bars above the plots (high  $\epsilon^{\text{sol}}$ ) correspond to oxidized PaAz, whereas the bars below the plots (low  $\epsilon^{\text{sol}}$ ) correspond to reduced PaAz. Titration curves of all residues of PaAz and pictures of PaAz indicating their spatial positions can be found in section A.2 of the Supporting Information.

of interaction energies lies in the necessity to assume the presence of multiple conformers of the sites in the definition of the dielectric boundaries,<sup>48</sup> where there is no formally exact and theoretically satisfactory method to correct for this effect. Empirical corrections proposed in the literature<sup>48</sup> were not employed in this study, but the effect was minimized by considering only surface-exposed, mobile residues as flexible and by removing energetically unfavorable side chain rotamers from the calculations (see Methods).

The titration curves of the His-35 and His-83 are depicted in the two central plots of Figure 4. In the following, protonation of a histidine side chain refers to the binding of a proton to one of the singly protonated forms of the histidine side chain to form the doubly protonated imidazolium form. For both histidine residues, the fully deprotonated imidazolate form is included in the

calculation but does not have a significant occupation probability in the studied pH range. The singly protonated tautomers that are populated by His-35 and His-83 in our simulations are consistent with experimental findings. The population of the  $N_{\delta}$  tautomer by the deprotonated His-83 and the  $N_{\epsilon}$  tautomer by the deprotonated His-35 comply with results from NMR experiments.<sup>65</sup> The population of the  $N_{\epsilon}$  tautomer by the deprotonated His-35 is also clearly evident from the available structural information at a pH value of 9.0<sup>12</sup> (see Figure 1b). The  $pK_{1/2}$  values of His-35 and His-83 are consistent with direct NMR measurements<sup>14,64–66</sup> and indirect fits to kinetic<sup>20,56</sup> and voltammetry data.<sup>21,58</sup> Even  $pK_{1/2}$  values that are measured directly with NMR are afflicted with a typical methodological uncertainty of  $\approx 1$  pH unit.<sup>56,68</sup> In some individual cases the uncertainty can reach higher values.<sup>56,68</sup> This uncertainty can



**Figure 5.** Reaction free energies for reactions that are involved in the coupling of protonation, reduction and conformational change in PaAz. Top row: The reaction free energies are plotted as functions of the pH value and the reduction potential of the solution. The reaction free energy is color coded (see color bar). Contours are plotted in constant intervals of 1.0 kcal/mol. Contour values are given in kilocalories per mole. Middle row: The reaction free energies are replotted from the top row as thermodynamically defined Henderson–Hasselbalch  $pK_a$  values calculated from  $pK_a = \text{pH} - \beta\Delta G^{\text{prot}}/(\ln 10)$ .<sup>70,71</sup> The tight coupling between the protonation of His-35 and the peptide flip leads to an apparent  $pK_a$  value of the peptide flip that is over wide ranges of solution pH and reduction potential very similar to the  $pK_a$  value for protonating His-35. The  $pK_a$  value is color coded (see color bar). Contours are plotted in constant intervals of 0.5 pH units. Contour values are given in pH units. Bottom row: The reduction free energy of the copper center is plotted as thermodynamically defined Nernst reduction potential calculated from  $\varepsilon = \varepsilon^{\text{sol}} - \Delta G^{\text{red}}/F$ .<sup>70</sup> The leftmost plot shows the reduction potential for the unconstrained system. The two central plots show the reduction potential for PaAz constrained to either the hpH or the lpH form. The rightmost plot shows the difference between the reduction potentials of PaAz in the lpH and hpH forms. Contours are plotted in constant intervals of 10 mV for the rightmost plot and in constant intervals of 20 mV for the other plots. Contour values are given in volts. The corresponding plots for other sites in PaAz can be found in section A.3 of the Supporting Information.

originate from interactions of the chosen reporter nucleus with other sites in the vicinity of the monitored site.<sup>68</sup> The indirectly determined experimental  $pK_{1/2}$  values slightly deviate from the NMR measured values. The overall agreement of our calculated  $pK_{1/2}$  values with the NMR data is better than with the indirectly determined data. The indirectly determined values rely on a fit of experimental data to mathematical models that assume the coupling of the reduction to one (His-35) or two (His-35 and His-83) protonatable sites. We show below that there are more than two protonatable sites that contribute to the coupling of protonation and reduction in PaAz. Thus, the NMR data are to be preferred as reference over the indirectly determined data. In agreement with both kinds of experiment, the protonation probability of His-35 is much more dependent on the reduction of the copper center than the protonation of His-83.

The rightmost plot of Figure 4 shows the occupation probability of the low-pH form of the peptide flip region as a function of the pH value and the reduction potential of the solution. We found that the conformational change of the peptide flip region depends on both parameters and thus on the protonation and reduction of PaAz. This finding is consistent with experimental

findings.<sup>12,16,69</sup> The titration curve of the peptide flip region at low reduction potentials has a midpoint at a value of about pH 8.0. This value is consistent with a value between pH 5.5 and 9.0 implied by the structural data<sup>12</sup> (see Figure 1). Our value is slightly higher than the value of 7.0, which can be deduced from the experimental equilibrium constant of  $\approx 1$  for reduced PaAz at pH 7.0 inferred from kinetic experiments.<sup>69</sup> Given the uncertainty in the experimental data, and the simplicity of the continuum electrostatic model, the agreement of our calculations with experiment is satisfactory.

The top row of Figure 5 shows plots of the reaction free energies corresponding to the titration curves in Figure 4. The reduction free energy of the copper center, the protonation free energies of His-35 and His-83, and the conformational free energy of the peptide flip show a complex dependency on the pH value and the reduction potential of the solution. In support of our previous findings,<sup>70–72</sup> it can be seen that the titration behavior of sites in a protein differs markedly from that of isolated model compounds in solution. In effect, the thermodynamics of protonation and reduction of sites in a protein cannot be described by ascribing a single  $pK_a$  value<sup>71,72</sup> or reduction

potential to these sites.<sup>70</sup> The coupled titration behavior of these sites is caused by the electrostatic interactions between sites in a protein.<sup>72</sup> These interactions can act over considerable distances especially if the interacting sites are charged. The coupling between the sites is analyzed in detail as follows.

The close similarity of the shapes of the titration and free energy curves for the protonation of His-35 and the peptide flip points to a tight coupling between the two reactions. This finding is in agreement with experimental findings.<sup>16,21</sup> The favorable effective interaction is caused by the complementary hydrogen-bonding properties of the peptide flip region and the side chain of His-35, where each of the bonding partners can either be a hydrogen bond acceptor or donor. The strong favorable microscopic interactions between pairs formed from a hydrogen bond donor and a hydrogen bond acceptor and the strong unfavorable microscopic interactions between pairs of two hydrogen bond donors or two hydrogen bond acceptors lead to the strongly coupled titration behavior of the sites. This coupling is so strong that PaAz occurs almost exclusively in either a high-pH form or a low-pH form. In the high-pH form (hpH), His-35 is deprotonated and the peptide flip region is in the high-pH conformation seen in PDB 5AZU<sup>12</sup> (see Figure 1b). In the low-pH form (lpH), His-35 is protonated and the peptide flip region is in the low-pH conformation seen in PDB 4AZU<sup>12</sup> (see Figure 1c). See Figure S6 of the Supporting Information for plots of the occupation probabilities of the two forms and of the residual probability.

The reduction probability of the copper center in the neutral pH range is strongly dependent on the pH value of the solution. Similarly, the protonation probabilities of His-35 and His-83 are dependent on the reduction potential of the solution in the neutral pH range. The stronger dependency in the case of His-35 can be rationalized from the strong electrostatic repulsion between the oxidized copper center and the protonated side chain of the nearby His-35. The much weaker dependence in the case of His-83 can be rationalized from its more remote location and consequently weaker electrostatic interaction with the copper center.

The protonation free energy can be expressed and plotted in terms of a thermodynamically defined Henderson–Hasselbalch  $pK_a$  value<sup>70,71</sup> (see the middle row of Figure 5). This representation of the protonation free energy makes especially clear that both sites exhibit a  $pK_a$  value that depends on both the pH value and the reduction potential of the solution. Also the much stronger dependence of the protonation of His-35 relative to the protonation of His-83 on the reduction potential of the solution is especially clearly seen in this representation of the protonation free energy.

Analogously, the reduction free energy of the copper center can be written in terms of a thermodynamically defined Nernst reduction potential<sup>70</sup> (see the bottom row of Figure 5). It can be seen that the reduction potential of the copper center does not depend on the reduction potential of the solution, as can be expected in the absence of other redox-active sites.<sup>70</sup> Furthermore, it can be seen that the reduction potential of the copper center depends on the pH value of the solution. This dependency results from the interaction with the surrounding protonatable sites.

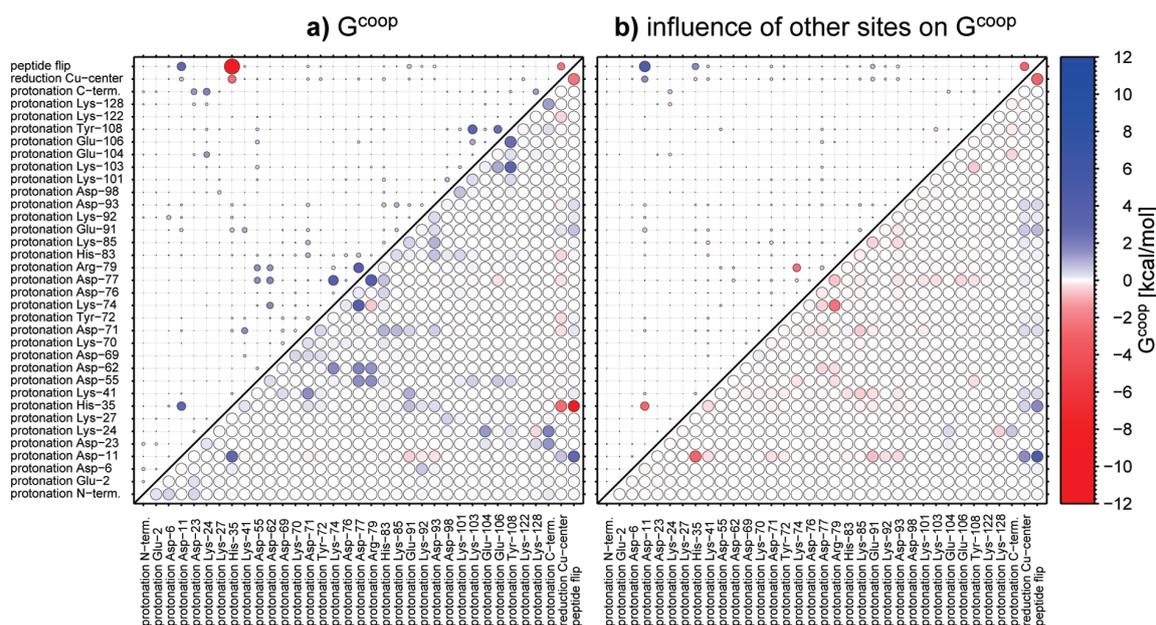
**Coupling between Protonation, Reduction, and Conformational Change on the Level of Individual Sites.** Following the investigation of the titration behavior of the individual sites, we analyzed the coupling between their reactions. Cooperativity free energies between reactions can be used to analyze their

coupling and provide a quantitative measure of the coupling strength.<sup>18,19</sup> The cooperativity free energy has a straightforward interpretation as effective relative interaction energy between the reacting sites.<sup>18,19</sup> The cooperativity free energy between a pair of events is conceptually identical to the effective interaction energy measured in a double mutant cycle experiment.<sup>73,74</sup> For our analysis, we considered all reactions of individual sites in PaAz that could be involved in the coupling between protonation, reduction, and conformational change. We considered the protonation reactions of all residues, the reduction reaction of the copper center, and the conformational change of the peptide flip region from the high-pH form to the low-pH form. The cooperativity free energies between all pairs of considered reactions in PaAz are shown in Figure 6a for a pH value of 7.0.

Not surprisingly, it can be seen that there is significant anticooperativity for most pairs of protonation reactions of sites that are in close distance, due to the electrostatic repulsion between the proton charges. A decrease of the effective repulsion between some pairs of protonation reactions can arise from the release of protons from nearby third sites.<sup>19,75,76</sup> This effect can even lead to an apparent attractive interaction between protons bound to pairs of interacting protonatable sites as in the case of His-35 and Asp-11 or Lys-74 and Arg-79. Also the favorable effective interaction between the protonation of almost all sites and the reduction of the copper center could be expected on the basis of the attractive electrostatic interaction between the complementary charges of the ligands. It can be seen that apart from His-35 and His-83 many other residues contribute to the cooperativity of protonation and reduction. The main contribution comes, however, from His-35, which is consistent with its location close to the copper center (see Figure 1) and earlier work.<sup>56,57,77</sup> The fit of experimental data to mathematical models that assume only one or two protonatable sites which are coupled to the reduction reaction implicitly include the contributions of all the other sites. Thus,  $pK_a$  values of the histidine residues inferred from such fits necessarily deviate somewhat from their true  $pK_a$  values. This finding explains the disagreement between  $pK_{1/2}$  values from direct measurements and indirect fits and why our calculated  $pK_{1/2}$  values agree better with direct NMR measurements.

The influence of third sites on the cooperativity free energy can be estimated by taking the difference between the cooperativity free energies calculated for the whole system and calculated in the absence of interactions with third sites (see Methods). The pairwise cooperativities between the isolated sites are equivalent to the relative microscopic interaction energies used in traditional two-state binding models (see, e.g., refs 45 and 32). The influence of the other sites on the cooperativity free energy can be seen to dampen the magnitude of the cooperativity free energies relative to the cooperativity free energy between the isolated sites in most, but not all, cases. Especially the cooperativity free energy between the peptide flip and the reduction of the copper center is increased by the surrounding sites. This effect will be analyzed in more detail below. The cooperativity free energies of the reduction of the copper center and the protonation of all protonatable sites is decreased by the influence of the surrounding sites. This dampening of the effective coupling between the reducible site and the protonatable sites contributes to the relatively low proton uptake upon reduction (see Figure 3c) and the overall pH dependence of the reduction potential of PaAz.

From Figure 6a, it can be seen that the coupling between protonation, reduction, and conformational change in PaAz is



**Figure 6.** Pairwise cooperativity free energies between reactions of sites in PaAz at pH 7.0 and  $\epsilon^{\text{sol}} = 0$ . The reactions are indicated by the labels on the plot axes. The symbols are colored according to the magnitude of the cooperativity free energy (see color bar). For easier spotting of cooperativities with a large magnitude, the symbols in the upper left triangles are scaled such that their area is proportional to the magnitude of the cooperativity free energy. See Figures S10–S12 of the Supporting Information for numeric values. (a) Cooperativity free energy for the whole system. (b) Difference between the cooperativity free energies for the whole system and for the isolated sites (all interactions with other sites set to zero) as a measure of the influence of surrounding sites on the cooperativity free energy.

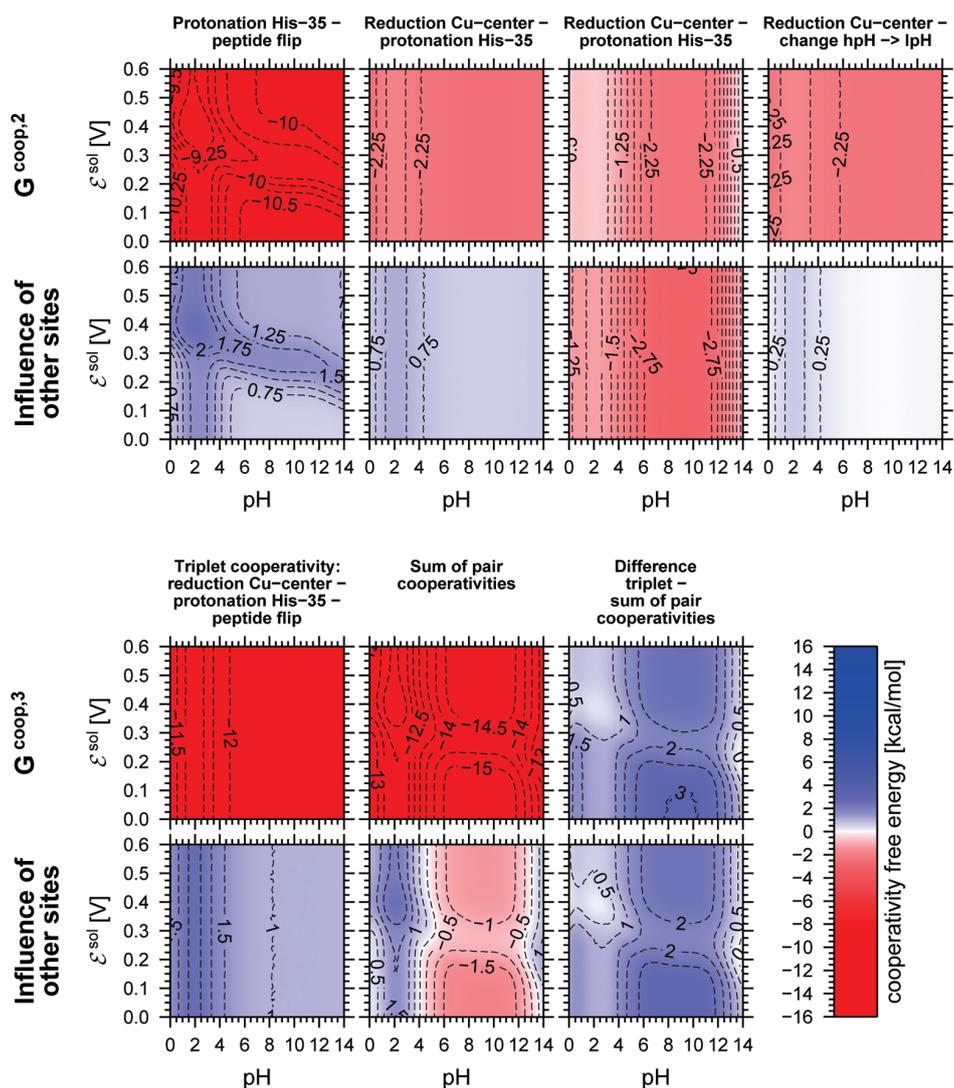
almost entirely contributed by only three sites. These sites are the copper center, His-35, and the peptide flip region. In the following, we concentrate on the cooperativity between the reactions of these sites for a more detailed analysis.

The cooperativity free energies between all possible pairs and the triplet formed from the reduction of the copper center, the protonation of His-35, and the peptide flip are plotted as functions of the pH value and the reduction potential of the solution in Figure 7. It can be seen that the pair cooperativity free energies do not depend strongly on the pH value and the reduction potential of the solution, but are instead nearly constant over wide regions of the parameter space. A similar behavior was also found for the other pair cooperativity free energies in Figure 6 (see Figure S13 of the Supporting Information).

The similarity of the titration behavior of His-35 and the peptide flip region already suggested a strong coupling between the protonation of His-35 and the conformational change of the peptide flip region (see Figure 4 and Figure 5). This suggestion is corroborated by the high magnitude of the cooperativity free energy between the two reactions over the whole range of ligand chemical potentials investigated (see Figure 7). Above, we also demonstrated a strongly pH-dependent titration behavior of the copper center, which points to the presence of at least one strongly interacting protonatable site. The cooperativity free energy between the protonation of His-35 and the reduction of the copper center adopts a nearly constant value of  $-2.3$  kcal/mol for pH values between 4.0 and 14.0 (see Figure 7). This effective interaction energy is responsible for the upshift of about 100 mV in the reduction potential of the copper center upon conversion from the high-pH to the low-pH form (see Figure 5, bottom right).

As already seen from Figure 6, the positive cooperativity between the reduction of the copper center and the peptide flip region is not caused by the direct microscopic interaction between the copper center and the peptide flip region. Taken separately, the electrostatic interaction between the two sites would even lead to weak anticooperativity between their reactions. The basis for the positive cooperativity between the two reactions are mediating, mainly electrostatic interactions of His-35 with the copper center and the peptide flip region. This can be demonstrated by computing the cooperativity free energy between the reduction of the copper center and a concerted reaction consisting of the protonation of His-35 and the peptide flip. We showed above that the coupling between these two reactions is so tight that PaAz is essentially always found in one of two forms under equilibrium conditions. The label  $\text{hpH} \rightarrow \text{lpH}$  above the plots in the upper right plots of Figure 7 indicates this concerted reaction that converts PaAz from the high-pH form to the low-pH form. It can be seen that there is nearly no influence of other sites on the cooperativity free energy between the reduction of the copper center and the concerted reaction. This finding proves that the positive cooperativity between the reduction of the copper center and the peptide flip is mediated by electrostatic interactions of both sites with His-35. This arrangement can be symbolized by the linear interaction scheme  $\text{BCC-His-35-PEP}$ , where BCC stands for blue copper center and PEP for the peptide flip region and a long hyphen symbolizes a strong coupling. That is, the interaction of the peptide flip region with His-35 is relayed to the copper center. Likewise, the interaction of the copper center with His-35 is relayed to the peptide flip region.

The bottom half of Figure 7 shows the triplet cooperativity free energy between the reduction of the copper center, the protonation of His-35, and the peptide flip in comparison to the

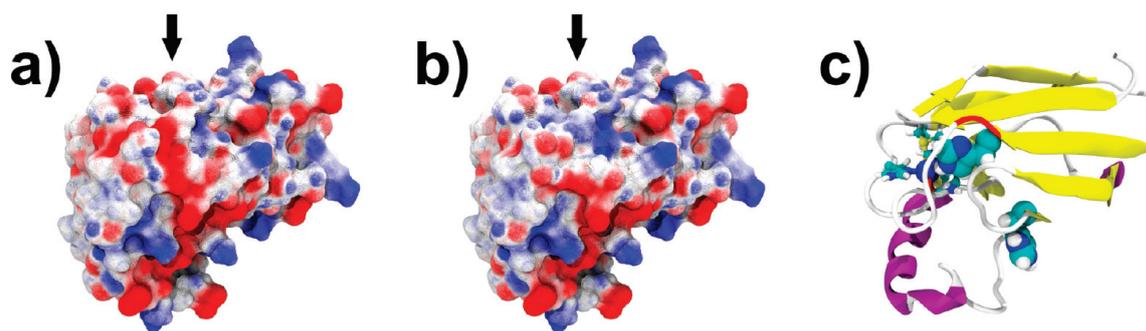


**Figure 7.** Cooperativity free energies between reactions involved in the coupling of protonation, reduction and conformational change in PaAz plotted as functions of the pH value and the reduction potential of the solution. Top: Pair cooperativity free energies between the reactions that are indicated above the curves. Bottom: The triplet cooperativity free energy between the reactions considered in the first three columns of the above block in comparison to the sum of the corresponding pair cooperativity free energies. The upper rows of the bottom and top blocks show the cooperativity free energy calculated for the whole system. The plots in the lower rows show the influence of the other sites on the cooperativity free energy (see Methods). The cooperativity free energy is color coded (see color bar). The upper and lower blocks are contoured in constant intervals of 0.25 and 0.5 kcal/mol, respectively. Contour values are given in kilocalories per mole. The corresponding plots for other sites in PaAz can be found in section A.4 of the Supporting Information.

sum of the corresponding pair cooperativity free energies. It can be seen that the triplet cooperativity free energy is nearly constant for the greatest part of the chemical potential space, except for a region of low variability in the acidic range of pH values ( $\text{pH} < 5$ ). Furthermore, the triplet cooperativity free energy is not equal to the sum of the pair cooperativity free energies but smaller in magnitude. Such a nonadditive behavior was also found in other systems.<sup>18</sup> Thus, the statistical mechanics of a multisite system can give rise to apparent many-body interactions even if the underlying microscopic energy function contains only pairwise interaction energies.<sup>18</sup> Comparison of the plots on the bottom right of Figure 7 shows that the deviation is dominated by the contributions of the surrounding sites. Thus, the reduced magnitude of

the triplet cooperativity is caused by the relaxation of degrees of freedom of other sites than those directly involved. This finding highlights the importance of using a detailed model that accounts for all protonatable sites, even though the individual contributions of most sites are small.

**Implications of Our Findings for the Physiological Role and Therapeutic Applications of PaAz.** The coupling of protonation reduction and conformational change in PaAz could be important for the physiological role of PaAz and for its role as a therapeutic agent. The physiological electron-transfer partners of PaAz are not known with certainty. Therefore, our discussion has to consider all possible factors that are known to affect electron-transfer reactions.



**Figure 8.** Electrostatic potential on the solvent accessible surface of reduced PaAz in (a) the high-pH form and (b) the low-pH form at pH = 7.0. The arrows at the plots indicate a region of variable electrostatic potential at the right rim of the hydrophobic patch above the peptide flip region. This region is termed His-35 patch and was suspected earlier to be involved in the mediation of ET. The color scale for the electrostatic potential reaches from red at  $-77$  mV to blue at  $+77$  mV ( $\pm 3k_B T/e^\circ$ ). The structures of PaAz were constructed by setting all sites to their most highly populated instances within the respective state. There are no significant conformational differences between the states apart from those in the peptide flip region. The electrostatic potentials and solvent accessible surfaces were computed with our in-house version<sup>25</sup> of the MEAD package.<sup>26,27</sup> (c) Structure of PaAz in the same orientation as in panels a and b for orientation. The backbone of Pro-36 and Gly-37 between which the peptide flip occurs is colored red. His-35, His-83, and the copper center are highlighted (cf. Figure 1). The figure was prepared with VMD<sup>23</sup> and Tachyon.<sup>24</sup> See Figure S15 of the Supporting Information for alternative views and equivalent plots for oxidized PaAz.

Electron-transfer reactions between proteins consist of several stages. The electron-transfer partners approach each other in a structurally heterogeneous encounter complex, followed by the formation of a structurally more defined productive complex.<sup>78</sup> Such a two-step binding scheme is believed to apply to all protein–protein binding reactions.<sup>78</sup> The formation of the productive complex can also involve rearrangements in the internal conformation of the binding partners.<sup>79,80</sup> ET reactions whose rate is affected by the steps that precede the actual electron-transfer step are termed coupled or gated.<sup>80</sup> Each of the separate steps can influence the rate for the overall electron-transfer reaction. In contrast, the thermodynamics of the electron-transfer reaction is solely determined by the difference in the reduction potentials of the electron-transfer partners.

We found that the conversion of PaAz from the high-pH form to the low-pH form leads to an increase in its reduction potential of about 100 mV (see Figure 5). This increase in reduction potential has been suggested before as a potential way of down-regulating the electron-transfer activity of PaAz if the pH value in the periplasm becomes too low.<sup>16,77</sup> The change in reduction potential can affect the rate of the actual electron-transfer step and the overall thermodynamics of an electron-transfer reaction. Furthermore, this conversion results in a marked change of the electrostatic potential at the surface of PaAz within the so-called His-35 patch<sup>56,77,81</sup> above His-35 and the peptide flip region (see Figure 8). The negative electrostatic potential of this region seen in the high-pH form is neutralized in the low-pH form. The His-35 patch is located in close distance to the hydrophobic patch, which is known to be the site of direct interaction in electron-transfer reactions of PaAz.<sup>13,82,83</sup> Thus, the change in the electrostatic surface potential could affect the thermodynamics and the kinetics of the binding steps that precede the electron transfer. The importance of electrostatic interactions in electron-transfer reactions of PaAz has been demonstrated experimentally.<sup>84</sup> The hydrophobic patch is most likely also involved in interactions of PaAz with the mammalian tumor suppressor p53 that takes place during the induction of apoptosis by PaAz.<sup>9,85</sup> Thus, the change in the electrostatic potential of PaAz caused by the conversion could influence its interaction with binding partners within its roles as electron carrier and as therapeutic agent.

A similar effect was proposed for the electron-transfer complex of ferredoxin NADP-reductase (FNR) and ferredoxin.<sup>86</sup> In this case, the deprotonation of a glutamate residue close to the redox center of FNR might induce the release of ferredoxin from the complex during the catalytic cycle of FNR.

Instead of a change in the pH value, also electrostatic interaction of PaAz with a binding partner could induce the conversion between the high-pH and low-pH forms. The cooperativity of this conversion with the reduction of the copper center would in effect also shift the reduction potential of PaAz. A possible influence of the electron-transfer partner on the conformational equilibrium of PaAz has been noted before by Jeuken et al. for the case of a metal electrode in voltammetry.<sup>21</sup> A similar situation has been found for the electron-transfer reaction between amicyanin and methylamine dehydrogenase.<sup>87</sup> The presence of methylamine dehydrogenase in the electron-transfer complex prevents the dissociation and protonation of one of the histidine ligands of the copper center of amicyanin. In effect, the reduction potential of the copper center of amicyanin is decreased and the electron transfer from amicyanin to methylamine dehydrogenase becomes more favorable.

If the conversion between the high-pH and low-pH forms has a regulatory function, another aspect of this conversion could be functionally important. Namely, the kinetics of the conversion is very slow.<sup>16,65</sup> Experimentally determined rates for the conversion range from 0.23 to 45 s<sup>-1</sup>.<sup>16,21</sup> In effect, a regulatory signal transferred to PaAz by means of this conversion would decay very slowly on a molecular time scale.

The physiological function of PaAz in *P. aeruginosa* and the identity of physiological redox partners are not yet clearly established. Initially, PaAz was believed to function as electron carrier in respiratory denitrification.<sup>1</sup> In this pathway, *P. aeruginosa* uses nitrite and other intermediates of the denitrification pathway as terminal electron acceptors under anaerobic conditions. PaAz was believed to function as electron donor to cytochrome *cd*<sub>1</sub> nitrite reductase.<sup>1,88,89</sup> The gene for azurin was, however, later found not to be essential for denitrification<sup>3</sup> and not located close to other denitrification genes in the genome of *P. aeruginosa*.<sup>90</sup> Instead, two soluble cytochromes *c* from the nitrite reductase gene cluster were found also to be able to donate electrons to nitrite reductase at similar rates.<sup>91</sup> These cytochromes *c* are

cytochrome  $c_{551}$  (NirM) and NirC. Cytochrome  $c_{551}$  has been reported to be a more efficient electron donor to nitrite reductase than PaAz.<sup>89,92</sup> Also possible roles of PaAz as obligatory electron donor to aromatic amine dehydrogenase or ethanol dehydrogenase have been ruled out.<sup>3</sup> PaAz has also been reported to accept electrons from cytochrome  $bc_1$  if the bacterium utilizes succinate as electron donor, where again cytochrome  $c_{551}$  can fulfill the same function.<sup>4</sup> PaAz expression was found to be induced under oxidative stress conditions and to contribute to the oxidative stress resistance.<sup>3</sup>

The finding that PaAz does not have an essential role as electron carrier in denitrification<sup>3</sup> does not necessarily mean that it never fulfills this function in vivo. The presence of multiple, seemingly redundant electron carriers that are able to fulfill the same function is not unprecedented. An example is found in photosynthesis of cyanobacteria. Both the roles of the electron donor and of the electron acceptor of photosystem I can be fulfilled by two different proteins in these bacteria.<sup>93–96</sup> The electron donor plastocyanin which utilizes a type-1 copper center is replaceable by cytochrome  $c_6$  under copper deficiency.<sup>96,97</sup> The electron acceptor ferredoxin which utilizes an iron–sulfur center can be replaced by flavodoxin under iron deficiency.<sup>96,98</sup> One might ask, if the redundancy of PaAz and the cytochromes  $c$  as electron donors to the nitrite reductase of *P. aeruginosa* does also confer resistance against copper or iron deficiency.

It seems also conceivable, that the physiological function of PaAz is not that of an electron carrier between a single electron donor and a single electron acceptor. *P. aeruginosa* possesses versatile metabolic capabilities<sup>99</sup> and a branched respiratory system with several alternative terminal oxidases.<sup>4,100,101</sup> If PaAz does not possess a strictly essential role, it seems conceivable that it fulfills a regulatory function. The ability of PaAz to interact with different redox partners could enable it to distribute electrons between different branches of the respiratory system. Such a role could also explain how PaAz contributes to the resistance of the bacterium against redox stress.<sup>3</sup> Vijgenboom and co-workers pointed out that an increased amount of PaAz could help to avoid toxic levels of nitric oxide which could accumulate in the cell if there is an excess of reducing substrates for nitrite reductase.<sup>3</sup>

In light of our findings, one might also ask whether such a function for PaAz would be modulated by the pH value of the solution and/or the ready availability of electron acceptors for PaAz. Instead of completely shutting down or slowing the electron flow between a single electron donor and a single electron acceptor, a different role could be proposed for the pH-dependent transition of PaAz between the high-pH and low-pH forms. The alteration of the electrostatic surface potential of PaAz that accompanies the transition could change the relative affinity of PaAz for different redox partners. This mechanism could form the basis for the proposed balancing of the electron flow through different branches of the respiratory system. Pseudospecific interaction surfaces of electron-transfer proteins have been proposed before to allow for electron transfer with different reaction partners in dependence on external conditions.<sup>102</sup> Besides by a decrease in pH, the transition between the high-pH form and the low-pH form is also favored by the reduction of the copper center (see above). Since this transition is very slow,<sup>16,21</sup> the following scenario seems perceivable. If electron acceptors for PaAz are readily available, the transition does not happen because PaAz is quickly reoxidized. If, on the other hand, the amount of electron donors for PaAz becomes large and electron acceptors are not readily available, PaAz might

undergo the transition from the high-pH form to the low-pH form. The accompanying increase in the reduction potential and the altered electrostatic surface potential might then redirect PaAz toward alternative electron acceptors. The resulting redirection of the electron flow toward alternative electron sinks could prevent harmful effects of excess reducing power such as those described by Vijgenboom et al.<sup>3</sup>

#### 4. CONCLUSIONS AND OUTLOOK

We investigated the cooperativity between protonation, reduction, and conformational change in PaAz and discussed possible implications for the function of PaAz. The hydrogen bonding between His-35 and the peptide flip region on the one hand and the electrostatic interaction of His-35 and the copper center on the other hand lead to an effective coupling between protonation, reduction, and conformational change. Furthermore, we found that the conversion between the high-pH form and the low-pH form leads to a marked alteration of the electrostatic potential at the surface of PaAz. This alteration might affect the interaction of PaAz with its binding partners and could thus be relevant for the roles of PaAz as electron carrier and as therapeutic agent. In turn, interactions of PaAz with binding partners can affect the reduction potential of PaAz by shifting the equilibrium between its high-pH and low-pH forms.

We also used PaAz as a model system to demonstrate that structure-based calculations of cooperativity free energies are useful in detecting and quantifying cooperativity between events in biomolecules. There is a large body of experimental literature on PaAz, which could be used to verify our results. The calculation of cooperativity free energies permitted us to add a quantitative description of the coupling between these events and to reveal some yet unknown details. We believe that calculations of cooperativity free energies can provide valuable contributions for the deciphering of mechanistic details in more complex systems that are less accessible to experiment. In principle, cooperativity free energy calculations have an experimental pendant in double and triple mutant cycle experiments.<sup>73,74</sup> Such experiments are, however, not applicable in the case of protein cofactors or if the mutation induces larger structural changes in the protein. In addition, it might be impracticable to carry out screening experiments for many different sites that could be mechanistically important. The identity of mechanistically important sites in complex bioenergetic systems is often far from obvious. Obvious candidates for an application of cooperativity free energy calculations are complex bioenergetic systems in which mechanistic details are still debated. Such systems are bacterial reaction centers,<sup>103</sup> photosystems I and II,<sup>93,94,104</sup> cytochrome  $c$  oxidase,<sup>105,106</sup> and the cytochrome  $bc_1$  and  $b_6f$  complexes.<sup>107–109</sup>

#### ■ ASSOCIATED CONTENT

Supporting Information. Additional plots and figures referenced in the text and atomic partial charges and intrinsic energies of model compounds used in this work. This material is available free of charge via the Internet at <http://pubs.acs.org>.

#### ■ AUTHOR INFORMATION

##### Corresponding Author

\*E-mail: [thomas.ullmann@uni-bayreuth.de](mailto:thomas.ullmann@uni-bayreuth.de) (R.T.U.); [matthias.ullmann@uni-bayreuth.de](mailto:matthias.ullmann@uni-bayreuth.de) (G.M.U.).

## ACKNOWLEDGMENT

We gratefully acknowledge useful discussions with Verónica I. Dumit. This work was supported by the Deutsche Forschungsgemeinschaft through Grant DFG UL 174/8-1.

## REFERENCES

- (1) (a) Horio, T. *J. Biochem.* **1958**, *45*, 195–205. (b) Horio, T. *J. Biochem.* **1958**, *45*, 267–279.
- (2) Adman, E. T. *Curr. Opin. Struct. Biol.* **1991**, *1*, 895–904.
- (3) Vijgenboom, E.; Busch, J. E.; Canters, G. W. *Microbiology* **1997**, *143*, 2853–2863.
- (4) Hasegawa, N.; Arai, H.; Igarashi, Y. *Biosci. Biotechnol. Biochem.* **2003**, *67*, 121–126.
- (5) (a) Schobert, M.; Jahn, D. *Int. J. Med. Microbiol.* **2010**, *300*, 549–556. Cystic Fibrosis: (b) Schobert, M.; Tielen, P. *Future Microbiol.* **2010**, *5*, 603–621.
- (6) Yoon, S. S. *J. Bacteriol. Virol.* **2010**, *40*, 59–66.
- (7) Yamada, T.; Hiraoka, Y.; Gupta, T. K. D.; Chakrabarty, A. M. *Cell Cycle* **2004**, *3*, 752–755.
- (8) Chaudhari, A.; Fialho, A. M.; Ratner, D.; Gupta, P.; Hong, C. S.; Kahali, S.; Yamada, T.; Haldar, K.; Murphy, S.; Cho, W.; et al. *Cell Cycle* **2006**, *5*, 1642–1648.
- (9) Goto, M.; Yamada, T.; Kimbara, K.; Horner, J.; Newcomb, M.; Gupta, T. K. D.; Chakrabarty, A. M. *Mol. Microbiol.* **2003**, *47*, 549–559.
- (10) Apiyo, D.; Wittung-Stafshede, P. *Biochem. Biophys. Res. Commun.* **2005**, *332*, 965–968.
- (11) Yamada, T.; Fialho, A. M.; Punj, V.; Bratescu, L.; Gupta, T. K. D.; Chakrabarty, A. M. *Cell. Microbiol.* **2005**, *7*, 1418–1431.
- (12) (a) Nar, H.; Messerschmidt, A.; Huber, R.; van de Kamp, M.; Canters, G. W. *J. Mol. Biol.* **1991**, *221*, 765–772. (b) Nar, H. Röntgenkristallographische Strukturauflösung von Azurin aus *Pseudomonas aeruginosa*, Ph.D. thesis, Technical University of Munich, Munich, Germany, 1992.
- (13) van de Kamp, M.; Floris, R.; Hali, F. C.; Canters, G. W. *Eur. J. Biochem.* **1990**, *194*, 109–118.
- (14) van de Kamp, M.; Canters, G. W.; Andrew, C. R.; Sanders-Loehr, J.; Bender, C. J.; Peisach, J. *Eur. J. Biochem.* **1993**, *218*, 229–238.
- (15) St Clair, C. S.; Ellis, W. R. J.; Gray, H. B. *Inorg. Chim. Acta* **1991**, *191*, 149–155.
- (16) Kalverda, A. P.; Ubbink, M.; Gilardi, G.; Wijmenga, S. S.; Crawford, A.; Jeuken, L. J. C.; Canters, G. W. *Biochemistry* **1999**, *38*, 12690–12697.
- (17) Bashford, D.; Karplus, M.; Canters, G. W. *J. Mol. Biol.* **1988**, *203*, 507–510.
- (18) Ben-Naim, A. *Cooperativity and Regulation in Biochemical Processes*; Kluwer Academic/Plenum: New York, 2001.
- (19) Ullmann, R. T.; Ullmann, G. M. GMCT—Monte Carlo Simulation Software for Macromolecular Receptors with Multiple Ligands, Membrane Potential and Conformational Flexibility. Submitted for publication.
- (20) Lapin, G. A.; Segal, M. G.; Weatherburn, D. C.; Henderson, R. A.; Sykes, A. G. *J. Am. Chem. Soc.* **1979**, *101*, 2302–2306.
- (21) Jeuken, L. J. C.; Wisson, L.-J.; Armstrong, F. A. *Inorg. Chim. Acta* **2002**, *331*, 216–223.
- (22) (a) Korzhnev, D. M.; Karlsson, V. Y.; Orekhov, B. G.; Billeter, M. *Protein Sci.* **2003**, *12*, 56–65. (b) Zhuravleva, A. V.; Korzhnev, D. M.; Kupce, E.; Arseniev, A. S.; Billeter, M.; Orekhov, V. Y. *J. Mol. Biol.* **2004**, *342*, 1599–1611.
- (23) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (24) Stone, J. An Efficient Library for Parallel Ray Tracing and Animation, M.Sc. thesis, Computer Science Department, University of Missouri—Rolla, Rolla, MO, 1998.
- (25) Ullmann, R. T.; Ullmann, G. M. Unpublished results.
- (26) Bashford, D. An object-oriented programming suite for electrostatic effects in biological molecules An experience report on the MEAD project. In *Scientific computing in object-oriented parallel environments*; Ishikawa, Y., Oldehoeft, R., Reynnders, J., Tholburn, M., Eds.; Springer: Berlin/Heidelberg, Germany, 1997; Vol. 1343, pp 233–240.
- (27) Bashford, D. *Frontiers in Bioscience* **2004**, *9*, 1082–1099.
- (28) Becker, T.; Ullmann, R. T.; Ullmann, G. M. *J. Phys. Chem. B* **2007**, *111*, 2957–2968.
- (29) Ferreira, A. M.; Bashford, D. *J. Am. Chem. Soc.* **2006**, *128*, 16778–16790.
- (30) Gunner, M. R.; Mao, J.; Song, Y.; Kim, J. *Biochim. Biophys. Acta* **2006**, *1757*, 942–968.
- (31) Nielsen, J. E.; McCammon, J. A. *Protein Sci.* **2003**, *12*, 1894–1901.
- (32) Ullmann, G. M.; Knapp, E. W. *Eur. Biophys. J.* **1999**, *28*, 533–551.
- (33) Ullmann, G. M.; Kloppmann, E.; Essigke, T.; Krammer, E.-M.; Klingen, A. R.; Becker, T.; Bombarda, E. *Photosynth. Res.* **2008**, *97*, 33–53.
- (34) (a) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931–967. (b) Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Acc.* **1998**, *99*, 391–403.
- (35) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- (36) Swart, M.; Ehlers, A. W.; Lammertsma, K. *Mol. Phys.* **2004**, *102*, 2467–2474.
- (37) Swart, M.; van Duijnen, P. Th.; Snijders, J. G. *J. Comput. Chem.* **2001**, *22*, 79–88.
- (38) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (39) Perdew, J. P.; Yue, W. *Phys. Rev. B* **1986**, *33*, 8800–8802.
- (40) Lopez, X.; Schaefer, M.; Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 5010–5018.
- (41) Brünger, A. T.; Karplus, M. *Proteins* **1988**, *4*, 148–156.
- (42) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (43) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evansack, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (44) (a) Dunbrack, R. L. J.; Cohen, F. E. *Protein Sci.* **1997**, *6*, 1661–1681. (b) Dunbrack, R. L. J. *Curr. Opin. Struct. Biol.* **2002**, *12*, 431–440.
- (45) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219–10225.
- (46) Bashford, D.; Case, D. A.; Dalvit, C.; Tennant, L.; Wright, P. E. *Biochemistry* **1993**, *32*, 8045–8056.
- (47) Klapper, I.; Fine, R.; Sharp, K. A.; Honig, B. H. *Proteins* **1986**, *1*, 47–59.
- (48) (a) Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. *Biophys. J.* **2002**, *83*, 1731–1748. (b) Song, Y.; Mao, J.; Gunner, M. R. *J. Comput. Chem.* **2009**, *30*, 2231–2247.
- (49) Goldstein, R. F. *Biophys. J.* **1994**, *66*, 293–324.
- (50) (a) Wang, F.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050–2053. (b) Wang, F.; Landau, D. P. *Phys. Rev. E* **2001**, *64*, No. 056101.
- (51) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (52) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (53) Ullmann, R. T.; Ullmann, G. M. *J. Phys. Chem. B* **2011**, *115*, 507–521.
- (54) Valleau, J. P.; Card, D. N. *J. Chem. Phys.* **1972**, *57*, 5457–5462.
- (55) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (56) Corin, A. F.; Bersohn, R.; Cole, P. E. *Biochemistry* **1983**, *22*, 2032–2038.
- (57) Ugurbil, K.; Mitra, S. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 2039–2043.
- (58) Monari, S.; Battistuzzi, G.; Dennison, C.; Borsari, M.; Ranieri, A.; Siwek, M. J.; Sola, M. *J. Phys. Chem. C* **2010**, *114*, 22322–22329.
- (59) van Pouderoyen, G.; Mazumdar, S.; Hunt, N. I.; Hill, A. O.; Canters, G. W. *Eur. J. Biochem.* **1994**, *222*, 583–588.
- (60) Battistuzzi, G.; Borsari, M.; Canters, G. W.; de Waal, E.; Loschi, L.; Warmerdam, G.; Sola, M. *Biochemistry* **2001**, *40*, 6707–6712.

- (61) Pascher, T.; Bergström, J.; Malmström, B. G.; Vännngård, T.; Lundberg, L. G. *FEBS Lett.* **1989**, *258*, 266–268.
- (62) Pascher, T.; Karlsson, B. G.; Nordling, M.; Malmström, B. G.; Vännngård, T. *Eur. J. Biochem.* **1993**, *212*, 289–296.
- (63) Pettigrew, G. W.; Leitch, F. A.; Moore, G. R. *Inorg. Chim. Acta* **1983**, *331*, 216–223.
- (64) (a) Hill, J. C.; Allen, H.; Leer, O.; Smith, B. E.; Ambler, R. P. *Biochem. Biophys. Res. Commun.* **1976**, *70*, 331–338. (b) Hill, H. A. O.; Smith, B. E. *J. Inorg. Biochem.* **1979**, *11*, 79–93.
- (65) (a) Uğurbil, K.; Norton, R. S.; Allerhand, A.; Bersohn, R. *Biochemistry* **1977**, *16*, 886–894. (b) Uğurbil, K.; Bersohn, R. *Biochemistry* **1977**, *16*, 3016–3023.
- (66) van de Kamp, M.; Hali, F. C.; Rosato, N.; Agro, A. F.; Canters, G. W. *Biochim. Biophys. Acta, Bioenerg.* **1990**, *1019*, 283–292.
- (67) Simonson, T. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243–252.
- (68) Webb, H.; Tynan-Connolly, B. M.; Lee, G. M.; Farrell, D.; O'Meara, F.; Søndergaard, C. R.; Teilum, K.; Hewage, C.; McIntosh, L. P.; Nielsen, J. E. *Proteins* **2010**, *79*, 685–702.
- (69) (a) Wilson, M. T.; Greenwood, C.; Brunori, M.; Antonini, E. *Biophys. J.* **1975**, *145*, 449–457. (b) Wilson, M. T.; Greenwood, C.; Brunori, M.; Antonini, E. *Biophys. J.* **1975**, *151*, 185–188.
- (70) Ullmann, G. M. *J. Phys. Chem. B* **2000**, *104*, 6293–6301.
- (71) Bombarda, E.; Ullmann, G. M. *J. Phys. Chem. B* **2010**, *114*, 1994–2003.
- (72) Ullmann, G. M. *J. Phys. Chem. B* **2003**, *107*, 1263–1271.
- (73) Carter, P. J.; Winter, G.; Wilkinson, A. J.; Fersht, A. R. *Cell* **1984**, *38*, 835–840.
- (74) Horovitz, A. *Folding Des.* **1996**, *1*, R121–R126.
- (75) Onufriev, A.; Case, D. A.; Ullmann, G. M. *Biochemistry* **2001**, *40*, 3413–3419.
- (76) Spassov, V.; Bashford, D. *Protein Sci.* **1998**, *7*, 2012–2025.
- (77) Silvestrini, M. C.; Brunori, M.; Wilson, M. T.; Darley-Usmar, V. M. *J. Inorg. Biochem.* **1981**, *14*, 327–338.
- (78) Ubbink, M. *FEBS Lett.* **2009**, *583*, 1060–1066.
- (79) Jeuken, L. J. C. *Biochim. Biophys. Acta, Bioenerg.* **2003**, *1604*, 67–76.
- (80) (a) Davidson, V. L. *Biochemistry* **1996**, *35*, 14036–14039. (b) Davidson, V. L. *Biochemistry* **2000**, *39*, 4924–4928. (c) Davidson, V. L. *Biochemistry* **2002**, *41*, 14633–14636.
- (81) Farver, O.; Blatt, Y.; Pecht, I. *Biochemistry* **1982**, *21*, 3556–3561.
- (82) van de Kamp, M.; Floris, R.; Hali, F. C.; Canters, G. W. *J. Am. Chem. Soc.* **1990**, *112*, 907–908.
- (83) (a) van Amsterdam, I. M. C.; Ubbink, M.; Jeuken, L. J. C.; Verbeet, M. P.; Einsle, O.; Messerschmidt, A.; Canters, G. W. *Chem.—Eur. J.* **2001**, *7*, 2398–2406. (b) van Amsterdam, I. M. C.; Ubbink, M.; Jeuken, L. J. C.; Verbeet, M. P.; Einsle, O.; Messerschmidt, A.; Canters, G. W. *Nat. Struct. Biol.* **2002**, *9*, 48–52. (c) van Amsterdam, I. M. C.; Ubbink, M.; Canters, G. W. *Inorg. Chim. Acta* **2002**, *331*, 296–392.
- (84) Sokerina, E. V.; Ullmann, G. M.; van Pouderooyen, G.; Canters, G. W.; Kostić, N. M. *J. Biol. Inorg. Chem.* **1999**, *4*, 111–121.
- (85) Xu, C.; Yin, J.; Zhao, B. *Sci. China, Ser. C: Life Sci.* **2010**, *53*, 1181–1188.
- (86) Dumit, V. I.; Essigke, T.; Cortez, N.; Ullmann, G. M. *J. Mol. Biol.* **2010**, *397*, 814–825.
- (87) Zhu, Z.; Cunane, L. M.; Chen, Z.; Durley, R. C. E.; Mathews, F. S.; Davidson, V. L. *Biochemistry* **1998**, *37*, 17128–17136.
- (88) Tordi, M. G.; Silvestrini, M. C.; Colosimo, A.; Tuttobello, L.; Brunori, M. *Biochem. J.* **1985**, *230*, 797–805.
- (89) Silvestrini, M. C.; Falcinelli, S.; Ciabatti, I.; Cutruzzolà, F.; Brunori, M. *Biochimie* **1994**, *76*, 641–654.
- (90) Vollack, K.-U.; Xie, J.; Härtig, E.; Römling, U.; Zumft, W. G. *Microbiol.* **1997**, *143*, 2853–2863.
- (91) Hasegawa, N.; Arai, H.; Igarashi, Y. *Biochem. Biophys. Res. Commun.* **2001**, *288*, 1223–1230.
- (92) Barber, D.; Parr, S. R.; Greenwood, C. *Biochem. J.* **1961**, *157*, 431–438.
- (93) Amunts, A.; Nelson, N. *Structure* **2009**, *17*, 637–650.
- (94) Fromme, R.; Grotjohann, I.; Fromme, P. Structure of Photosystems I and II. In *Structural Biology of Membrane Proteins*; Grishshamer, R., Buchanan, S. K., Eds.; The Royal Society of Chemistry: London, 2006; Chapter 18, pp 320–348.
- (95) Rogers, L. J. Ferredoxins, Flavodoxins, and Related Proteins: Structure, Function, and Evolution. In *Cyanobacteria: A Comprehensive Review*; Fay, P., van Baalen, C., Eds.; Elsevier: Amsterdam, The Netherlands, 1987; pp 35–67.
- (96) Ullmann, G. M. Simulation and Analysis of Docking and Molecular Dynamics of Electron-Transfer Protein Complexes., Ph.D. thesis, Freie Universität Berlin, Berlin, 1998
- (97) Ullmann, G. M.; Hauswald, M.; Jensen, A.; Kostić, N. M.; Knapp, E. W. *Biochemistry* **1997**, *36*, 16187–16196.
- (98) Ullmann, G. M.; Hauswald, M.; Jensen, A.; Knapp, E. W. *Proteins* **2000**, *38*, 301–309.
- (99) Poole, R. K.; Cook, G. M. Redundancy of Aerobic Respiratory Chains in Bacteria? Routes, Reasons and Regulation. In *Advances in Microbial Physiology*; Poole, R. K., Ed.; Academic Press: New York, 2000; Vol. 43, pp 165–224.
- (100) Matsushita, K.; Yamada, M.; Shinagawa, E.; Adachi, O.; Ameyama, M. *J. Bacteriol.* **1980**, *141*, 389–392.
- (101) Stover, C. K.; Pham, X. Q.; Erwin, A. L.; Mizoguchi, S. D.; Warrenner, P.; Hickey, M. J.; Brinkman, F. S. L.; Hufnagle, W. O.; Kowalik, D. J.; Lagrou, M.; et al. *Nature* **2000**, *406*, 959–964.
- (102) Williams, P. A.; Fulop, V.; Leung, Y.-C.; Chan, C.; Moir, J. W. B.; Howlett, G.; Ferguson, S. J.; Radford, S. E.; Hajdu, J. *Nat. Struct. Biol.* **1995**, *2*, 975–982.
- (103) Okamura, M. Y.; Paddock, M. L.; Graige, M. S.; Feher, G. *Biochim. Biophys. Acta, Bioenerg.* **2000**, *1458*, 148–163.
- (104) Renger, G.; Renger, T. *Photosynth. Res.* **2008**, *98*, 53–80.
- (105) Chakrabarty, S.; Namslauer, I.; Brzezinski, P.; Warshel, A. *Biochim. Biophys. Acta, Bioenerg.* **2011**, *1807*, 413–426.
- (106) Capitanio, G.; Martino, P. L.; Capitanio, N.; Papa, S. *Biochim. Biophys. Acta, Bioenerg.* **2011**, in press (DOI: 10.1016/j.bbabi.2011.02.004).
- (107) Mulikdjanian, A. Y. *Biochim. Biophys. Acta, Bioenerg.* **2010**, *1797*, 1858–1868.
- (108) Crofts, A. R. *Annu. Rev. Physiol.* **2004**, *66*, 689–733.
- (109) Hunte, C.; Palsdottir, H.; Trumpower, B. *FEBS Lett.* **2003**, *545*, 39–46.



# Chapter 7

## Manuscript D:

Thermodynamics of Transport through  
the Ammonium Transporter Amt-1  
Investigated with Free Energy  
Calculations

R. Thomas Ullmann, Susana L. A. Andrade and G. Matthias Ullmann  
J. Phys. Chem. B 2011, 116, 9690-9703  
doi: 10.1021/jp305440f

# Thermodynamics of Transport Through the Ammonium Transporter Amt-1 Investigated with Free Energy Calculations

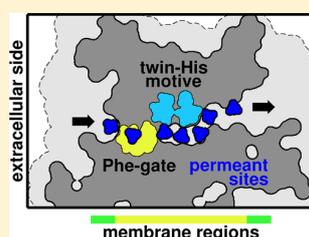
R. Thomas Ullmann,<sup>\*,†</sup> Susana L. A. Andrade,<sup>\*,‡</sup> and G. Matthias Ullmann<sup>\*,†</sup>

<sup>†</sup>Structural Biology/Bioinformatics, University of Bayreuth, Universitätsstrasse 30, BGI, 95447 Bayreuth, Germany

<sup>‡</sup>Institute of Organic Chemistry and Biochemistry and BIOS Centre for Biological Signalling Studies, Albert-Ludwigs-Universität Freiburg, Albertstrasse 21, 79104 Freiburg, Germany

## S Supporting Information

**ABSTRACT:** Amt-1 from *Archaeoglobus fulgidus* (AfAmt-1) belongs to the Amt/Rh family of ammonium/ammonia transporting membrane proteins. The transport mode and the precise microscopic permeation mechanism utilized by these proteins are intensely debated. Open questions concern the identity of the transported substrate (ammonia and/or ammonium) and whether the transport is passive or active. To address these questions, we studied the overall thermodynamics of the different transport modes as a function of the environmental conditions. Then, we investigated the thermodynamics of the underlying microscopic transport mechanisms with free energy calculations within a continuum electrostatics model. The formalism developed for this purpose is of general utility in the calculation of binding free energies for ligands with multiple protonation forms or other binding forms. The results of our calculations are compared to the available experimental and theoretical data on Amt/Rh proteins and discussed in light of the current knowledge on the physiological conditions experienced by microorganisms and plants. We found that microscopic models of electroneutral and electrogenic transport modes are in principle thermodynamically viable. However, only the electrogenic variants have a net thermodynamic driving force under the physiological conditions experienced by microorganisms and plants. Thus, the transport mechanism of AfAmt-1 is most likely electrogenic.



thermodynamics of possible permeation mechanisms investigated with free energy calculations  
electroneutral  $\text{NH}_3$  uniport  
electrogenic  $\text{NH}_4^+$  uniport  
electrogenic  $\text{NH}_3/\text{H}^+$  symport

## INTRODUCTION

Amt-1 from *Archaeoglobus fulgidus* (AfAmt-1) belongs to the widespread family of Amt/Rh transmembrane transport proteins. The ammonium transport (Amt) proteins supply microorganisms and plants with ammonium as the most directly utilizable nitrogen source.<sup>1–11</sup> The human Rhesus (Rh) proteins are implicated in ammonium excretion and pH regulation functions.<sup>11–13</sup>

The precise molecular mechanism of substrate conduction through Amt/Rh proteins is not yet known. It is debated whether the uncharged ammonia or the charged ammonium ion is the species that is actually translocated by the proteins.<sup>2–11,14</sup> Traditionally, it was thought that the transported species is  $\text{NH}_4^+$ .<sup>1</sup> This opinion was grounded on several lines of evidence. First, the  $\text{pK}_a$  value of 9.25 makes  $\text{NH}_4^+$  much more abundant than  $\text{NH}_3$  at physiological pH values around 7.0. Second, lipid membranes possess a significant permeability for the apolar  $\text{NH}_3$  molecule, which seemed to make a dedicated transport protein unnecessary. Third and possibly most important, the charge carried by the ammonium ion makes it possible to drive its electrogenic transmembrane transport by the electric transmembrane potential. This possibility would enable a cell to take up ammonium also at very low concentrations of ammonium in the extracellular medium and to compensate the loss of ammonia through

unspecific permeation of the cell membrane in a process called cyclic retention.<sup>1</sup>

The assumption of an active transport mechanism was called into question by some physiological experiments whose interpretation led to the hypothesis that Amt/Rh proteins actually facilitate the passive, electroneutral transport of the uncharged ammonia molecule.<sup>11,15,16</sup> These experiments were supported by the first X-ray crystal structures of AmtB from *Escherichia coli* (EcAmtB) that revealed a transmembrane pore with a hydrophobic central region whose permeation by the charged ammonium ion was considered to be thermodynamically prohibitive.<sup>8,11,16,17</sup> According to this hypothesis,  $\text{NH}_4^+$  would be recruited at the extracellular side leaving the proton behind in the extracellular phase. The resulting ammonia molecule would then permeate the transmembrane pore and be reprotonated at the intracellular side.

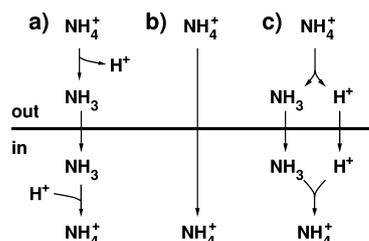
However, as noted in our earlier work, an electroneutral uniport of  $\text{NH}_3$  poses the bioenergetic problem of how a cell can retrieve ammonia/ammonium at low extracellular concentrations of the substrate.<sup>4,18</sup> Reports of membrane potential dependent uptake currents in plant and microbial Amts support the view of an electrogenic transport that would involve  $\text{NH}_4^+$

Received: June 3, 2012

Revised: July 13, 2012

Published: July 17, 2012

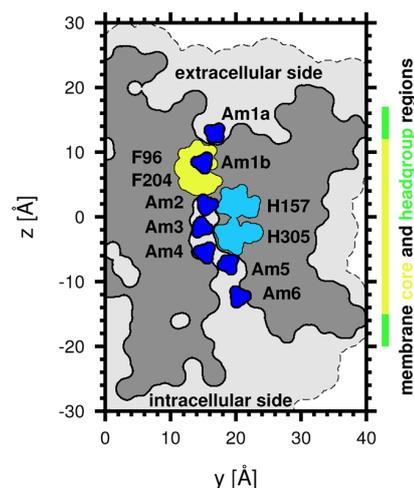
either in a uniport of  $\text{NH}_4^+$  or a symport of  $\text{NH}_3$  and  $\text{H}^+$ .<sup>19–27</sup> A membrane potential dependence of the ammonium uptake activity was also reported for Rh type transporters,<sup>28–30</sup> but partly discarded as too weak to be significant.<sup>28,29</sup> The finding that the transport activity of the Rh proteins was stimulated by an extracellular alkalization led to the interpretation that  $\text{NH}_3$  must be involved in either a uniport of  $\text{NH}_3$ <sup>11,13,28,29</sup> or a symport of  $\text{NH}_3$  and  $\text{NH}_4^+$ .<sup>30</sup> A summarizing scheme of the possible transport mechanisms considered to date<sup>4,5,31</sup> is shown in Figure 1.<sup>4</sup>



**Figure 1.** Mechanistic scheme of transport modes that could be operative in AfAmt-1.<sup>4</sup> (a) At physiological pH values ammonium will be almost exclusively in the cationic form, such that translocation of  $\text{NH}_3$  implies extracellular deprotonation and intracellular reprotonation. Electrogenic transport of ammonium is either a net uniport of  $\text{NH}_4^+$  as depicted in panel b or a symport of  $\text{NH}_3/\text{H}^+$  as depicted in panel c. Both electrogenic transport modes result in a net transport of  $\text{NH}_4^+$ . During a uniport the ammonium ion is translocated as an intact entity. In contrast, during a symport the ammonium ion is separated into an ammonia molecule and a proton, which may be translocated along different routes, and recombined on the inner side of the membrane.

Figure 2 shows the structure of one monomer within the AfAmt-1 trimer highlighting the transmembrane pore and putative intermediate sites passed by the permeant.<sup>18</sup> The transmembrane pore can be subdivided into a central hydrophobic pore lumen and two wider and less hydrophobic vestibule regions at the intracellular and extracellular ends of the pore. The extracellular vestibule and the pore lumen are separated by a narrow constriction that is formed by two conserved phenylalanine residues (F96 and F204). The so-called Phe-gate has been attributed a role in the dehydration and subsequent deprotonation of the ammonium ion either directly at or close to the Phe-gate.<sup>5,32,33</sup> The pore lumen is lined by hydrophobic residues except for a conserved pair of histidine residues (H157 and H305) termed the twin-His motif. The twin-His residues are arranged in a nearly coplanar geometry, with their  $\text{N}\delta$  atoms within hydrogen-bonding distance. A role of the twin-His motif in the conduction mechanism has been suggested based on the structural data and mutagenesis experiments. This role could either be the providing of hydrogen bond interactions to the permeant<sup>16,34</sup> or as transient proton acceptor in the deprotonation of the ammonium ion.<sup>5,17,35</sup>

Most previous theoretical studies of substrate permeation through Amt/Rh proteins concentrated on different aspects of substrate permeation through EcAmtB.<sup>14,32,33,36–49</sup> One work studied the thermodynamics of substrate permeation through Rh50 from *Nitrosomonas europaea* (NeRh50).<sup>50</sup> In this work, we studied the thermodynamics of the different possible transport mechanisms of AfAmt-1. We extend upon previous theoretical studies by including the effects of substrate



**Figure 2.** Structure of Amt-1 from *Archaeoglobus fulgidus* (AfAmt-1). The extracellular side is shown at the top and the intracellular side at the bottom. The figure shows a single monomer. The boundaries of the membrane core (yellow) and headgroup regions (green) are indicated by the bars on the right-hand side of the plot. The white regions at the top and the bottom denote the solvent phases. The light-gray outer region of Amt-1 is a projection of the solvent inaccessible volume of the transporter trimer into a plane perpendicular to the membrane. The darker inner region is a projection of a thin slice of Amt-1 into the same projection plane. The slice plane is slightly tilted with respect to the membrane normal to follow the course of the transmembrane pore. The positions of the investigated permeant positions (blue), the twin-histidine motif (cyan) and the Phe-gate (light yellow) are indicated.

concentration, electrochemical transmembrane gradients, proton-coupled binding equilibria, and competitive binding of different ligand species. Our free energy calculations are based on a microstate description of AfAmt-1 within a continuum electrostatics/molecular mechanics model.<sup>51,52</sup> Such a description can provide valuable insight into biomolecular function and has already been successfully applied to other membrane proteins like bacteriorhodopsin, bacterial and plant photosynthetic reaction centers, cytochrome  $bc_1$ , quinol-fumarate reductase, and cytochrome  $c$  oxidase.<sup>52–57</sup>

## METHODS

**Microstate Description of the Receptor–Ligand System.** Our model is based on a microstate description of the system.<sup>51,58,59</sup> A microstate is defined by a particular instance occupied by each site. The term instance describes the combination of a particular set of atomic partial charges, a particular number of bound ligands of each ligand type and a particular set of atomic coordinates. The energy of a microstate  $\mathbf{n}$  is given by<sup>51</sup>

$$E_{\mathbf{n}}^{\text{micro}} = \sum_{i=1}^{N^{\text{sites}}} (E_{i,k}^{\text{int},r} - \sum_m \nu_{i,k,m} \bar{\mu}_m) + \sum_{i=1}^{N^{\text{sites}}} \sum_{j=1}^{j<i} W_{i,k,j,l} \quad (1)$$

where  $N^{\text{sites}}$  is the number of sites and  $\mathcal{L}$  is the number of ligand types.  $E_{i,k}^{\text{int},r}$  is the intrinsic energy of site  $i$  in instance  $k$ , where the superscript  $r$  designates the receptor environment. The stoichiometric coefficient  $\nu_{i,k,m}$  indicates the number of ligands of type  $m$  bound to instance  $k$  of site  $i$  and  $\bar{\mu}_m$  is the electrochemical potential of this ligand in the adjacent solution.

$W_{i,k,j,l}$  is the interaction energy of sites  $i$  and  $j$  in their respective instances  $k$  and  $l$ .

The intrinsic energy in the receptor environment  $E_{i,k}^{\text{intr}}$  is calculated relative to the intrinsic energy of an appropriate model compound in solution  $E_{i,k}^{\text{intr,m}60,61}$

$$E_{i,k}^{\text{intr},r} = E_{i,k}^{\text{intr},m} + \Delta E_{i,k}^{\text{intr}} \quad (2)$$

The intrinsic energy shift  $\Delta E_{i,k}^{\text{intr}}$  upon transfer from the solution to the receptor environment is obtained from a continuum-electrostatics/molecular mechanics model.<sup>51,52,62,63</sup>

The definition of an equilibrium state requires that each site binds its ligands exclusively from one of the membrane sides only and that no transfer of ligands takes place between the membrane sides.<sup>62,63</sup> Thus, the connectivity of each site to one of the membrane sides has to be determined. In the transmembrane pore of AfAmt-1, the Phe-gate separates the membrane sides and thus defines the connectivity of the pore-lining residues. The permeant sites Am1a and Am1b are connected to the extracellular side. The rest of the permeant sites and the twin-His residues are connected to the intracellular side. All other sites are assigned to either membrane side based on their exposure to either solvent phase or their involvement in hydrogen bonding networks (Table S1 of the Supporting Information).

An additional complication arises if a ligand can undergo binding reactions with other ligands, e.g., if the ligand itself is protonatable. In such a case all of the binding equilibria involving the macromolecular receptor, the ligand, and the proton are coupled. The corresponding expressions for the ligand chemical potentials in our case are given below.

**Electrochemical Potentials of the Ligands.** The chemical potential of a ligand  $i$  is given by

$$\mu_i = \mu_i^\circ + \beta^{-1} \ln a_i \quad (3)$$

where  $\mu_i^\circ$  and  $a_i$  are the standard chemical potential and the activity of the ligand, respectively. The electrochemical potential of the ligand  $i$  extends the chemical potential by the energy of the ligand in the electrostatic potential of the surrounding solution

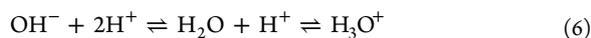
$$\bar{\mu}_i = \mu_i + z_i F \Delta \psi \quad (4)$$

where  $z_i$  is the formal charge of the ligand,  $F$  is the Faraday constant, and  $\Delta \psi$  is the electric transmembrane potential. By convention, the Heaviside step function  $\Theta$  adopts a value of 0 if the ligand resides in the extracellular phase and a value of 1 if the ligand resides the intracellular phase.<sup>64,65</sup>

In our case, there is a total of six different ligand species comprising the proton and the permeant species ammonia, ammonium, hydroxyl ion, water, and hydronium ion. The reaction equations for the corresponding protonation equilibria are given by



and



The equilibrium condition demands that the sum of the electrochemical potentials of all species is equal for each stage of the protonation equilibria eqs 5 and 6. Thus, we can write the electrochemical potential of each ligand species as the sum of the electrochemical potentials of the fully deprotonated ligand and of the protons bound by the species

$$\bar{\mu}_{\text{NH}_4^+} = \bar{\mu}_{\text{NH}_3} + \bar{\mu}_{\text{H}^+} \quad (7)$$

$$\bar{\mu}_{\text{H}_2\text{O}} = \bar{\mu}_{\text{OH}^-} + \bar{\mu}_{\text{H}^+} \quad (8)$$

$$\bar{\mu}_{\text{H}_3\text{O}^+} = \bar{\mu}_{\text{OH}^-} + 2\bar{\mu}_{\text{H}^+} \quad (9)$$

Consequently, there are three independent ligand electrochemical potentials ( $\bar{\mu}_{\text{H}^+}$ ,  $\bar{\mu}_{\text{OH}^-}$ , and  $\bar{\mu}_{\text{NH}_3}$ ) for each membrane side.

The activity of the proton  $a_{\text{H}^+}$  is given by the definition of the pH as

$$a_{\text{H}^+} \equiv 10^{-\text{pH}} \quad (10)$$

Similar to the pH, we define the functions pO and pN, as measures of the total activity of all water and ammonia protonation forms, respectively

$$a_{\text{NH}_3} + a_{\text{NH}_4^+} \equiv 10^{-\text{pN}} \quad (11)$$

$$a_{\text{OH}^-} + a_{\text{H}_2\text{O}} + a_{\text{H}_3\text{O}^+} \equiv 10^{-\text{pO}} \quad (12)$$

If we assume that the activity coefficients of all protonation forms are reasonably close to a value of 1.0, pN and pO will be independent of the pH value. With this assumption, we can express pN and pO as functions of the total concentration of all ammonia and water species, respectively

$$\text{pN} \approx -\log_{10} \left[ \frac{c_{\text{NH}_3} + c_{\text{NH}_4^+}}{c^\circ} \right] \quad (13)$$

$$\text{pO} \approx -\log_{10} \left[ \frac{c_{\text{OH}^-} + c_{\text{H}_2\text{O}} + c_{\text{H}_3\text{O}^+}}{c^\circ} \right] \quad (14)$$

where the standard concentration is, by definition, given by  $c^\circ \equiv 1$  mol/L. Our assumption should be fulfilled reasonably well under physiologically relevant conditions, i.e., for small concentrations of ammonia and ammonium  $c_{\text{NH}_3} + c_{\text{NH}_4^+} \ll 1$  mol/L and pH values far from the extreme ends of the pH range. The activities of the individual protonation forms of ammonia and water can then be expressed as functions of pN and pO. The activity of  $\text{NH}_3$  is given by

$$a_{\text{NH}_3} = \frac{10^{-\text{pN}}}{1 + \exp[-\beta(\mu_{\text{NH}_4^+}^\circ - \mu_{\text{NH}_3}^\circ - \mu_{\text{H}^+}^\circ)]} \quad (15)$$

The activity of the  $\text{OH}^-$  ion is given by

$$a_{\text{OH}^-} = 10^{-\text{pO}} / \left\{ 1 + \exp[-\beta(\mu_{\text{H}_2\text{O}}^\circ - \mu_{\text{OH}^-}^\circ - \mu_{\text{H}^+}^\circ)] + \exp[-\beta(\mu_{\text{H}_3\text{O}^+}^\circ - \mu_{\text{OH}^-}^\circ - 2\mu_{\text{H}^+}^\circ)] \right\} \quad (16)$$

The total chemical potentials of the ammonia and the hydroxyl ion can be calculated via eq 3. The chemical potentials of the other permeant species follow from eqs 7–9. A detailed derivation of eqs 15 and 16 can be found in section A of the Supporting Information. Plots of the chemical potential of all permeant species as a function of the pH value of the solution can be found in Figure S1 of the Supporting Information. The standard chemical potentials of the permeant species are available from experiment (see Table S4 of the Supporting Information). In our calculations, we assume the total concentration of all water species to be fixed at 55.5 mol/L as calculated from the density and the molecular weight of water ( $\rho_{\text{H}_2\text{O}} = 1$  kg/L,  $M_{\text{H}_2\text{O}} = 18.015$  g/mol).

**Simulation Setup.** Our simulations are based on the crystal structure of AfAmt-1 in the native form (PDB code 2B2F).<sup>18</sup> Hydrogen atoms were added with HBUILD<sup>66</sup> in CHARMM<sup>67</sup> and their positions were subsequently energy minimized using the CHARMM force field.<sup>68</sup> The intrinsic energies and interaction energies were computed from a continuum electrostatics/molecular mechanics model implemented in our program GCEM<sup>69</sup> based on our modified version of the MEAD library.<sup>55,70</sup> All MC simulations were carried out with our program suite GMCT.<sup>51</sup> Equilibrium probabilities of binding forms for individual sites were computed with the Metropolis MC method.<sup>71</sup> Free energy calculations were performed with the free energy perturbation method<sup>72</sup> in our recently presented generalization<sup>73</sup> combined with the Bennett acceptance ratio method.<sup>74</sup> Details of our computational method are specified in the remaining parts of this section.

**Structure Preparation.** The structure of AfAmt-1 in the native form (PDB code 2B2F) was used as basis for our calculations.<sup>18</sup> Hydrogen atoms were added with HBUILD<sup>66</sup> in CHARMM<sup>67</sup> and their positions were subsequently energy minimized using the CHARMM force field as described below.<sup>68</sup> Two additional rotamer positions were added for each hydroxyl or sulfhydryl hydrogen atom by varying the corresponding torsion angle in steps of 120°. Four alternative hydrogen positions for protonated forms of carboxylic acids were added to represent the syn and trans configurations of the dissociable proton at each carboxyl oxygen atom. The corresponding angles and bond lengths were taken from ref 69.

The positions of the permeant sites (see Figure 2) were determined from putative binding sites obtained from experimental and theoretical studies at AfAmt-1<sup>18</sup> and the close homologue EcAmtB.<sup>16,17,32,33,48</sup> Am1a corresponds to the ammonium recruitment site between W137 and S208 identified in the crystallographic studies.<sup>16–18,35</sup> This recruitment site has also been verified in several independent MD studies of EcAmtB.<sup>32,33,36,37,39,41,44,47,48</sup> Am1b was modeled based on observations in MD simulations of EcAmtB, where ammonium was sandwiched by the side chains of the two phenylalanine residues forming the Phe-gate.<sup>32,33,39,49</sup> Similar conformations are also observed in computational studies of a complex formed by ammonium and two benzene molecules in aqueous solution.<sup>75</sup> This binding site was proposed to promote dehydration of the ammonium ion and consequently to facilitate its deprotonation either directly at this site or at Am2.<sup>32,33,39</sup> Am2 was placed at the corresponding ammonia position proposed for EcAmtB (PDB 1U7G).<sup>16</sup> Am3 and Am4 were placed at xenon binding sites as found in the xenon-pressurized structure of AfAmt-1 (PDB 2B2J).<sup>18</sup> These sites closely comply with the putative ammonia or water binding sites identified in the transmembrane pore lumen of AfAmt-1<sup>18</sup> and EcAmtB.<sup>16,17,35</sup> Am5 was placed at the position of a water molecule (water 414) found in PDB 2B2F directly below H305 at the lower end of the hydrophobic pore lumen.<sup>18</sup> Am6 was modeled based on a proposed site of ammonia reprotonation in the intracellular vestibule of EcAmtB between the side chains of the residues that are equivalent to D300 and S263 of AfAmt-1.<sup>16,32,33,48</sup> The permeant positions were energy minimized while setting all atomic partial charges to zero to avoid biasing their protonation state. The positions of the permeant heavy atoms were constrained harmonically with a force constant of 0.5 kcal/(mol Å<sup>2</sup>) during the minimization. The coordinates of all hydrogen atoms were included in the energy minimization. All protein heavy atoms except for those of the side chains of

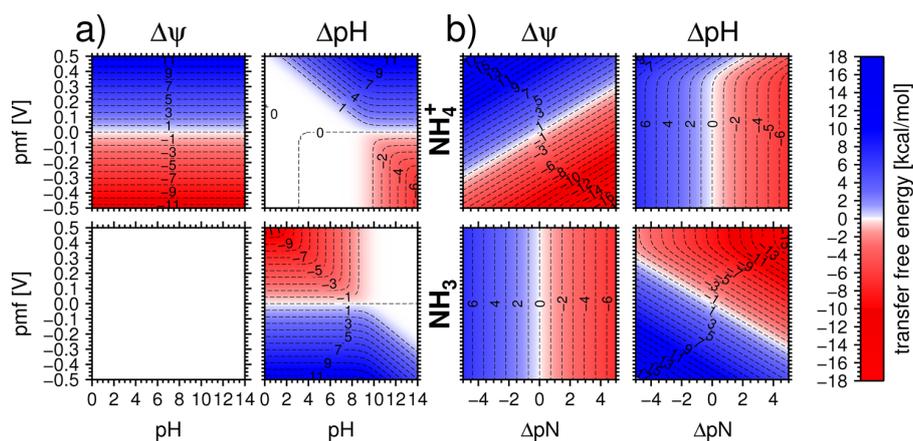
F96 and F204 were kept fixed during the minimization. Additional rotamers were added for all permeant sites by randomly rotating the initially generated coordinates in space resulting in an 8-fold increase in the number of rotamers.

**Calculation of Intrinsic Energies and Interaction Energies.** The intrinsic energies and interaction energies were computed from a continuum electrostatics/molecular mechanics model implemented in GCEM.<sup>76</sup> We considered a single monomer explicitly with all titratable sites, whereas the other two monomers of the trimer are represented implicitly by their dielectric regions. All aspartate, histidine, glutamate, lysine, arginine, cysteine, and tyrosine residues and the termini were considered as protonatable sites. The permeant sites in the transmembrane pore were modeled with binding forms for each of the considered permeants and a ligand-free binding form as described above.

We used a detailed charge model with explicit hydrogen positions for all protonatable sites. Atomic partial charges for standard forms of amino-acid residues were taken from the CHARMM22 parameter set.<sup>68</sup> Atomic partial charges of nonstandard forms of amino acid residues were taken from ref 69. Model compounds of protonatable amino acids include the entire residue plus the directly neighboring CHARMM charge groups belonging to the backbone of the preceding and succeeding amino acids to ensure charge neutrality and to mimic an *N*-formyl,*N*-methylamide blocked amino acid compounds.<sup>61</sup> The intrinsic energies of the model compounds for protonatable amino acid residues in aqueous solution were calculated from pK<sub>a</sub> values of appropriate model compounds taken from the literature<sup>52,77</sup> as described in ref 69. Atomic partial charges for the permeant sites were obtained from density functional theory calculations with the ADF program<sup>78,79</sup> (functionals VWN<sup>80</sup> and BP86<sup>81,82</sup> with a QZ4P basis set). The atomic partial charges were calculated with the multipole derived charge analysis method<sup>83</sup> (Table S2 of the Supporting Information). The intrinsic energies for the permeants in solution were computed from experimental data (Table S3 of the Supporting Information). The Lennard-Jones parameters of ammonia/ammonium were taken from the CHARMM22 force field (model compounds methylamine/methylammonium).<sup>68</sup> The Lennard-Jones parameters for the water species were taken from the TIP3P water model.<sup>84</sup>

Conformational energies of the sites were computed using the CHARMM22 force field,<sup>68</sup> and added to the intrinsic energies. We used bonded terms involving atoms of the site, and Lennard-Jones interaction energies within the site and between the site and the background (i.e., parts of the protein not belonging to any site). Lennard-Jones interactions were also added to the site–site interaction energies.

MEAD uses a finite-difference method on cubic grids to solve the linearized Poisson–Boltzmann equation.<sup>85,86</sup> The dielectric constant of the protein including the permeant sites was set to 4.<sup>55</sup> The dielectric constant of the solvent including protein cavities and unoccupied space of the transmembrane pore was set to 80. The position of the membrane was assigned according to the OPM database<sup>87</sup> and visual inspection of PDB 2B2F. The dielectric constant of the membrane core, representing the hydrophobic lipid tails, was set to 2 and the dielectric constant of the polar headgroup region was set to 20. The membrane core region extends from  $z = -15$  to  $+12$  Å in the coordinates defined by PDB 2B2F (see Figure 2). The headgroup regions have a thickness of 5 Å. The temperature was set to 298.15 K. The dielectric boundary between solute



**Figure 3.** Transfer free energies for the import of ammonia (lower row) and ammonium (upper row). The transfer free energy is color coded (see color bar) and indicated by isocontours drawn in constant intervals of 1.0 kcal/mol (contour values given in kcal/mol). (a) Transfer free energies as a function of the intracellular pH and the proton-motive force. (b) Transfer free energies as a function of the transmembrane pN difference ( $\Delta pN = pN^{\text{in}} - pN^{\text{out}}$ ) and the pmf. (a and b) The pmf in the left column consists entirely of an electric transmembrane potential ( $\text{pmf} = \psi^{\text{in}} - \psi^{\text{out}}$ ). The pmf in the right column consists entirely of a transmembrane pH difference ( $\text{pmf} \approx -59 \text{ mV } \Delta \text{pH}$ ).

and solvent was calculated using a water probe sphere of 1.4 Å radius and the atomic radii (1.0 Å for H, 1.55 Å for N, 1.7 Å for C, 1.5 Å for O, and 1.8 Å for S). The ionic strength was set to 0.15 M. The thickness of the ion exclusion layer was set to 2.0 Å.

Electrostatic potentials were computed using the focusing technique<sup>88</sup> with four nested cubic grids. The grids for the computation of the electrostatic solvation and interaction energies had grid spacings of 2.0 Å, 0.5 Å, 0.2 Å and 0.15 Å, respectively. The outer grid had a grid length of 101 points and was centered on the geometric center of the protein. The following grids were centered on the geometric center of the site. The second grid had a grid length of 241 points. The third grid had a grid length of 345 points. The inner grid had a grid length that was adjusted for each instance of each site separately to fit the dimensions of the site plus 15 Å in each direction. The same grids were used for the model compound and the site in the protein. The grids for the computation of the electrostatic transmembrane potential<sup>64</sup> had grid spacings of 2.0, 0.5, 0.35, and 0.2 Å, respectively. The corresponding grid lengths were 121, 373, 387, and 425 points. All grids were centered on the geometric center of the protein.

**Monte Carlo Simulations.** All MC simulations were carried out with our program suite GMCT.<sup>51</sup> The temperature was set to 298.15 K. The interaction energy cutoffs for pair and triplet moves were set to 1.0 and 2.0 kcal/mol.

Equilibrium probabilities of binding forms for individual sites were computed with the Metropolis MC method.<sup>71,89,90</sup> We used 5000 MC scans for the equilibration and  $10^5$  MC scans for the production run.

Free energy calculations were performed with the free energy perturbation method<sup>72</sup> in our recently presented generalization.<sup>73</sup> We used staging<sup>91</sup> with nine alchemical intermediate states evenly distributed along the transformation coordinate. The Bennett acceptance ratio method was used to minimize the statistical error of the free energy estimates.<sup>74</sup> Each free energy calculation consisted of multiple simulations according to the random single-move simulation scheme.<sup>73</sup> The number of simulations was increased until the statistical error of the free energy estimate was smaller than 0.01 kcal/mol. Each separate

simulation consisted of 1000 MC scans for equilibration and 10 000 to 50 000 MC scans for production.

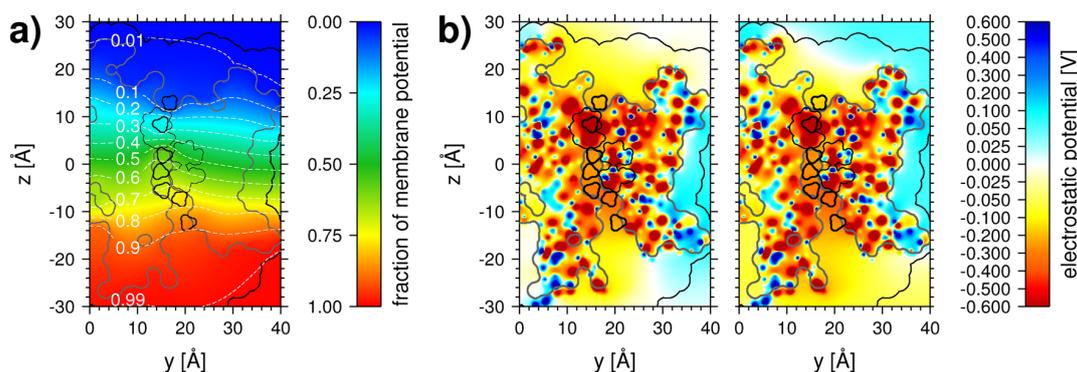
## RESULTS AND DISCUSSION

**Bioenergetics of Transmembrane Transport.** In this section, we study the thermodynamics of the transmembrane transport of ammonia and ammonium from an overall perspective without regard to mechanistic details. The free energy for the import of a permeant is given by the difference in the electrochemical potential of the permeant between the inside phase and the outside phase<sup>64,92–94</sup>

$$\Delta G^{\text{transfer}} = \bar{\mu}^{\text{in}} - \bar{\mu}^{\text{out}} \quad (17)$$

The electrochemical potentials of ammonia and ammonium are calculated as described in the Methods section. The electrochemical potential of the permeant species depends on the activity of the permeant species and the electrostatic potential in the respective bulk solvent phase. Ammonia and ammonium are interconvertible through protonation as described by eq 5. This protonation equilibrium leads to a pH-dependence of the activities of ammonia and ammonium and thus of their electrochemical potentials (Figure S1 of the Supporting Information). Thus, there are three factors that determine the electrochemical potentials of ammonia and ammonium in a given solvent phase. The first factor is the total activity of ammonium and ammonia that is measured by the pN value defined by eq 11. The second factor is the relative abundance of the two protonation forms which is determined by the pH value via  $(a_{\text{NH}_3}/a_{\text{NH}_4^+}) = 10^{\text{pH}-\text{p}K_a}$ , where  $\text{p}K_a = 9.25$ . The third factor is the electrostatic potential in the respective bulk solvent phase determined by the electric transmembrane potential.

Figure 3 shows plots of the transfer free energy for transport of ammonia or ammonium from the extracellular phase to the intracellular phase. Figure 3a shows plots of the transfer free energy as a function of the intracellular pH and the proton-motive force (pmf). Figure 3b contains the corresponding plots at a fixed intracellular pH value of 7.0 as functions of the concentration gradient of ammonia/ammonium across the membrane and the pmf. The pmf is the electrochemical potential difference of the proton across the membrane



**Figure 4.** Electrostatic potential distribution in a cross-section through AfAmt-1 as function of the transverse position  $z$  and the lateral position  $y$  in the projection plane. The orientation of the protein is the same as in Figure 2 and the electrostatic potentials are plotted in the same slice plane. (a) The fraction of the electric transmembrane potential (electric distance) as a function of the coordinates  $y$  and  $z$ . The fraction is color coded (see color bar) and indicated by isocontours. (b) The electrostatic potential under typical conditions (pH 7.0 and  $\Delta\psi = -0.12$  V). The structure of AfAmt-1 was constructed by setting all sites to their most highly populated instances under the specified conditions. The electrostatic potential is color coded (see color bar). left: electrostatic potential contribution due to the protein charge distribution alone. right: total electrostatic potential due to the protein charge distribution and the electric transmembrane potential.

expressed in terms of a voltage difference acting on the charge of the proton<sup>65,92</sup>

$$\text{pmf} = -\frac{\ln 10}{\beta F} \Delta\text{pH} + \Delta\psi \quad (18)$$

where  $F$  is the Faraday constant,  $\Delta\text{pH} = \text{pH}^{\text{in}} - \text{pH}^{\text{out}}$ , and  $\Delta\psi = \psi^{\text{in}} - \psi^{\text{out}}$ . Here,  $\psi$  is the electrostatic potential in the respective bulk solvent phase (far from the membrane). A negative value of the pmf corresponds to the physiological direction of the pmf ( $\psi$  inside negative, outside higher proton activity than inside), while a positive value of the pmf indicates a nonphysiological direction of the pmf. Equation 18 provides a possibility to express the chemical component  $\Delta\text{pH}$  and the electrical component  $\Delta\psi$  of the proton-motive in a common unit. In this way, we can directly compare the effect of the two pmf components on the transfer free energy. Analogously to  $\Delta\text{pH}$ , the concentration gradient of the permeants across the membrane is expressed as  $\Delta\text{pN} = \text{pN}^{\text{in}} - \text{pN}^{\text{out}}$ . A negative value of  $\Delta\text{pN}$  indicates that the permeant is less abundant in the outside phase than in the inside phase. This case corresponds to the physiologically relevant situation for microorganisms and plants that have to acquire ammonium as nitrogen source at low ammonium concentrations in the surrounding medium.<sup>1</sup> Typical ammonium concentrations inside the cell are  $\sim 1$  mM, whereas the ammonium concentration in the surrounding medium is typically significantly lower.<sup>1,14,19,23,95,96</sup>

The left column of Figure 3a shows the effect of the electric transmembrane potential on the transfer free energy. If the electrostatic potential is lower (more negative) in the inside phase than in the outside phase, the import of the positively charged ammonium ion will be thermodynamically favored. In contrast, the import of the uncharged ammonia molecule is unaffected by the electrostatic potential difference between the solvent phases.

The right column of Figure 3a shows the effect of the transmembrane pH difference on the transfer free energy. If the pH value of the outside phase is lower than that of the inside phase, the import of  $\text{NH}_4^+$  will be thermodynamically favored. This favorable effect of the pH gradient arises because a fraction of the ammonium ions will be deprotonated in the more

alkaline inner phase leading to a free energy gain. However, the magnitude of this energy gain will be very small under physiologically relevant conditions where, on both membrane sides, the pH value is close to 7.0 and ammonium is the predominant protonation form. In contrast, the import of  $\text{NH}_3$  is hindered by a transmembrane pH difference in the physiological direction. The free energy cost arises because a part of the imported ammonia is formed by deprotonation of ammonium at the more acidic extracellular side and reprotonated at the more alkaline intracellular side. Thus, the ammonia uniport is, in this case, thermodynamically equivalent to a net antiport of an ammonium ion and a proton, where the ammonium ion is imported and the proton is exported against the direction of the pmf.<sup>4</sup>

Figure 3b shows plots of the transfer free energy for the import of ammonia or ammonium at a fixed intracellular pH value of 7.0 as a function of the transmembrane concentration gradient of the permeant and the pmf. The permeant concentration gradient  $\Delta\text{pN}$  favors the permeant import if the permeant concentration in the outer phase is higher than in the inner phase. Under physiologically relevant conditions for bacteria and plants, however, the concentration of the permeants is lower in the outer phase than in the inner phase. Consequently, the permeant concentration gradient will actually hinder the permeant import by making a positive contribution to the transfer free energy. This unfavorable free energy contribution has to be compensated by coupling the permeant import to the pmf as a driving force. A physiological pmf cannot drive the net import of ammonia as can be seen from the lower row of Figure 3a. However, a physiological pmf can drive the net import of ammonium as can be seen from the upper row of Figure 3a.

In summary, a net uptake of nitrogen at low concentrations of ammonia/ammonium in the extracellular medium is only possible as pmf-driven electrogenic import. A pmf driven transport mode couples the thermodynamically unfavorable import of ammonia to the thermodynamically favorable import of the proton.<sup>4,96</sup> A net electrogenic transport mode can also involve the transport of a part of the substrate as ammonia and another part as ammonium. The import is likely predominantly driven by the electric transmembrane potential since the transfer free energy of ammonium is much more sensitive to

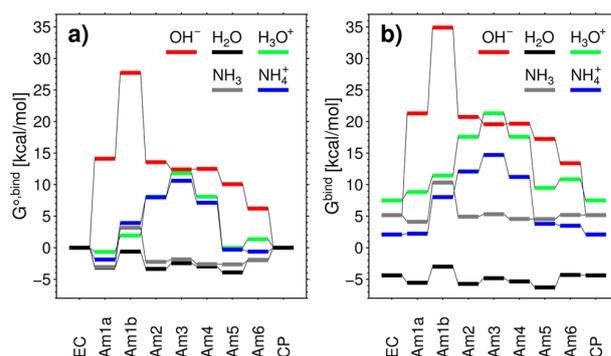
this component of the pmf than to the pH gradient (Figure 3b). Furthermore,  $\Delta\psi$  is also the dominant pmf component at the membranes of bacteria<sup>1,97</sup> and plant root cells.<sup>19</sup>

**Electrostatic Potential Across AfAmt-1.** In this section, we investigate the possible role of the electrostatic potential in the transport mechanism of AfAmt-1. Figure 4 shows plots of the electrostatic potential distribution across AfAmt-1. Contributions arising from the electric transmembrane potential and the protein charge distribution are shown separately and added up at pH 7.0 and  $\Delta\psi = -120$  mV. Figure 4a shows that the transmembrane potential does not exhibit a simple linear dependence on the coordinate transverse to the membrane. This nontrivial dependence is caused by the complex shape of the molecule and the resulting distribution of the dielectric regions.<sup>64</sup> A physiological transmembrane potential arises from the separation of positively and negatively charged ions by the membrane, where a net positive charge resides on the extracellular side and a net negative charge of the same magnitude resides on the intracellular side. Thus, the electrostatic transmembrane potential ranges from  $-1/2\Delta\psi$  in the bulk extracellular phase to  $+1/2\Delta\psi$  in the bulk intracellular phase. The contribution of the electric transmembrane potential to the total electrostatic potential is much lower than the contribution of the protein electrostatic potential for realistic values of  $\Delta\psi$ .

The protein electrostatic potential is negative in the whole transmembrane pore region (Figure 4b). The electrostatic potential in the vestibule regions and the lumen of the pore can thus aid in the stabilization of a positively charged permeant (see also Figures S2 to S4 of the Supporting Information). The negative electrostatic potential at the extracellular vestibule is also likely to increase the rate of substrate uptake from the extracellular phase by attracting positively charged ammonium ions. The negatively charged side chain of D149 makes the largest contribution to the negative electrostatic potential at the extracellular vestibule. A role of the equivalent residue D160 in EcAmtB in the stabilization of the ammonium ion at the extracellular vestibule has been proposed earlier by Luzhkov et al.<sup>36</sup>

**Thermodynamics of Permeant Binding.** In this section, we investigate the different factors that influence the thermodynamics of transferring the permeants from the bulk solvent phases to the permeant sites in the transporter pore. The free energy difference associated with this transfer can be formalized as binding free energy of a particular permeant species at a given permeant site in the transporter pore. We defined the unbound state of a permeant site by occupation of its ligand-free binding form. The bound state is defined by occupation of one of the forms of the permeant site with the respective permeant species bound. By comparing the binding free energies for the different permeant species, we can investigate possible implications for the transport mechanism of AfAmt-1.

Figure 5 shows the binding free energies and standard binding free energies at all permeant sites for all permeants considered (pH 7.0,  $\Delta\psi = 0$  mV, pN 3). The standard binding free energy (Figure 5a) extrapolates the binding free energy to a virtual standard concentration of 1.0 mol/L. The standard binding free energy for the ligand species  $x$  can be calculated from the binding free energy for this ligand by removing the explicit dependence of the binding free energy on the activity of the ligand



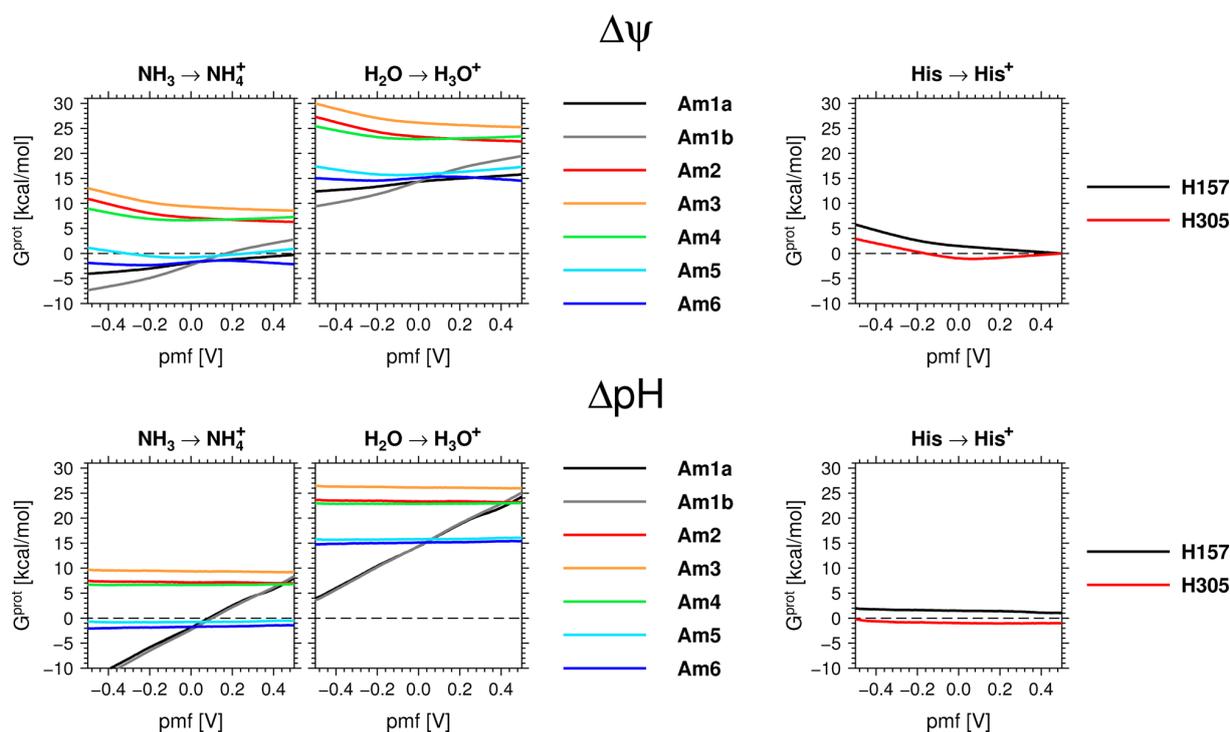
**Figure 5.** Standard binding free energy (a) and binding free energy (b) for different permeants at the transmembrane pore sites. The pH is set to 7.0 on both membrane sides. No pmf is applied. The pN is set to 3.0 on both membrane sides. Connecting lines between the free energy levels are just a guide to the eye and do not imply transition state energies.

$$G_x^{o,bind} = G_x^{bind} - \beta^{-1} \ln a_x \quad (19)$$

In general, there is also an implicit dependence of the standard binding free energy on the ligand activity if there are other, interacting binding sites for the same ligand.<sup>98–100</sup> However, since the permeant sites are almost always occupied by an uncharged water molecule under physiologically relevant conditions, the interactions between the permeant sites are limited. Consequently, the implicit dependence of the standard binding free energy on the ligand activity is small in our case.

The standard binding free energy is directly comparable to potential of mean force profiles which are often used in MD studies to analyze the thermodynamics of transmembrane transport. Potential of mean force profiles for the permeation of ammonia and ammonium along the transmembrane pore of homologues of AfAmt-1 obtained from MD studies have been reported.<sup>32,39,48–50</sup> We compared the potential of mean force at the positions equivalent to our permeant sites to our standard binding free energies. The potential of mean force profiles for the permeation of NH<sub>3</sub> through EcAmtB<sup>32,39,48</sup> and NeRh50<sup>50</sup> are comparable to the profile of our standard binding free energies. The corresponding profiles for the permeation of NH<sub>4</sub><sup>+</sup> obtained from the MD studies show a significantly less favorable thermodynamics of permeation than our standard binding free energies. This discrepancy arises most likely from the presence of water in the pore lumen in our simulation and its absence in the molecular dynamics studies. In their recent work, employing a polarizable force field, Lamoureux et al. found even more favorable standard binding free energies for ammonium in the hydrophobic pore lumen than found here.<sup>49</sup> This finding is based on the explicitly included polarization effect whose predicted magnitude is, however, surprisingly large.

Upon transfer from the bulk solvent into the transmembrane pore, the permeant is partially desolvated. This desolvation leads to a positive contribution to the binding free energy, often termed desolvation penalty, which disfavors binding. This effect is especially pronounced for charged permeant species. Water present in the pore lumen can partially compensate for the lost hydration shell and thus reduce the free energy cost for desolvation of the charged permeant. The desolvation penalty in our simulations is alleviated in this way because the surrounding permeant sites are almost always occupied by



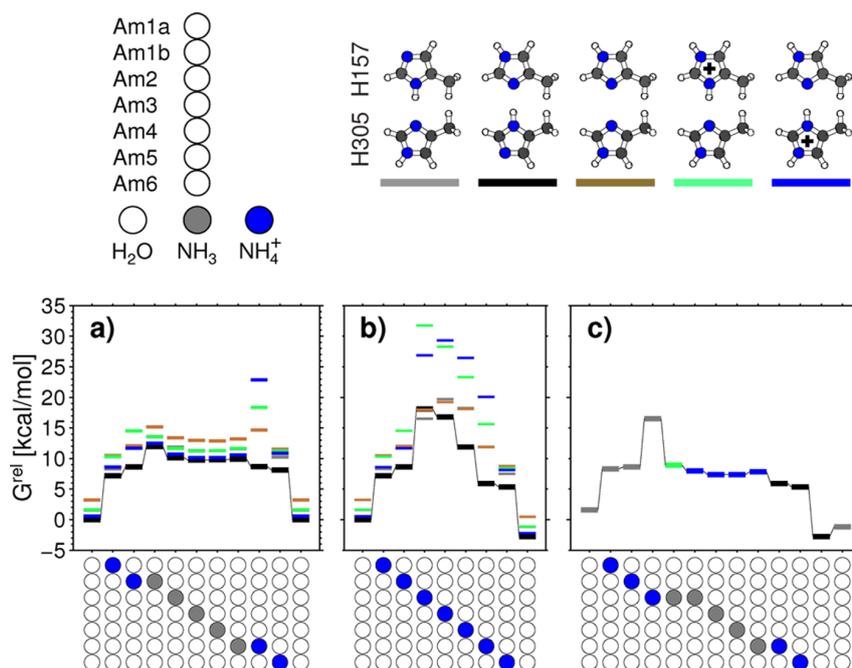
**Figure 6.** Protonation free energy for the twin-His motif and the permeants ammonia and water at the transmembrane pore sites as functions of a proton-motive force. Upper block: pmf consists entirely of an electric transmembrane potential. The pH value is set to 7.0 on both membrane sides. Lower block: pmf consists entirely of a transmembrane pH difference. The intracellular pH value is set to 7.0. The extracellular pH value is given by  $7.0 + \text{pmf}/59$  mV. For both blocks, the pN is set to 3.0 on both membrane sides.

water. Note that the possibility of empty binding sites is included in our model, i.e., the water molecules are free to leave the transporter pore. The question whether or not the pore lumen is occupied by water is thus a very important point in determining the electrostatic barrier for permeation of ammonium through the pore. The importance of water for the stabilization of the protonated  $\text{NH}_4^+$  ion relative to the deprotonated  $\text{NH}_3$  molecule has already been pointed out by Bostick and Brooks<sup>32,33</sup> and Bernèche et al.<sup>5,46</sup> Although intraluminal water is not observed in most of the MD simulations reporting the potential of mean force profiles, other MD studies show stable<sup>46,49</sup> or transient<sup>36,37</sup> hydration of the pore lumen. Most notably, a recent MD simulation of EcAmtB with a polarizable force field also showed stable hydration of the pore.<sup>49</sup> Thus, the protein–water interactions might be influenced by the chosen water model.<sup>101,102</sup> Along these lines, a recent work studied the role of the water model in binding free energy calculations and found that a more detailed 5-site water model like the one used here yielded more reliable predictions than older, three-site water models like those used in the MD studies that did not show hydration of the hydrophobic pore lumen of EcAmtB.<sup>103</sup>

The standard binding free energy directly measures the difference between the interactions of the permeant with the protein environment and the bulk solvent. Thus,  $G^{\text{o,bind}}$  can be used to analyze whether AfAmt-1 has an intrinsic selectivity for a certain permeant species over other permeant species. The standard binding free energies reveal that AfAmt-1 is selective for  $\text{NH}_4^+$  over  $\text{H}_3\text{O}^+$  at the vestibule regions (Am1a, Am1b, and Am7) and at the center of the pore lumen (Am3 and Am4). This selectivity is based on the higher desolvation penalty for

the smaller and more polar  $\text{H}_3\text{O}^+$  ion. In contrast, a selectivity of AfAmt-1 for  $\text{NH}_3$  over  $\text{H}_2\text{O}$  is not apparent from our results, which is perhaps not surprising given the similar physicochemical properties of these molecules. Such a selectivity might, however, also arise from differences in the free energy barriers between the permeant positions which are not included in our calculations. Binding of the  $\text{OH}^-$  ion is very unfavorable because of the high desolvation penalty and the unfavorable electrostatic interaction with the protein (Figure 4b).

The binding free energy extends the standard binding free energy by adding the effect of the actual activities of the different permeant species (Figure 5b). The binding free energy is more favorable if the permeant is more abundant in the respective solvent phase and less favorable if the permeant is less abundant. The activity of the different protonation forms of ammonia and water is determined by the pH value of the respective solvent phase (see Methods section). At neutral pH values,  $\text{NH}_4^+$  is much more abundant than  $\text{NH}_3$ . Similarly,  $\text{H}_2\text{O}$  is the predominant protonation form of water, whereas  $\text{H}_3\text{O}^+$  and  $\text{OH}^-$  are very rare. Thus, the contribution of the permeant activity to the binding free energy strongly favors the binding of  $\text{NH}_4^+$  over that of  $\text{NH}_3$  and the binding of  $\text{H}_2\text{O}$  over that of  $\text{H}_3\text{O}^+$  and  $\text{OH}^-$ . In addition, the total activity of the water species will normally be much higher than that of the ammonia species (typically  $c_{\text{NH}_3} + c_{\text{NH}_4^+} < 10^{-3}$  mol/L while  $c_{\text{H}_2\text{O}} \approx 55.5$  mol/L). Due to its high abundance, the binding of water is much more favorable than the binding of any other permeant species at all permeant sites. Consequently, the permeant sites are almost always occupied by water and only occasionally by the other permeant species. Thus, it seems unlikely that the transport mechanism of AfAmt-1 involves the binding of more



**Figure 7.** Free energy level scheme for states of the permeant sites and the twin-His motif occurring during different proposals for the permeation mechanism. The species present at the permeant sites (Am1a to Am6) are indicated at the horizontal axis (gray, NH<sub>3</sub>; blue, NH<sub>4</sub><sup>+</sup>; white, H<sub>2</sub>O). The free energy of each state is plotted relative to the free energy of the state with all permeant sites occupied by water, H157 adopting the *N* $\delta$ -protonated tautomer and H305 adopting the *N* $\delta$ -protonated tautomer. The energy levels are color coded according to the configuration adopted by the twin-His motif as defined by the legend. Connecting lines between the free energy levels are just a guide to the eye and do not imply transition state energies. The pH is set to 7.0 on both membrane sides. The pN is set to 3.0 on both membrane sides.  $\Delta\psi$  is set to  $-0.12$  V. (a) Electroneutral NH<sub>3</sub> uniport; (b) electrogenic NH<sub>4</sub><sup>+</sup> uniport not involving protonation state changes of the twin-His motif; (c) electrogenic NH<sub>3</sub>/H<sup>+</sup> symport involving protonation state changes of the twin-His motif. It is not clear how the restoration of the twin-His motif's configuration between the last two stages of this mechanistic scheme will occur microscopically. The restoration might involve a transient reprotonation of the twin-His motif, proton transfer along the water molecules in the pore lumen and/or a transient reorientation of the histidine side chains.

than one molecule of the permeants ammonia or ammonium per transporter monomer at the same time. Crystallographic studies showed residual electron density in the pore lumen of EcAmtB.<sup>16,17,34,35</sup> As noted earlier by Zheng et al.<sup>17</sup> and ourselves,<sup>18</sup> it is impossible to reliably distinguish between water and ammonia as occupants of the permeant sites on the basis of the experimentally determined electron density. Bernèche and co-workers pointed out the importance of the high water concentration for the question whether the permeant sites are occupied by ammonium/ammonia or water.<sup>5,17</sup> These considerations and the findings from our calculations argue for water as predominant occupant of the permeant sites at least for AfAmt-1 and possibly also for its structurally highly similar homologues. Plots of the binding free energies for all permeant species at all permeant sites as a function of the pmf components can be found in section F of the Supporting Information.

**Protonation Free Energies of the Permeant Sites and the twin-His Motif.** In this section, we study the protonation free energies of ammonia and water at the permeant sites and of the twin-His residues. Figure 6 shows plots of these protonation free energies as functions of an electric transmembrane potential or a transmembrane pH gradient. The settings chosen in the calculation of the protonation free energies are the same as those used in the calculation of the permeant binding free energies above. The protonation free energies of the twin-His residues and of water and ammonia at the intracellular permeant sites (Am2 to Am7) are only weakly

dependent on the pH gradient because the intracellular pH value is constant. The protonation free energies of NH<sub>3</sub> and H<sub>2</sub>O at the extracellular vestibule (Am1a and Am1b) are roughly linearly dependent on the pH gradient. This dependence stems from the dependence of the extracellular pH value, and thus of the extracellular proton activity, on the pH gradient.

The effect of the electric transmembrane potential on the protonation free energies arises from its influence on the intrinsic energies of the binding sites and on the proton electrochemical potentials of the solvent phases. The transmembrane potential favors the release of protons to the solvent phase with the more negative potential and the uptake of protons from the solvent phase with the more positive potential. This is why the transmembrane potential has a qualitatively opposite effect on the protonation free energy of intracellular and extracellular binding sites. The effect of the transmembrane potential increases with the distance of the binding site from the respective outer surface of the protein (Figure 4a). This dependence can be understood from the difference in the electrostatic potential between the respective bulk solvent phase and the position of the binding site. A larger electrostatic potential difference will lead to a larger magnitude of the energy that is gained or spent in transferring the proton from the bulk solvent to the binding site.

The protonation free energy of water is less favorable than the protonation free energy of ammonia at all permeant sites. This difference is mainly caused by the lower intrinsic proton affinity of water relative to ammonia. An additional

contribution arises from the higher desolvation penalty for the smaller and more polar hydronium ion relative to the ammonium ion. The protonation free energy of the permeants at the vestibular permeant sites (Am1a, Am1b, Am6, and Am7) is much more favorable than at the luminal permeant sites (Am2, Am3, and Am4). The protonation free energy of water at the luminal permeant sites is so unfavorable that it seems unlikely that a hydronium ion at the luminal sites is involved in the conduction mechanism. In contrast, the lower protonation free energies of ammonia let it seem possible that an ammonium ion can be transiently stabilized at the intraluminal permeant sites. However, the protonation free energies indicate that the permanent protonation of ammonia is only favorable at the vestibular permeant sites but not at the intraluminal permeant sites. This finding is consistent with the findings of previous theoretical studies on EcAmtB<sup>32,33,36–48</sup> and NeRh50<sup>50</sup> and the binding of xenon to the intraluminal permeant sites of xenon-pressurized AfAmt-1 crystals.<sup>18</sup>

The twin-His residues are the only titratable sites of the protein whose protonation probability is strongly dependent on the binding of ammonia/ammonium to the permeant sites (see plots in section E of the Supporting Information). The possibilities of an active role of the twin-His motif as transient proton acceptor for the substrate ammonium<sup>5,17,34,35</sup> or a passive role as hydrogen bond donor to the permeants passing the pore lumen have been noted earlier.<sup>16</sup> The protonation free energies of the twin-His residues have small magnitudes over the whole range of conditions studied. The small free energy differences between the protonated and deprotonated forms of these residues indicate that transient changes between the protonation states are easily possible. Thus, the twin-His motif could play an active role in a possible electrogenic conduction mechanism of AfAmt-1. This possibility is investigated further in the next section.

**Thermodynamics of Possible Permeation Mechanisms.** For investigating the thermodynamics of possible permeation mechanisms, we concentrate on the permeant sites and the two histidine residues forming the twin His-motif. We considered microscopic states for these sites, while all other sites can freely equilibrate. For the twin-His motif, we considered all five thermodynamically viable permutations of the histidine protonation forms as depicted in the legend of Figure 7. We computed the free energy of all states with one or two permeants other than water relative to the fully water occupied state with the thermodynamically most favorable protonation form of the twin His motif (H157 singly protonated at the N $\epsilon$  atom and H305 singly protonated at the N $\delta$  atom). The full sets of computed relative free energies of the states are depicted in Figures S26 to S29 of the Supporting Information. Occupation of more than two sites at the same time by other permeant species than water is thermodynamically very unfavorable and thus unlikely to be involved in the permeation mechanism. Figure 7 shows free energy profiles for the possible permeation mechanisms considered herein. We did not calculate free energy barriers between the intermediate states of the permeation mechanism. Potential of mean force profiles for the permeation of NH<sub>3</sub> and NH<sub>4</sub><sup>+</sup> through EcAmtB are available from MD studies.<sup>32,33,48,49</sup> These profiles indicate that the barriers for permeant translocation between the permeant sites are small relative to the overall electrostatic barrier for the permeation of the pore lumen by the ammonium ion. Thus, we believe that the

inclusion of these barriers is of minor importance for evaluating the thermodynamic viability of the transport mechanisms.

The parameters in the calculations were set to standard values that could be encountered at the membrane of microorganisms. The pH value was set to 7.0 on both membrane sides. The electric transmembrane potential was set to  $-0.12$  V. A transmembrane pH gradient was not applied because its effect on the transport and binding thermodynamics is much smaller than that of the electric transmembrane potential (see above). Typically,  $\Delta\psi$  is also the dominant component of the pmf.<sup>1,19,97</sup> The pN value was set to 3.0 on both membrane sides. Figure 7 shows the free energy profiles of possible conduction mechanisms. Taking into account the permeant concentrations and the electrochemical transmembrane gradients has a significant effect on the free energy profiles. For comparison, Figure S30 of the Supporting Information shows the same free energy profiles at standard total concentration of the permeants and at zero electrochemical transmembrane gradients.

Figure 7a shows the electroneutral NH<sub>3</sub> uniport mechanism first proposed in microscopic detail by Khademi et al. on the basis of the hydrophobicity of the transmembrane pore.<sup>16</sup> This mechanism was supported by previous theoretical studies on EcAmtB<sup>32,33,36–48</sup> and NeRh50<sup>50</sup> which concentrated on the ammonia or ammonium uniport mechanisms. The NH<sub>3</sub> uniport mechanism does not involve protonation state changes of the twin-His motif. Ammonium is recruited at Am1a and deprotonated somewhere close to Am1b. The proton is left behind in the extracellular phase. The neutral ammonia molecule permeates the hydrophobic pore lumen consisting of Am2, Am3 and Am4, and is reprotonated with a proton from the intracellular phase at Am5. Finally, the ammonium ion is released to the intracellular phase. The reasoning for the deprotonation of NH<sub>4</sub><sup>+</sup> at Am1b and its reprotonation at Am5 is based on the high protonation free energy of ammonia at the intervening intraluminal permeant sites.

Figure 7b shows an electrogenic NH<sub>4</sub><sup>+</sup> uniport mechanism that does not involve protonation state changes of the twin-His motif (net transport of NH<sub>4</sub><sup>+</sup>). The ammonium ion is recruited from the extracellular phase at Am1a, permeates the whole transmembrane pore passing the intermediate permeant sites and is released from Am7 to the intracellular phase. The permeation of the transmembrane pore is associated with an electrostatic energy barrier of ca. 15 kcal/mol.

Figure 7c shows a possible electrogenic NH<sub>3</sub>/H<sup>+</sup> symport mechanism (net transport of NH<sub>4</sub><sup>+</sup>). The most important features of this mechanism have been proposed by Bernèche and co-workers.<sup>5</sup> Ammonium is recruited from the extracellular phase at Am1a, loses its hydration shell at Am1b and is transferred to Am2. At Am2, the ammonium ion is deprotonated by H157. An escape of the NH<sub>3</sub> molecule back to the intracellular vestibule might be hindered by the preferential permeability of the Phe gate for NH<sub>4</sub><sup>+</sup> relative to NH<sub>3</sub> reported by Bernèche and co-workers.<sup>5</sup> The protonation of the twin-His motif is stabilized by the hydrogen bond between the two central N $\delta$  nitrogen atoms. This hydrogen bond also enables the twin-His motif to easily change between the two protonated configurations (green and blue in Figure 7) via proton transfer between the two central N $\delta$  nitrogen atoms. In this way, the proton is transferred from H157 to H305 at the intracellular side of the pore. The neutral ammonia molecule permeates the hydrophobic pore lumen and is reprotonated by H305 at Am5. Finally, the ammonium ion is transferred to Am6

and released to the intracellular phase. This mechanism has a slightly lower electrostatic barrier for the permeation of the pore lumen than the  $\text{NH}_4^+$  uniport mechanism (ca. 12 kcal/mol). As noted by Bernèche et al.,<sup>5</sup> this mechanism poses the problem of how to reset the tautomeric forms of the twin-His residues to those occupied initially. The reset could involve a transient reprotonation of the twin-His motif from the intracellular phase. In addition, the restoration might involve proton transfer along the water molecules in the pore lumen in a Grotthuss mechanism and/or a transient reorientation of the histidine side chains. The uncertainty about the precise mechanism of the reset makes it impossible at present to quantify the size of the associated free energy barrier with confidence. Based on gas-phase QM calculations, Lamoureux et al. estimate the free energy barrier for this reset to be no larger than 15 kcal/mol if the reset involves water molecules in the pore and a transiently formed  $\text{OH}^-$ .<sup>49</sup> As discussed above, the formation of a  $\text{OH}^-$  ion is thermodynamically very unfavorable within the protein environment (see Thermodynamics of Permeant Binding section). Within our model, the barrier for the transient formation of the  $\text{OH}^-$  ion would be about 22 kcal/mol (Figures S26 and S28 of the Supporting Information). This high free energy barrier might, however, be lowered if polarization and quantum effects are explicitly included. A reset mechanism that involves reorientation of the twin-His side chains is found by Lamoureux et al. to be unlikely because the associated barrier is larger than 20 kcal/mol.

Which transport mechanism is now likely to be operative? The low overall free energy barrier of 12 kcal/mol associated with the electroneutral  $\text{NH}_3$  uniport mechanism indicates that the mechanism is kinetically feasible. An important counter argument against an  $\text{NH}_3$  uniport as exclusively operative mechanism is the lack of a net thermodynamic driving force under physiologically relevant conditions (see Bioenergetics of Transmembrane Transport section). In contrast, an  $\text{NH}_3$  uniport is even hindered if a physiological pH gradient is applied in addition to the electric transmembrane potential (Figure 3); that is, ammonia would be exported instead of imported. With a pH gradient of  $\Delta\text{pH} = 1$  pH unit, the ammonia import would be hindered by 1.3 kcal/mol. In contrast, an electrogenic transport mechanism involving  $\text{NH}_4^+$  can be driven by the electric transmembrane potential (Figure 3).

Amt/Rh proteins of microorganisms, plants and mycorrhizal fungi seem to be expressed only if the availability of nitrogen, sources is low, i.e., if the ammonium concentration in the medium is low.<sup>3,104–107</sup> This finding underlines the necessity for these life forms to drive the ammonium import by the electric transmembrane potential as pointed out earlier by ourselves and other authors.<sup>1,4,14,18</sup> The free energy schemes for the electrogenic mechanisms indicate a sizable total free energy barrier of 15–18 kcal/mol for the permeation of the transporter pore. The main contribution to this free energy barrier is the cost of the electrostatic desolvation of the ammonium ion upon transfer from the aqueous solvent to the hydrophobic pore lumen (12–15 kcal/mol). The desolvation penalty might be lowered further when explicitly accounting for electronic polarization, which is especially significant in the interaction of aromatic residues with  $\text{NH}_4^+$ .<sup>49,75</sup> The electrogenic mechanisms seem to be thermodynamically feasible especially when considering the high optimal growth temperature of *A. fulgidus* of ca. 80 °C.<sup>108,109</sup> The similar magnitude of the free energy barriers for the electrogenic mechanism considered and

the uncertainty about the free energy barrier associated with the last step of the  $\text{NH}_3/\text{H}^+$  symport mechanism do not allow us to prefer one of the mechanism on the basis of the free energy profiles.

Although our calculations concentrated on AfAmt-1 and did not include other homologues, we do believe that many arguments are transferable to other homologues. This belief is based on the high structural similarity of Amt and Rh proteins revealed by the available crystal structures.<sup>13,16–18,35,110,111</sup> An electrogenic transport was reported on the basis of experiments on plant<sup>19–24</sup> and microbial Amts<sup>25–27</sup> and on human Rh proteins<sup>30</sup> implying the involvement of  $\text{NH}_4^+$  as substrate. In seeming contradiction, an involvement of  $\text{NH}_3$  as substrate was implied by experiments reported for EcAmtB<sup>16</sup> and some Rh proteins.<sup>28,29,112,113</sup> These experiments indicated an intracellular alkalinization accompanying the transport which was interpreted to originate from reprotonation of the transported  $\text{NH}_3$  in the intracellular phase.<sup>16</sup> The significance of these findings for the mechanism of Amt/Rh proteins is, however, debated.<sup>4,31</sup> As pointed out earlier by Musa-Aziz et al., an alkalinization of the intracellular phase occurs already if a fraction of the overall imported ammonia/ammonium is transported as neutral ammonia.<sup>11,114</sup> In addition, the experiments showed that the transport was stimulated by slight increases in the extracellular pH value while keeping a pH value well below the  $\text{pK}_a$  of ammonium. Under these conditions, a pH increase will only have a significant effect on the abundance and availability of  $\text{NH}_3$  while the abundance of  $\text{NH}_4^+$  is merely affected. Bakouh et al. observed sensitivity of the transport through human RhCG to the transmembrane potential and to the extracellular pH value.<sup>30</sup> These findings lead them to propose that  $\text{NH}_3$  and  $\text{NH}_4^+$  are directly involved in the transport mechanism. The function of molecular systems is inherently stochastic. Thus, the mechanistic schemes shown in Figure 7 may be seen as limiting cases. The electroneutral  $\text{NH}_3$  uniport may also contribute to the total permeant flux across the membrane as long as a parallel electrogenic transport supplies an overall thermodynamic driving force.

The prominent role attributed to the twin-his motif in the proposed  $\text{NH}_3/\text{H}^+$  symport mechanism is consistent with the high degree of its conservation among Amt/Rh proteins.<sup>2,5,34,115</sup> The histidine residues can in some cases be substituted by glutamate or aspartate<sup>34,115</sup> which might be able to perform the same function.<sup>49</sup> There are, however, also some rare cases as in some Rh proteins where no equivalent of the twin-His motif is present.<sup>12</sup>

From our calculations, it can not be excluded that different transport mechanisms are operative in different Amt/Rh proteins as advocated by some workers.<sup>9,10,116,117</sup> It is, however, also conceivable that the Amt/Rh proteins can support different transport modes. The transport mechanism actually operational would then depend on the driving force available in a particular system. More work is required to gain certainty in this respect.

## CONCLUSIONS AND OUTLOOK

We investigated the thermodynamics of different possible transport mechanisms of AfAmt-1. Both, electrogenic and electroneutral permeation mechanisms were found to be thermodynamically viable. A net transport of substrate across the membrane requires the presence of a thermodynamic driving force.<sup>64,92,93</sup> Microorganisms like *A. fulgidus* and plants have to acquire ammonium at low external ammonium concentrations.<sup>1,14,19,23,95,96</sup> We found that, in this case, the

driving force for the substrate import can only be provided by the electric transmembrane potential implying an electrogenic transport mode. We computed free energy profiles for different electrogenic transport mechanisms and found that  $\text{NH}_4^+$  uniport and  $\text{NH}_3/\text{H}^+$  symport are thermodynamically viable. Important factors that allow the permeation of the pore lumen by ammonium are the negative electrostatic potential of the protein and water present in the pore lumen. These factors alleviate the free energy cost of desolvating the ammonium ion upon transfer from the solvent to the hydrophobic pore lumen. The possible  $\text{NH}_3/\text{H}^+$  symport mechanism involves the genetically conserved twin-His motif as intermediate acceptor of the cotransported proton. Further experimental and theoretical work is needed to discern between the different possible electrogenic mechanisms and to obtain a deeper understanding of their microscopic details. In particular, it would be interesting to investigate the stoichiometry of protons and ammonia transported as functions of the physical conditions such as pH value, electric transmembrane potential, and transmembrane pH gradient. Theoretical methods that provide time information could be used to study microscopic details of the transport kinetics. Extensions of our microstate description to model the kinetics of molecular systems have already been applied successfully and might be used fruitfully in future studies of AfAmt-1.<sup>58,118–121</sup>

The formalism developed herein for the treatment of ligands with multiple binding forms is generally applicable. Often, the ligand itself can occur in multiple protonation forms which leads to a coupling of the protonation equilibria of both binding partners and the receptor–ligand binding equilibrium. The coupling of protein–ligand binding equilibria to the protonation equilibria of the binding partners occurring in such cases is currently receiving increasing attention.<sup>57,122–134</sup> Our approach is thus also interesting within a wider scope of binding free energy calculations in general, for example in combination with a constant pH  $\lambda$ -dynamics method.<sup>135–137</sup>

## ■ ASSOCIATED CONTENT

### ● Supporting Information

Parameters used in the simulations, derivations of expressions for the chemical potentials and the transfer free energies of the investigated permeants, and additional figures referenced in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [thomas.ullmann@uni-bayreuth.de](mailto:thomas.ullmann@uni-bayreuth.de); [andrade@bio.chemie.uni-freiburg.de](mailto:andrade@bio.chemie.uni-freiburg.de); [matthias.ullmann@uni-bayreuth.de](mailto:matthias.ullmann@uni-bayreuth.de).

### Notes

The authors declare no competing financial interest.

## ■ REFERENCES

- (1) Kleiner, D. *FEMS Microbiol. Lett.* **1985**, *32*, 87–100.
- (2) Howitt, S. M.; Udvardi, M. K. *BBA-Biomembr.* **2000**, *1465*, 152–170.
- (3) Wirén, N.; Merrick, M. Regulation and function of ammonium carriers in bacteria, fungi, and plants. In *Molecular mechanisms controlling transmembrane transport*; Springer: Berlin, 2004; Vol. 9, pp 95–120.
- (4) Andrade, S. L. A.; Einsle, O. *Mol. Membr. Biol.* **2007**, *24*, 357–365.
- (5) Lamoureux, G.; Javelle, A.; Baday, S.; Wang, S.; Bernèche, S. *Transfus. Clin. Biol.* **2010**, *17*, 168–175.
- (6) Khademi, S.; Stroud, R. M. *Physiol.* **2006**, *21*, 419–429.
- (7) Khademi, S.; Stroud, R. M. Gas channels for ammonia. In *Structural biology of membrane proteins*; The Royal Society of Chemistry: London, 2006; pp 212–234.
- (8) Winkler, F. *Pflugers Arch. Eur. J. Phys.* **2006**, *451*, 701–707.
- (9) Ludewig, U.; Neuhäuser, B.; Dynowski, M. *FEBS Lett.* **2007**, *581*, 2301–2308.
- (10) Ludewig, U.; von Wirén, N.; Rentsch, D.; Frommer, W. *Genome Biol.* **2001**, *2*, reviews1010.1–reviews1010.5.
- (11) Boron, W. F. *Exp. Physiol.* **2010**, *95*, 1107–1130.
- (12) Conroy, M. J.; Bullough, P. A.; Merrick, M.; Avent, N. D. *Br. J. Haematol.* **2005**, *131*, 543–551.
- (13) Gruswitz, F.; Chaudhary, S.; Ho, J. D.; Schlessinger, A.; Pezeshki, B.; Ho, C.-M.; Sali, A.; Westhoff, C. M.; Stroud, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 9638–9643.
- (14) Booger, F. C.; Ma, H.; Bruggeman, F. J.; van Heeswijk, W. C.; García-Contreras, R.; Molenaar, D.; Krab, K.; Westerhoff, H. V. *FEBS Lett.* **2011**, *585*, 23–28.
- (15) Soupene, E.; He, L.; Yan, D.; Kustu, S. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 7030–7034.
- (16) Khademi, S.; O'Connell, J.; Remis, J.; Robles-Colmenares, Y.; Miercke, L. J. W.; Stroud, R. M. *Science* **2004**, *305*, 1587–1594.
- (17) Zheng, L.; Kostrewa, D.; Bernèche, S.; Winkler, F. K.; Li, X.-D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17090–17095.
- (18) Andrade, S. L. A.; Dickmanns, A.; Ficner, R.; Einsle, O. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14994–14999.
- (19) Wang, M. Y.; Glass, A. D. M.; Shaff, J. E.; Kochian, L. V. *Plant Physiol.* **1994**, *104*, 899–906.
- (20) Ludewig, U.; von Wirén, N.; Frommer, W. B. *J. Biol. Chem.* **2002**, *277*, 13548–13555.
- (21) Mayer, M.; Dynowski, M.; Ludewig, U. *Biochem. J.* **2006**, *396*, 431–437.
- (22) Mayer, M.; Schaaf, G.; Mouro, I.; Lopez, C.; Colin, Y.; Neumann, P.; Cartron, J.-P.; Ludewig, U. *J. Gen. Physiol.* **2006**, *127*, 133–144.
- (23) Søgaard, R.; Alsterfjord, M.; MacAulay, N.; Zeuthen, T. *Pflugers Arch. Eur. J. Phys.* **2009**, *458*, 733–743.
- (24) Ortiz-Ramirez, C.; Mora, S. I.; Trejo, J.; Pantoja, O. *J. Biol. Chem.* **2011**, *286*, 31113–31122.
- (25) Siewe, R. M.; Weil, B.; Burkovski, A.; Eikmanns, B. J.; Eikmanns, M.; Krämer, R. *J. Biol. Chem.* **1996**, *271*, 5398–5403.
- (26) Meier-Wagner, J.; Nolden, L.; Jakoby, M.; Siewe, R.; Krämer, R.; Burkovski, A. *Microbiol.* **2001**, *147*, 135–143.
- (27) Walter, B.; Kuspert, M.; Ansorge, D.; Krämer, R.; Burkovski, A. *J. Bacteriol.* **2008**, *190*, 2611–2614.
- (28) Westhoff, C. M.; Ferreri-Jacobia, M.; Mak, D.-O. D.; Foskett, J. K. *J. Biol. Chem.* **2002**, *277*, 12499–12502.
- (29) Ludewig, U. *J. Physiol.* **2004**, *559*, 751–759.
- (30) Bakouh, N.; Benjelloun, F.; Hulin, P.; Brouillard, F.; Edelman, A.; Chérif-Zahar, B.; Planelles, G. *J. Biol. Chem.* **2004**, *279*, 15975–15983.
- (31) Javelle, A.; Lupo, D.; Li, X.-D.; Merrick, M.; Chami, M.; Ripoche, P.; Winkler, F. K. *J. Struct. Biol.* **2007**, *158*, 472–481.
- (32) Bostick, D. L.; Brooks, C. L., III *PLoS Comput. Biol.* **2007**, *3*, e22.
- (33) Bostick, D. L.; Brooks, C. L., III *Biophys. J.* **2007**, *92*, L103–L105.
- (34) Javelle, A.; Lupo, D.; Zheng, L.; Li, X.-D.; Winkler, F. K.; Merrick, M. *J. Biol. Chem.* **2006**, *281*, 39492–39498.
- (35) Javelle, A.; Lupo, D.; Ripoche, P.; Fulford, T.; Merrick, M.; Winkler, F. K. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5040–5045.
- (36) Luzhkov, V. B.; Almláf, M.; Nervall, M.; Åqvist, J. *Biochemistry* **2006**, *45*, 10807–10814.
- (37) Nygaard, T. P.; Rovira, C.; Peters, H.; Jensen, M. Ø. *Biophys. J.* **2006**, *91*, 4401–4412.
- (38) Liu, Y.; Hu, X. *J. Phys. Chem. A* **2006**, *110*, 1375–1381.

- (39) Lin, Y.; Cao, Z.; Mo, Y. *J. Am. Chem. Soc.* **2006**, *128*, 10876–10884.
- (40) Cao, Z.; Mo, Y.; Thiel, W. *Angew. Chem., Int. Ed.* **2007**, *46*, 6811–6815.
- (41) Yang, H.; Xu, Y.; Zhu, W.; Chen, K.; Jiang, H. *Biophys. J.* **2007**, *92*, 877–885.
- (42) Ishikita, H.; Knapp, E.-W. *J. Am. Chem. Soc.* **2007**, *129*, 1210–1215.
- (43) Ishikita, H. *FEBS Lett.* **2007**, *581*, 4293–4297.
- (44) Lin, Y.; Cao, Z.; Mo, Y. *J. Phys. Chem. B* **2009**, *113*, 4922–4929.
- (45) Wang, J.; Yang, H.; Zuo, Z.; Yan, X.; Wang, Y.; Luo, X.; Jiang, H.; Chen, K.; Zhu, W. *J. Phys. Chem. B* **2010**, *114*, 15172–15179.
- (46) Lamoureux, G.; Klein, M. L.; Bernèche, S. *Biophys. J.* **2007**, *92*, L82–L84.
- (47) Nygaard, T. P.; Alfonso-Prieto, M.; Peters, G. H.; Jensen, M. Ø.; Rovira, C. *J. Phys. Chem. B* **2010**, *114*, 11859–11865.
- (48) Akgun, U.; Khademi, S. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3970–3975.
- (49) Wang, S.; Orabi, E. A.; Baday, S.; Bernèche, S.; Lamoureux, G. *J. Am. Chem. Soc.* **2012**, *134*, 10419–10427.
- (50) Hub, J. S.; Winkler, F. K.; Merrick, M.; de Groot, B. L. *J. Am. Chem. Soc.* **2010**, *132*, 13251–13263.
- (51) Ullmann, R. T.; Ullmann, G. M. *J. Comput. Chem.* **2012**, *33*, 887–900.
- (52) Ullmann, G. M.; Knapp, E. W. *Eur. Biophys. J.* **1999**, *28*, 533–551.
- (53) Ullmann, G. M.; Kloppmann, E.; Essigke, T.; Krammer, E.-M.; Klingen, A. R.; Becker, T.; Bombarda, E. *Photosynth. Res.* **2008**, *97*, 33–53.
- (54) Klingen, A. R.; Bombarda, E.; Ullmann, G. M. *Photochem. Photobiol. Sci.* **2006**, *5*, 588–596.
- (55) Bashford, D. *Front. Biosci.* **2004**, *9*, 1082–1099.
- (56) Gunner, M. R.; Mao, J.; Song, Y.; Kim, J. *Biochim. Biophys. Acta* **2006**, *1757*, 942–968.
- (57) Antosiewicz, J. M.; Shugar, D. *Mol. Biosyst.* **2011**, *7*, 2923–2949.
- (58) Becker, T.; Ullmann, R. T.; Ullmann, G. M. *J. Phys. Chem. B* **2007**, *111*, 2957–2968.
- (59) Ferreira, A. M.; Bashford, D. *J. Am. Chem. Soc.* **2006**, *128*, 16778–16790.
- (60) Warshel, A. *Biochemistry* **1981**, *20*, 3167–3177.
- (61) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219–10225.
- (62) Calimet, N.; Ullmann, G. M. *J. Mol. Biol.* **2004**, *339*, 571–589.
- (63) Bombarda, E.; Becker, T.; Ullmann, G. M. *J. Am. Chem. Soc.* **2006**, *128*, 12129–12139.
- (64) Roux, B. *Biophys. J.* **1997**, *73*, 2980–2989.
- (65) Azzone, G.; Benz, R.; Bertl, A.; Colombini, M.; Crofts, A.; Dilley, R.; Dimroth, P.; Dutton, P.; Felle, H.; Harold, F.; et al. *BBA-Bioenerg.* **1993**, *1183*, 1–3.
- (66) Brünger, A. T.; Karplus, M. *Proteins* **1988**, *4*, 148–156.
- (67) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (68) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (69) Ullmann, R. T.; Ullmann, G. M. *J. Phys. Chem. B* **2011**, *115*, 10346–10359.
- (70) Bashford, D. An object-oriented programming suite for electrostatic effects in biological molecules An experience report on the MEAD project. In *Scientific computing in object-oriented parallel environments*; Ishikawa, Y., Oldehoeft, R., Reynders, J., Tholburn, M., Eds.; Springer: Berlin, Germany, 1997; Vol. 1343, pp 233–240.
- (71) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H. *J. Chem. Phys.* **1953**, *21*, 1087–1092.
- (72) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
- (73) Ullmann, R. T.; Ullmann, G. M. *J. Phys. Chem. B* **2011**, *115*, 507–521.
- (74) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (75) Orabi, E. A.; Lamoureux, G. *J. Chem. Theory Comp.* **2012**, *8*, 182–193.
- (76) Ullmann, R. T.; Ullmann, G. M. unpublished results.
- (77) Bashford, D.; Case, D. A.; Dalvit, C.; Tennant, L.; Wright, P. E. *Biochemistry* **1993**, *32*, 8045–8056.
- (78) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931–967.
- (79) Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Acc.* **1998**, *99*, 391–403.
- (80) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- (81) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (82) Perdew, J. P.; Yue, W. *Phys. Rev. B* **1986**, *33*, 8800–8802.
- (83) Swart, M.; Th., V. D. P.; Snijders, J. G. *J. Comput. Chem.* **2001**, *22*, 79–88.
- (84) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (85) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *186*, 671–679.
- (86) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1987**, *9*, 327–335.
- (87) Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. *Bioinformatics* **2006**, *22*, 623–625.
- (88) Klapper, I.; Fine, R.; Sharp, K. A.; Honig, B. H. *Proteins* **1986**, *1*, 47–59.
- (89) Beroza, P.; Fredkin, D. R.; Okamura, M. Y.; Feher, G. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5804–5808.
- (90) Rabenstein, B.; Ullmann, G. M.; Knapp, E. W. *Eur. Biophys. J.* **1998**, *27*, 626–637.
- (91) Valleau, J. P.; Card, D. N. *J. Chem. Phys.* **1972**, *57*, 5457–5462.
- (92) Mitchell, P. *BBA-Bioenerg.* **2011**, *1807*, 1507–1538.
- (93) Rottenberg, H. *BBA-Bioenerg.* **1979**, *549*, 225–253.
- (94) Skulachev, V. P. *Eur. J. Biochem.* **1992**, *208*, 203–209.
- (95) Britto, D. T.; Glass, A. D.; Kronzucker, H. J.; Siddiqi, M. Y. *Plant Physiol.* **2001**, *125*, 523–526.
- (96) Schmidt, L.; Look, C.; Bock, E.; Jetten, M. S. M. *Microbiology* **2004**, *150*, 1405–1412.
- (97) Kashket, E. R. *Annu. Rev. Microbiol.* **1985**, *39*, 219–242.
- (98) Ullmann, G. M. *J. Phys. Chem. B* **2000**, *104*, 6293–6301.
- (99) Ullmann, G. M. *J. Phys. Chem. B* **2003**, *107*, 1263–1271.
- (100) Bombarda, E.; Ullmann, G. M. *J. Phys. Chem. B* **2010**, *114*, 1994–2003.
- (101) Huggins, D. J. *J. Chem. Phys.* **2012**, *136*, 064518.
- (102) Huggins, D. J. *J. Comput. Chem.* **2012**, *33*, 1383–1392.
- (103) Fadda, E.; Woods, R. J. *J. Chem. Theory Comp.* **2011**, *7*, 3391–3398.
- (104) Kleiner, D. *FEBS Lett.* **1985**, *187*, 237–239.
- (105) Gazzarrini, S.; Lejay, L.; Gojon, A.; Ninnemann, O.; Frommer, W. B.; von Wirén, N. *Plant Cell* **1999**, *11*, 937–948.
- (106) Rawat, S. R.; Silim, S. N.; Kronzucker, H. J.; Siddiqi, M. Y.; Glass, A. D. M. *Plant J.* **1999**, *19*, 143–152.
- (107) Pérez-Tienda, J.; Testillano, P. S.; Balestrini, R.; Fiorilli, V.; Azzón-Aguilar, C.; Ferrol, N. *Fungal Genet. Biol.* **2011**, *48*, 1044–1055.
- (108) Beeder, J.; Nilsen, R. K.; Rosnes, J. T.; Torsvik, T.; Lien, T. *Appl. Environ. Microbiol.* **1994**, *60*, 1227–1231.
- (109) Klenk, H.-P.; Clayton, R. A.; Tomb, J.-F.; White, O.; Nelson, K. E.; Ketchum, K. A.; Dodson, R. J.; Gwinn, M.; Hickey, E. K.; Peterson, J. D.; et al. *Nature* **1997**, *390*, 364–370.
- (110) Li, X.; Jayachandran, S.; Nguyen, H.-H. T.; Chan, M. K. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 19279–19284.
- (111) Lupo, D.; Li, X.-D.; Durand, A.; Tomizaki, T.; Cherif-Zahar, B.; Matassi, G.; Merrick, M.; Winkler, F. K. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 19303–19308.
- (112) Ripoche, P.; Bertrand, O.; Gane, P.; Birkenmeier, C.; Colin, Y.; Cartron, J.-P. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17222–17227.
- (113) Mouro-Chanteloup, I.; Cochet, S.; Chami, M.; Genetet, S.; Zidi-Yahiaoui, N.; Engel, A.; Colin, Y.; Bertrand, O.; Ripoche, P. *PLoS ONE* **2010**, *5*, e8921.
- (114) Musa-Aziz, R.; Jiang, L.; Chen, L.-M.; Behar, K.; Boron, W. J. *Membr. Biol.* **2009**, *228*, 15–31.
- (115) Hall, J. A.; Kustu, S. *Proc. Natl. Acad. Sci. U.S.A.* **2011**.

- (116) Tremblay, P.-L.; Hallenbeck, P. C. *Mol. Microbiol.* **2009**, *71*, 12–22.
- (117) McDonald, T.; Dietrich, F.; Lutzoni, F. *Mol. Biol. Evol.* **2011**, *29*, 51–60.
- (118) Till, M. S.; Essigke, T.; Becker, T.; Ullmann, G. M. *J. Phys. Chem. B* **2008**, *112*, 13401–13410.
- (119) Pflock, T. J.; Oellerich, S.; Southall, J.; Cogdell, R. J.; Ullmann, G. M.; Köhler, J. *J. Phys. Chem. B* **2011**, *115*, 8813–8820.
- (120) Pflock, T. J.; Oellerich, S.; Krapf, L.; Southall, J.; Cogdell, R. J.; Ullmann, G. M.; Köhler, J. *J. Phys. Chem. B* **2011**, *115*, 8821–8831.
- (121) Bombarda, E.; Ullmann, G. M. *Faraday Discuss.* **2011**, *148*, 173–193.
- (122) Clark, T. J. *Comput. Aided Mol. Des.* **2010**, *24*, 605–611.
- (123) Malde, A.; Mark, A. J. *Comput. Aid. Mol. Des.* **2011**, *25*, 1–12.
- (124) Donnini, S.; Villa, A.; Groenhof, G.; Mark, A. E.; Wierenga, R. K.; Juffer, A. H. *Proteins* **2009**, *76*, 138–150.
- (125) Aguilar, B.; Anandakrishnan, R.; Ruscio, J. Z.; Onufriev, A. V. *Biophys. J.* **2010**, *98*, 872–880.
- (126) Wright, E.; Serpersu, E. H. *J. Thermodyn. Catal.* **2011**, *2*, 105.
- (127) Aleksandrov, A.; Simonson, T. *J. Comput. Chem.* **2010**, *31*, 1550–1560.
- (128) Aleksandrov, A.; Proft, J.; Hinrichs, W.; Simonson, T. *ChemBioChem* **2007**, *8*, 675–685.
- (129) Park, M.-S.; Gao, C.; Stern, H. A. *Proteins* **2011**, *79*, 304–314.
- (130) Sondergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comp.* **2011**, *7*, 2284–2295.
- (131) Muddana, H.; Daniel Varnado, C.; Bielawski, C.; Urbach, A.; Isaacs, L.; Geballe, M.; Gilson, M. *J. Comput. Aid. Mol. Des.* **2012**, *26*, 475–487.
- (132) Natesan, S.; Subramaniam, R.; Bergeron, C.; Balaz, S. *J. Med. Chem.* **2012**, *55*, 2035–2047.
- (133) ten Brink, T.; Exner, T. *J. Comput. Aided Mol. Des.* **2010**, *24*, 935–942.
- (134) Czodrowski, P. *Bioorgan. Med. Chem.* **2012**, doi: 10.1016/j.bmc.2012.03.009.
- (135) Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmüller, H. *J. Chem. Theory Comp.* **2011**, *7*, 1962–1978.
- (136) Wallace, J. A.; Shen, J. K. *J. Chem. Theory Comp.* **2011**, *7*, 2617–2629.
- (137) Goh, G. B.; Knight, J. L.; Brooks, C. L. *J. Chem. Theory Comp.* **2012**, *8*, 36–46.



## **Chapter 8**

### **Manuscript E:**

Simulation of the Electron Transfer  
between the Tetraheme Subunit and the  
Special Pair of the Photosynthetic  
Reaction Center Using a Microstate  
Description

Torsten Becker, R. Thomas Ullmann and G. Matthias Ullmann  
J. Phys. Chem. B 2007, 111, 2957-2968  
doi: 10.1021/jp066264a

## Simulation of the Electron Transfer between the Tetraheme Subunit and the Special Pair of the Photosynthetic Reaction Center Using a Microstate Description

Torsten Becker, R. Thomas Ullmann, and G. Matthias Ullmann\*

Structural Biology/Bioinformatics, University of Bayreuth, Universitätsstrasse 30,  
BGI,95447 Bayreuth, Germany

Received: September 25, 2006; In Final Form: December 14, 2006

Charge transfer through biological macromolecules is essential for many biological processes such as, for instance, photosynthesis and respiration. Protons or electrons are transferred between titratable residues or redox-active cofactors, respectively. Transfer rates between these sites depend on the current charge configuration of neighboring sites. Here, we formulate the kinetics of charge-transfer systems in a microstate formalism. A unique transfer rate constant can be assigned to the interconversion of microstates. Mutual interactions between sites participating in the transfer reactions are naturally taken into account. The formalism is applied to the kinetics of electron transfer in the tetraheme subunit and the special pair of the reaction center of *Blastochloris viridis*. It is shown that continuum electrostatic calculations can be used in combination with an existing empirical rate law to obtain electron-transfer rate constants. The re-reduction kinetics of the photo-oxidized special pair simulated in a microstate formalism is shown to be in good agreement with experimental data. A flux analysis is used to follow the individual electron-transfer steps.

### Introduction

Electron transfer, often coupled to proton transfer, is one of the fundamental processes of biochemistry. In photosynthesis or oxidative phosphorylation, for instance, the transfer of electrons along a chain of redox-active sites enables biochemical systems to convert light or energy stored in chemical compounds into energy forms that can be used for other biochemical reactions. Impressive progress has been made over the last decades in understanding the underlying processes even on an atomic level. This progress was supported by the availability of structures of the involved proteins.<sup>1,2</sup> The structural details allow to interpret data obtained by titration and kinetic experiments in terms of redox-active groups and transfer events between them. This detailed analysis gives a first idea about the microscopic picture of the biological function of these proteins.<sup>3</sup> Available structures also enable computational investigation of these proteins, thus providing a theoretical insight into their molecular mechanism. Transfer processes require the knowledge of the equilibrium energetics of the involved states and the barriers connecting them. According to Marcus theory, the barrier determining a given electron-transfer rate depends on the free-energy difference between the donor state and the acceptor state, on the response of the surrounding media to changes of the charge distribution, the so-called reorganization energy, and on the electronic coupling between the donor state and the acceptor state.<sup>4,5</sup> Empirical models based on Marcus theory proved to be very successful in reproducing electron-transfer rates for various biological electron-transfer systems.<sup>6–9</sup> For systems with well-known equilibrium energetics, even simulations on complete electron-transfer systems have been performed.<sup>10,11</sup> In those simulations, data not directly accessible

to experiment, such as the reorganization energies or interactions between redox-active groups, have been estimated from kinetic experiments. Electrostatic calculations have been shown to provide good estimates of redox potentials and of the interactions among sites in a protein<sup>12,13</sup> as well as estimates of the reorganization energy.<sup>14</sup> These quantities are required to calculate electron-transfer rates. To gain theoretical insight into transfer processes, it is desirable to simulate electron-transfer systems without referring to system-specific data derived experimentally.

It is known from experiments and theoretical calculations that the mutual interactions between sites of the system are crucial for an understanding of equilibrium and kinetic properties.<sup>15–17</sup> In theoretical calculations of protonation and redox properties of proteins, these mutual interactions are reflected by describing the system in terms of microstates, i.e., a state of the system where every protonatable or redox-active site has a well-defined charge form.<sup>13</sup> Titration properties for single sites are given by thermal averaging of the contribution of these microstates. Kinetic simulations, in contrast, commonly refer to transition events between individual sites.<sup>11,18</sup> Although intuitive, such a description has a major drawback since, in general, it is not possible to assign a unique rate constant to a charge-transfer reaction between two given sites. The charge configuration of the surrounding sites may significantly influence the charge-transfer rate constant. Thus, in principle, there are as many charge-transfer rate constants between two given sites as there are charge configurations of neighboring sites. In the present work, we formulate kinetic equations of a transfer system in terms of the microstates already known from titration calculations. This approach resolves the ambiguities for calculating rate constants between sites. The rate constants are calculated using an empirical rate law.<sup>8</sup> All energetic parameters required for rate calculations are obtained from electrostatic calculations.

\* Author to whom correspondence should be addressed. Fax: +49-921-55-3544. E-mail: Matthias.Ullmann@uni-bayreuth.de.

To test the accuracy of this theoretical approach, we simulate the electron-transfer kinetics of the multiheme subunit of bacterial photosynthetic reaction center (RC) and the so-called special pair (SP): a chlorophyll dimer.

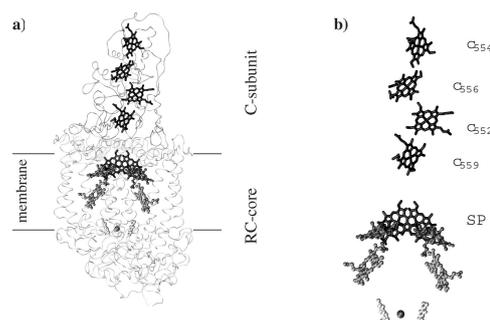
The RC is an integral membrane protein, which couples the oxidation of a soluble cytochrome *c* or an iron sulfur protein on the periplasmic side of the membrane to the reduction of a quinone at the cytoplasmic side.<sup>3,19</sup> Structures are available for various systems,<sup>20–22</sup> and the equilibrium energetics as well as transfer kinetics have been accessed in numerous studies.<sup>23–26</sup> The core of all reaction centers is formed by three subunits labeled H, M, and L.<sup>3</sup> Following the photoinduced excitation of the SP an electron is transferred via several redox cofactors to the quinone  $Q_B$ . Re-reduction of the SP, i.e., the reduction after the photo-oxidation, is facilitated either directly by a soluble protein transport protein (e.g., *Rhodobacter sphaeroides*) or via an additional C-subunit (e.g., *Blastochloris viridis*). The C-subunit, if present, contains four heme cofactors forming a transfer chain along the membrane normal. Electrons enter the C-subunit via a diffusing electron transport protein, which probably binds close to the outermost heme group.<sup>27</sup>

Photoinduced oxidation of the SP offers a well-defined initialization that can be controlled experimentally. Light absorption at various frequencies allows to follow the changes in the redox form of various cofactors. Electron transfer in the C-subunit is, thus, a process well-suited for kinetic analysis. Consequently, this subsystem was probed not only by equilibrium titration<sup>23,24</sup> but also by kinetic experiments.<sup>25,28</sup> With decreasing distance from the SP, the midpoint potentials of the hemes show a peculiar low–high–low–high pattern. Initial preparation of the system in various redox forms revealed a strong dependence of the kinetics of the re-reduction of the SP on the initial state of the system. As shown in ref 25, the reduction of the SP is the slowest if only the highest potential heme is reduced (heme  $c_{559}$ ). The re-reduction of the SP becomes faster upon reducing additional hemes and is 2 times faster if the three highest potential hemes are reduced. Since there is strong evidence that the re-reduction of the SP always occurs from heme  $c_{559}$ ,<sup>28</sup> this reaction is one example where charge-transfer rates between given sites depend significantly on the charges of their neighboring sites. An analysis using experimental redox potentials in combination with calculated interaction energies strongly suggests that the electrostatic interaction between heme  $c_{559}$  and heme  $c_{552}$  is responsible for the 2-fold increase in the rate.<sup>28</sup>

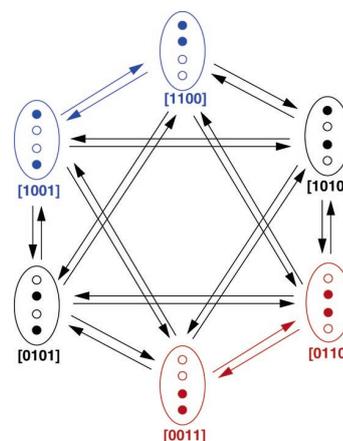
In this article, we present a coherent scheme to simulate complex charge-transfer reactions in biological systems using coupled differential equations. Charge transfer is described as a transition from one microstate of the transfer system to another. Mutual interactions between sites are naturally taken into account within the presented formalism. Unique transfer rate constants can be assigned to transitions between microstates. Electron-transfer rates are estimated in the framework of Marcus theory. Equilibrium electrostatic calculations are used to access the energetics and mutual interactions between sites. The formalism is applied to electron transfer between the four heme cofactors and the SP of the RC of *B. viridis*. The simulated re-reduction kinetics of the photo-oxidized SP are in good agreement with experimental data.

## Theory

It is common to describe charge transfer in proteins as a series of transfer events between groups involved in the transfer reaction.<sup>11,18</sup> For example in electron transfer between the



**Figure 1.** (a) Reaction center (RC) with the C-subunit of *Blastochloris viridis*. (b) Redox cofactors of the RC. The four hemes of the C-subunit form a near-linear transfer chain along the membrane normal. Electrons are transferred along this chain to reduce the special pair, SP.



**Figure 2.** Charge-transfer system described by microstates,  $\vec{x} = (x_1, \dots, x_4)$ , where 1 or 0 denotes a reduced or oxidized site, respectively. The state transfer reactions colored blue and red represent charge transfer from site 4 to site 2, differing in the charge configuration of surrounding sites.

C-subunit and the SP of the RC, such a description corresponds to the attempt to estimate rate constants for the transfer between the heme cofactors and the SP directly. This approach, however, bears some problems as can be seen for the system depicted in Figure 2. This system has four redox-active sites and several possible transitions between them. The state of the system is described by a four-dimensional vector defining a microstate. The elements of this vector are either 1 or 0, indicating whether the sites are reduced or oxidized, respectively. Each transition represents the transfer of an electron from one redox-active site to another. Examining, for example, the two transfer reactions  $[1001] \rightarrow [1100]$  and  $[0011] \rightarrow [0110]$  (blue and red in Figure 2, respectively), one can see that both events represent an electron transfer from site four to site two. Nevertheless, these transfer events are obviously not the same since they differ in the charge configuration of the other sites. Thus for the transfer from site four to site two, there are not one but several rate constants, depending on the number of possible configurations of the remaining sites. Figure 2, however, already suggests an alternative description of transfer reactions which will be outlined below.

**System Description.** Although, in general, no unique rate constant can be assigned to a charge transfer between two given sites of a multisite system, there is a unique transition between two given microstates. Thus, the ambiguities inherent for rates between individual sites can be resolved if transitions between

microstates of the system are considered. A description based on microstates is already well established in theoretical calculations of equilibrium redox and protonation probabilities of proteins<sup>12,13,29,30</sup> and was proposed for charge-transfer systems by Bashford.<sup>31,32</sup> Here, we use the concept of microstates to describe the kinetics of a charge-transfer system. To keep the notation simple, we formulate the theory for a system of redox-active groups only, i.e., an electron-transfer system. Treating proton transfer or coupled systems is, however, straightforward.

We consider a system with  $N$  redox-active sites. Each state of the system can be written as an  $N$ -dimensional vector  $\vec{x} = (x_1, \dots, x_N)$ , where  $x_i$  is 0 or 1 if site  $i$  is oxidized or reduced, respectively. A transfer event is described as a transition from one state vector to another. In the most general description, transitions are possible between any two given states. In practice, external constraints on the system may restrict the number of possible transitions. Such restrictions might concern the number of electrons in the system; i.e., if no electrons are allowed to enter or leave the system, then only transitions between state vectors with equal numbers of reduced sites will be allowed. If concerted transfer reactions cannot occur, then only transitions that represent a single electron transfer are considered. Such transfer events have the form  $(x_1, \dots, x_i, \dots, x_j, \dots, x_n) \rightarrow (x_1, \dots, \bar{x}_i, \dots, \bar{x}_j, \dots, x_n)$  where  $x_i$  and  $\bar{x}_i$  denote the redox state of site  $i$  before and after the transfer event, respectively.

Each state of the system has a well-defined energy that depends on the energetics of the individual sites and the interaction between sites. The energy of a state  $\vec{x}$  is given by<sup>33</sup>

$$G(\vec{x}) = \sum_{i=1}^N (x_i - x_i^\circ) F(E_i^{\text{intr}} - E) + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (x_i - x_i^\circ)(x_j - x_j^\circ) W_{ij} \quad (1)$$

where  $F$  is the Faraday constant,  $x_i$  denotes the redox state of the site  $i$  in state  $\vec{x}$ ,  $x_i^\circ$  is the reference form of site  $i$ ,  $E_i^{\text{intr}}$  is the redox potential that site  $i$  would have if all other sites are in their reference form (intrinsic redox potential),  $E$  is the reduction potential of the solution, and  $W_{ij}$  represents the interaction of site  $i$  with site  $j$ .

**Equilibrium Properties.** Equilibrium properties of a physical system are completely determined by the energies of its states. To keep the notation concise, states will be numbered by Greek indices; i.e., for state energies,  $G_\nu \equiv G(\vec{x})$ . For site indices, the roman letters  $i$  and  $j$  will be used.

The equilibrium probability of a single state is given by

$$P_\nu^{\text{eq}} = \frac{e^{-\beta G_\nu}}{Z} \quad (2)$$

with  $\beta = 1/kT$  and  $Z$  being the partition function of the system

$$Z = \sum_{\nu=1}^M e^{-\beta G_\nu} \quad (3)$$

The sum runs over all  $M$  possible states. Properties of single sites can be obtained from eq 2 by summing up the individual contributions of all states. For example, the probability of site  $i$  being reduced is given by

$$\langle x_i \rangle = \sum_{\nu} x_{\nu,i} P_\nu^{\text{eq}} \quad (4)$$

where  $x_{\nu,i}$  denotes the redox form of site  $i$  in the charge state  $\nu$ . For small systems, this sum can be evaluated explicitly. For larger systems, Monte Carlo techniques can be invoked to determine these probabilities.

For a system of interacting sites, the probabilities  $\langle x_i \rangle$  can show a complex shape, thus rendering the assignment of midpoint potentials difficult or even meaningless.<sup>16</sup> The energy differences between microstates, however, remain well-defined and thus form a convenient basis to describe the system.

**Time Evolution.** In the microstate description put forward in this article, charge-transfer events are described as transitions between well-defined microstates of a system. We simulate the time dependence of the population of each microstate using a master equation

$$\frac{d}{dt} P_\nu(t) = \sum_{\mu=1}^M k_{\nu\mu} P_\mu(t) - \sum_{\mu=1}^M k_{\mu\nu} P_\nu(t) \quad (5)$$

where  $P_\nu(t)$  denotes the probability that the system is in charge state  $\nu$  at time  $t$  and  $k_{\nu\mu}$  denotes the probability per unit time that the system will change its state from  $\mu$  to  $\nu$ . The summation runs over all possible states  $\mu$ . Simulating charge transfer by eq 5 assumes that these processes can be described as a (stochastic) Markov process. This assumption implies that the probability of a given charge transfer only depends on the current state of the system and not on the way the system reached this state. Our approach excludes systems with strongly coupled microstates, e.g., delocalized charge systems, where the charge distribution can only be adequately described by a coherent superposition of various microstates. Delocalization of charges one mainly expects for strongly coupled electron-transfer systems due to the small mass of the electron. In biological systems, however, one often encounters long-range electron transfer, which is accurately described in the framework of Marcus theory. Thus, biological electron-transfer systems are well described in the nonadiabatic picture, i.e., in the weak-coupling limit.

Equation 5 is formally equivalent to kinetic rate equations known from physical chemistry. However, it should be noted that eq 5 describes the time evolution of the probability distribution of microstates of the system. For these microstates, as argued above, energies  $G_\nu$  and transition probabilities  $k_{\nu\mu}$  can be assigned unambiguously. The correct time-dependent probability of finding a single site in the reduced form can be obtained by summing up individual contributions from the time-dependent probabilities  $P_\nu(t)$

$$\langle x_i \rangle(t) = \sum_{\nu} x_{\nu,i} P_\nu(t) \quad (6)$$

Equation 5 is a coupled system of linear differential equations with constant coefficients. As explained in more detail in the Methods section, there is an analytical solution for nondegenerate systems that can be written as

$$P_\nu(t) = \sum_{\mu} c_{\mu} v_{\mu,\nu} e^{\alpha_{\mu} t} \quad (7)$$

where  $\alpha_{\mu}$  is the  $\mu$ th eigenvalue of the matrix associated with eq 5,  $v_{\mu,\nu}$  is the  $\nu$ th element of the  $\mu$ th eigenvector of this matrix, and  $c_{\mu}$ 's are integration constants determined from the initial condition of the system.

**Flux Analysis.** For analyzing a complex charge-transfer system, it is of particular interest to follow the flow of charges

through the system, i.e., the charge flux. The flux from state  $\nu$  to state  $\mu$  is determined by the population of state  $\nu$  times the probability per unit time that state  $\nu$  will change into state  $\mu$ , i.e., by  $k_{\nu\mu}P_\nu(t)$ . The net flux between states  $\nu$  and  $\mu$  is thus given by

$$J_{\nu\mu}(t) = k_{\nu\mu}P_\mu(t) - k_{\mu\nu}P_\nu(t) \quad (8)$$

In the following, this flux between two states will be termed interstate flux. The interstate flux (eq 8) is positive if there is a net flux from state  $\mu$  to state  $\nu$ .

In the common case, where the transition between  $\nu$  and  $\mu$  represents the transfer of a single charge between site  $i$  and site  $j$ , this interstate flux corresponds to a flux between site  $i$  and site  $j$ ,  $J_{ij}^{(\nu\mu)} \equiv J_{\nu\mu}$ . The index  $(\nu\mu)$  here indicates that this contribution to the total flux between sites  $i$  and  $j$  is related to the interstate flux  $J_{\nu\mu}$ . In general, there will be several interstate fluxes contributing to the charge transfer between sites  $i$  and  $j$ . The total net flux between these sites  $J_{ij}$  due to the interstate fluxes  $J_{\nu\mu}$  is therefore given by the sum over interstate fluxes

$$J_{ij} = \sum_{(\nu\mu)} J_{ij}^{(\nu\mu)} = \sum_{(\nu\mu)} J_{\nu\mu} \quad (9)$$

Here the summation index  $(\nu\mu)$  indicates that the sum has to be taken over all contributing interstate fluxes. The fluxes  $J_{ij}$  will be referred to as intersite fluxes. Equipped with these elementary fluxes between pairs of sites, the complete flux network of the system can be deduced.

**Calculation of the Rate Constants  $k_{\nu\mu}$ .** The outlined theory is directly applicable to a large class of reaction systems such as, for example, proton and electron transfer in proteins. The determination of the rate constants  $k_{\nu\mu}$  will, however, be specific for the particular reactions that should be simulated. For electron-transfer systems, continuum electrostatic calculations in combination with existing empirical rate laws<sup>6,8</sup> can be used to obtain electron-transfer rates in good agreement with experimental data. Three factors mainly govern the rate constants of biological electron-transfer reactions: the energy difference between the donor state and the acceptor state, the environmental polarization (reorganization energy), and the electronic coupling between the redox sites. The energy barrier for the transfer process is given in the framework of Marcus theory as

$$\Delta G^\ddagger = \frac{(\Delta G^\circ + \lambda)^2}{4\lambda} \quad (10)$$

where  $\Delta G^\circ$  is the energy difference between the donor state and the acceptor state and  $\lambda$  is the so-called reorganization energy. The electronic coupling between the redox sites is commonly accounted for by a distance-dependent exponential function  $A \exp(-\beta(R - R_0))$  where  $R$  is the edge-to-edge distance between cofactors,  $R_0$  represents a Van der Waals contact distance, and  $A$  represents an optimal rate. These aspects of biological electron transfer have been successfully combined to formulate a heuristic rate law applicable to long-range electron transfer<sup>6,8</sup>

$$\begin{aligned} \log(k_{\text{ex}}) &= 13 - 0.6(R - 3.6) - 3.1 \frac{(\Delta G^\circ + \lambda)^2}{\lambda} \\ \log(k_{\text{en}}) &= 13 - 0.6(R - 3.6) - 3.1 \frac{(-\Delta G^\circ + \lambda)^2}{\lambda} - \frac{\Delta G^\circ}{0.06} \end{aligned} \quad (11)$$

where  $k_{\text{ex}}$  and  $k_{\text{en}}$  are the rate constants for exothermic and endothermic electron-transfer reactions, respectively.

The free energy  $\Delta G^\circ$  for a transition between two states  $\nu$  and  $\mu$  can be calculated within the electrostatic model using eq 1. The reorganization energy  $\lambda$  contains two contributions,  $\lambda = \lambda_o + \lambda_i$ , where  $\lambda_o$  is the solvent reorganization energy and  $\lambda_i$  accounts for changes of the nuclear degrees of freedom between the donor site and the acceptor site.  $\lambda_o$  was shown to be accessible to equilibrium calculations<sup>34</sup> and formulated especially suited for Poisson–Boltzmann calculations<sup>14</sup>

$$\lambda_o = -\frac{1}{2} \sum_i^K (\phi_{\text{ad}}^{\text{opt}}(\vec{r}_i) - \phi_{\text{ad}}(\vec{r}_i)) \Delta q_i^{\text{ad}} \quad (12)$$

$\Delta q_i^{\text{ad}}$  is the change in charge of atom  $i$  when going from the donor to the acceptor state. The potentials  $\phi_{\text{ad}}^{\text{opt}}$  and  $\phi_{\text{ad}}$  are generated by the charge distribution  $\Delta\rho = \rho_a - \rho_d$  in a low (opt) and a high dielectric environment, respectively. Here,  $\rho_a$  and  $\rho_d$  denote the charge distributions of the acceptor and donor states, respectively. The permittivity constant for the low dielectric environment reflects the electronic polarizability while the permittivity constant for the high dielectric environment accounts for the nuclear and electronic polarizabilities. The solvent reorganization energy is given by the difference in solvation free energies of the charge distribution  $\Delta\rho$  between a low and a high dielectric environment.  $\lambda_i$ 's can be estimated by density functional theory (DFT) calculations but are often found to be significantly smaller than the solvent reorganization energy.<sup>35–38</sup> The inner sphere reorganization energy can be calculated from density functional theory as

$$\lambda_i = E_{\text{bond}}(\vec{r}_d, \rho_a) - E_{\text{bond}}(\vec{r}_a, \rho_a) \quad (13)$$

where  $E_{\text{bond}}$  is the total DFT energy of the two sites and  $\vec{r}_d$  and  $\vec{r}_a$  correspond to the optimized geometries of the sites in the donor and acceptor states, respectively. Hence, the inner sphere reorganization energy is given by the difference in bonding energy between the donor and the acceptor geometries while the sites are kept in their acceptor state charge distribution.<sup>37,38</sup>

In this formulation, the total reorganization energy depends only on the charge difference between two states. Thus, the electron transfer between sites  $i$  and  $j$  is always connected with the same reorganization energy irrespective of the states that are converted into each other.

## Methods

**Structures and Parameters.** In the calculations, we used the structure of the RC of *Blastochloris viridis* (PDB code 1PRC)<sup>39</sup> having a resolution of 2.3 Å. Hydrogens were added with HBUILD<sup>40</sup> in CHARMM<sup>41</sup> and subsequently minimized using the CHARMM force field.<sup>42</sup>

The atomic partial charges for most atoms were taken from the CHARMM force field. The partial charges of the hemes, the special pair, and the linked residues were obtained from density functional calculations (functionals VWN<sup>43</sup> and PW91<sup>44</sup>) using the ADF program package.<sup>45</sup> The electrostatic potentials obtained from the density functional calculation were fitted using the CHELPG algorithm<sup>46</sup> combined with a singular value decomposition.<sup>47</sup> The charges of all other cofactors are the same as those in previous calculations.<sup>48–50</sup>

**Continuum Electrostatic Calculations and Equilibrium Redox Titration.** All of the electrostatic calculations were done

using the program suite MEAD,<sup>12,51</sup> which uses a finite difference method to determine the electrostatic potential. For the calculations of the intrinsic redox potentials and the interaction energies, the dielectric constant of the protein and the dielectric constant of water were set to 4.0 and 80.0, respectively. The ionic strength was set to 0.1 M. For the protein, the electrostatic potential was calculated by focusing using three grids of 141<sup>3</sup>, 121<sup>3</sup>, and 121<sup>3</sup> grid points and grid spacings of 2.0, 1.0, and 0.25 Å, respectively. For the model compounds, the electrostatic potential was calculated by focusing using two grids of 121<sup>3</sup> grid points and grid spacings of 1.0 and 0.25 Å. The first grid was centered on the protein or the model compound; the other grids were centered on the titratable group.

To calculate the energy of the different redox states, we first determined the protonation of the different protonatable residues in the protein using continuum electrostatics and Monte Carlo simulations.<sup>52</sup> We then fixed the highest populated protonation state to calculate the energy of the different redox states. The solution redox potentials of the special pair, of the bis-histidinyl heme, and of the metioninyl-histidinyl heme were adjusted to fit the equilibrium redox behavior as closely as possible. The equilibrium titration curves and the populations of the different redox states in dependence on the solution redox potential have been calculated by a statistical mechanics averaging.

**Reorganization Energies.** For the calculation of the solvent reorganization energy, we used a dielectric constant of 1.0 for the cofactor, 2.0 for the electronic dielectric constant, 4.0 for the total dielectric constant of the protein, and 80.0 for the water. In the calculation of the reaction field potential in the high dielectric environment, an ionic strength of 0.1 M was considered. The electrostatic potential was calculated by focusing using three grids of 181<sup>3</sup>, 181<sup>3</sup>, and 301<sup>3</sup> grid points and grid spacings of 2.0, 1.0, and 0.25 Å, respectively. The first grid was centered on the geometric center of the protein; the other two grids were centered on the geometric centers of the cofactors between which the electron transfer takes place.

Inner sphere reorganization energies were computed from density functional calculations (functionals VWN<sup>43</sup> and Becke Perdew<sup>53</sup> with a TZ2P basis set) using the ADF program.<sup>45</sup> The interaction between the sites influences the inner sphere reorganization energy  $\lambda_i$  only slightly. Therefore,  $\lambda_i$  was calculated as the sum of the contributions of the two cofactors involved in the transfer reaction.<sup>38,54</sup> The model sites consisted of the metal ions, the porphyrins, and the side chains axially coordinating the metal ions or binding the heme porphyrins. The amino acid side chains were cut at the C $_{\alpha}$  atom, and their C $_{\beta}$  atoms were fixed in their crystal structure positions. The heme propionates were cut off and substituted by hydrogen atoms. The phytol tails of the special pair were truncated to methyl groups. Geometry optimizations with stringent convergence criteria were performed for each model site in its reduced and oxidized forms. Starting from the geometry-optimized structures, single-point calculations were performed to obtain the bonding energies of the model systems having the optimized geometry of one redox form and the respective opposite charge distribution.

**Solution of the Differential Equation.** The master equation in eq 5 can be rewritten as

$$\begin{pmatrix} \frac{dP_1(t)}{dt} \\ \vdots \\ \frac{dP_v(t)}{dt} \\ \vdots \\ \frac{dP_M(t)}{dt} \end{pmatrix} = \begin{pmatrix} -\sum_{\mu=1}^M k_{\mu 1} & \cdots & k_{1\mu} & \cdots & k_{1M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ k_{v1} & \cdots & -\sum_{\mu=1}^M k_{\mu v} & \cdots & k_{vM} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ k_{M1} & \cdots & k_{Mv} & \cdots & -\sum_{\mu=1}^M k_{\mu M} \end{pmatrix} \begin{pmatrix} P_1(t) \\ \vdots \\ P_v(t) \\ \vdots \\ P_M(t) \end{pmatrix} \quad (14)$$

where  $M$  is the number of states. For simplicity, we call the probability vector  $\mathbf{p}$ , the matrix of rate constants  $\mathbf{A}$ , and its elements  $a_{\nu\mu}$ ; i.e., eq 14 becomes

$$\frac{d\mathbf{p}(t)}{dt} = \mathbf{A}\mathbf{p}(t) \quad (15)$$

The diagonal elements  $a_{\nu\nu}$  of matrix  $\mathbf{A}$  are the negative of the sums over all of the rate constants  $k_{\mu\nu}$  diminishing the population of state  $\nu$ . The diagonal elements  $a_{\nu\nu}$ , therefore, represent a decay rate for the population of state vector  $\nu$ . The off-diagonal element  $a_{\nu\mu}$  is the rate constant  $k_{\nu\mu}$  for the conversion of state  $\mu$  to state  $\nu$ .

Equation 14 is a homogeneous system of first-order linear ordinary differential equations and can be solved formally as

$$\mathbf{p}(t) = \exp(\mathbf{A}t)\mathbf{p}(0) \quad (16)$$

A detailed balance criterium ensures that matrix  $\mathbf{A}$  can be diagonalized, and if all eigenvalues are nondegenerate, then the solution can be written in the especially simple form

$$\mathbf{p}(t) = \sum_{\mu=1}^M c_{\mu} \mathbf{v}_{\mu} e^{\alpha_{\mu} t} \quad (17)$$

where  $\alpha_{\mu}$  and  $\mathbf{v}_{\mu}$  are the eigenvalue and the corresponding eigenvector of matrix  $\mathbf{A}$ , respectively, and  $c_{\mu}$ 's are constants that can be determined from the initial concentrations  $\mathbf{p}$  at  $t = 0$  (i.e., all the terms  $e^{\alpha_{\mu} t} = 1$ )

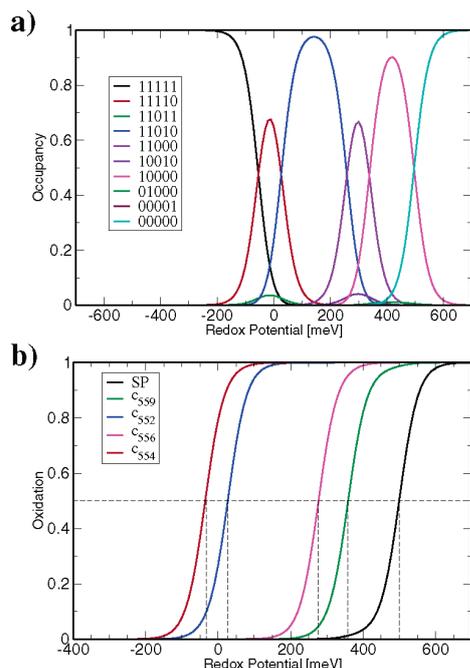
$$\mathbf{p}(0) = \mathbf{V}\mathbf{c} \quad (18)$$

where  $\mathbf{V}$  is a matrix containing the eigenvectors of  $\mathbf{A}$ . For the general case of degenerate eigenvalues, eq 17 has to be slightly modified.<sup>55</sup> For a more detailed discussion of master equations and their general behaviors, see, for example, refs 55 and 56.

To solve the system of differential equations in eq 14 in the form of eq 17, the eigenvalues and the eigenvectors of matrix  $\mathbf{A}$  were determined using the routine DGEEV of the LAPACK library.<sup>57</sup> The coefficients  $c_{\mu}$  are determined by solving eq 18 using standard numerical methods (lower and upper triangular decomposition).<sup>58</sup>

## Results and Discussion

In this article, we use a microstate formalism to simulate the kinetics of electron transfer between the C-subunit and the SP



**Figure 3.** (a) Occupancies of significantly populated microstates, where 1 denotes reduced whereas 0 indicates the oxidized state. The state vector is given in the order (SP, C<sub>559</sub>, C<sub>552</sub>, C<sub>556</sub>, C<sub>554</sub>). (b) Titration curves for individual sites were calculated from eq 4. The dashed lines indicate the midpoint potentials as measured in titration experiments.

**TABLE 1: Calculated Equilibrium Midpoint Potentials Compared to Experimental Values**

$E_m$ (meV)	SP	C <sub>559</sub>	C <sub>552</sub>	C <sub>556</sub>	C <sub>554</sub>
calculated	500	359	27	278	-35
Fritzsch et al. <sup>24</sup>	500	370	10	300	-60
Dracheva et al. <sup>23</sup>	500	380	20	310	-60

of the RC. Poisson–Boltzmann electrostatics in combination with an empirical rate law provide transition probabilities between microstates. These rate constants are used to simulate the kinetic behavior of this system using coupled differential equations. The resulting re-reduction kinetics of the special pair SP reproduces the redox state dependence seen in experiments and are shown to be in quantitative agreement with the kinetic data obtained in ref. 25.

**Midpoint Potentials.** A prerequisite for a good description of the kinetics is to accurately account for equilibrium properties. Therefore, we compute first the equilibrium midpoint potentials for all four hemes and the special pair SP using Poisson–Boltzmann electrostatics.

Figure 3 depicts the occupancies of those microstates that were found to be significantly populated. Experimentally observed Nernst-like titration curves are obtained for individual sites by summing up the contributions of these microstates as stated in eq 4. Table 1 lists the calculated midpoint potentials, which are compared to the experimental values. A good overall agreement is achieved; the characteristic high–low–high–low pattern with respect to the distance from the special pair is clearly reproduced, and the numerical values of all hemes are within the range of accuracy commonly achieved by electrostatic theory.<sup>59</sup> The same intrinsic redox potentials and interaction energies that allow us to reproduce the redox-titration behavior of the RC are used to calculate microscopic redox potentials (Table 2). These microscopic redox potentials are required to compute rate constants.

**TABLE 2: Intrinsic Redox Potentials and Interaction Energies for the SP and the Four Heme Groups of the C-Subunit of *B. viridis*<sup>a</sup>**

	$E^{\text{intr}}$	SP	C <sub>559</sub>	C <sub>552</sub>	C <sub>556</sub>	C <sub>554</sub>
SP	451	0				
C <sub>559</sub>	286	39	0			
C <sub>552</sub>	23	6	67	0		
C <sub>556</sub>	170	1	7	51	0	
C <sub>554</sub>	-36	0.1	1	5	58	0

<sup>a</sup> All values are given in meV.

**TABLE 3: Solvent and Inner Sphere Reorganization Energies and Cofactor Distances**

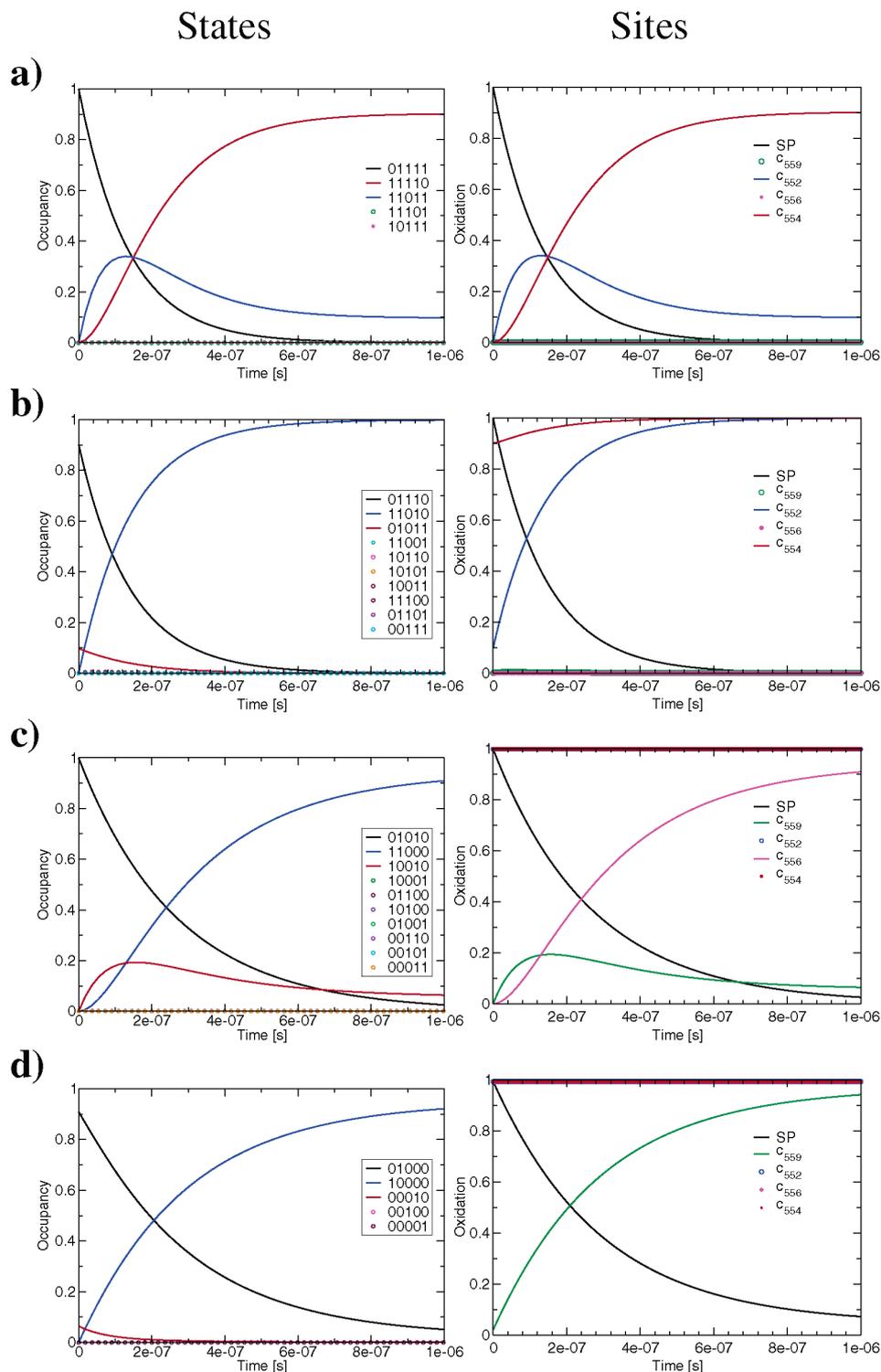
site 1	site 2	$\lambda_o$ (eV)	$\lambda_i$ (eV)	$R$ (Å)
SP	C <sub>559</sub>	0.51	0.115	12.1
SP	C <sub>552</sub>	0.58	0.100	24.6
SP	C <sub>556</sub>	0.59	0.115	37.4
SP	C <sub>554</sub>	0.59	0.115	50.9
C <sub>559</sub>	C <sub>552</sub>	0.44	0.063	6.9
C <sub>559</sub>	C <sub>556</sub>	0.57	0.076	21.5
C <sub>559</sub>	C <sub>554</sub>	0.59	0.076	35.4
C <sub>552</sub>	C <sub>556</sub>	0.46	0.063	8.1
C <sub>552</sub>	C <sub>554</sub>	0.57	0.063	21.5
C <sub>556</sub>	C <sub>554</sub>	0.41	0.076	7.1

**Reorganization Energies.** The next important parameter determining electron-transfer rate constants is the reorganization energy. Table 3 lists the calculated reorganization energies and the edge-to-edge distances between cofactors. The total reorganization energy  $\lambda = \lambda_o + \lambda_i$  varies between 0.48 – 0.66 eV. Experimental values for the reorganization energies for the electron transfer from the C-subunit to the SP and within the C-subunit do not exist. Reorganization energies obtained on other biological systems have been reported in the range between 0.2 and 1.2 eV.<sup>14,60–62</sup> Thus, although the reorganization energies obtained for our system cannot directly be compared to experimental data, the values obtained from our calculations are in agreement with literature values reported for biological transfer systems.

**Kinetics.** The midpoint potentials in Table 1 cannot directly be used to obtain the reaction free energy of a transfer event as necessary for eq 11. Midpoint potentials reflect the energetic cost to reduce a particular site while all sites are in equilibrium with the solution reduction potential. In contrast, in eq 11 the free energy for transferring an electron refers to the free-energy difference between two states differing in their charge configuration (microstate) by exactly the transferred electron. The latter energy can be directly related to a microscopic equilibrium constant, while the relation between a midpoint potential and a real equilibrium constant is not straightforward.<sup>16</sup>

To demonstrate that Poisson–Boltzmann electrostatics can be used in a microstate formalism to obtain valid descriptions of the kinetic properties of electron-transfer systems, we compare the re-reduction kinetics of the SP to the experimentally observed half-life of the oxidized state.<sup>25</sup> Ortega et al. exposed the reaction center of *B. viridis* to different redox potentials, thus preparing the system in charge configurations with 4, 3, and 2 electrons distributed over the system consisting of the four hemes and the SP. The re-reduction kinetics of the SP were measured after photoinduced oxidation.

To mimic this experimental setup, four simulations were performed. First, all four hemes were set to their reduced forms, the SP was set to its oxidized form, and the kinetics of this system were simulated according to eq 7. The kinetics of this first re-reduction reaction were not measured experimentally. The next three simulations started from the equilibrium distribu-



**Figure 4.** Left column depicting the time-dependent probability distribution of microstates after photo-oxidation of the SP simulated by eq 5. The state vector is given in the order (SP, c<sub>559</sub>, c<sub>552</sub>, c<sub>556</sub>, c<sub>554</sub>). The associated oxidation probabilities of the four hemes and the SP are depicted in the right column. Data shown in symbols do not significantly differ from either 0 or 1. (a–d) The initial setup consists of 4, 3, 2, or 1 electrons distributed among the four hemes, respectively. Initial distributions for the microstates were taken from an equilibrium distribution prior to photo-oxidation of the SP.

tion of the previous run and setting the SP to its oxidized state, thus resembling a system with 4, 3, or 2 electrons present prior to instantaneous photo-oxidation of the SP.

An overview of the complete kinetics of the system is shown in Figure 4. Each state is characterized by a state vector where 1 denotes a reduced site and 0 an oxidized site. The order of

sites is given from left to right: SP, heme c<sub>559</sub>, heme c<sub>552</sub>, heme c<sub>556</sub>, and heme c<sub>554</sub>. This order corresponds to the spatial arrangement of the redox sites along the membrane normal. Figures 4a–d correspond to simulations with 4, 3, 2, and 1 electrons present after photo-oxidation of the SP, respectively. On the left side of Figure 4, the time-dependent probability

**TABLE 4: Half-Life of the Photo-oxidated State of the SP Dependent on the Number of Electrons in the System**

$t_{1/2}$ (s)	4	3	2	1
experiment <sup>25</sup>		$115 \times 10^{-9}$	$190 \times 10^{-9}$	$230 \times 10^{-9}$
simulation	$93 \times 10^{-9}$	$99 \times 10^{-9}$	$187 \times 10^{-9}$	$210 \times 10^{-9}$

distribution of the accessible microstates is shown. The corresponding oxidation probabilities for the heme active sites and the SP are shown on the right side. In all simulations, it is observed that only a limited number of microstates contributes significantly to the probability distribution in the pico- to microsecond time scale.

To obtain half-lives for the oxidized state of the SP that can be compared to the experimental data, we performed an exponential fit on the re-reduction curves of the SP. For all our simulations a one-exponential fit was sufficient to obtain good agreement with the simulation data. The seemingly nonexponential character of the SP re-reduction in Figures 4c and 4d originates from the fact that the SP is not fully reduced in equilibrium; i.e., the re-reduction curve reaches a constant level above zero. This behavior was taken into account by fitting the re-reduction kinetics of the SP to the function

$$f(t) = \frac{1}{1+a}(a + e^{-\alpha t}) \quad (19)$$

where  $\alpha$  is the decay constant of the photo-oxidized form of the SP and  $a/(1+a)$  represents its equilibrium probability. Table 4 lists the half-lives derived from this one-exponential fit of our simulation data compared to the half-lives of the very fast component obtained by Ortega et al. All values are within a factor of 1.5 compared to the experimental values. Furthermore, the increase of the lifetime upon decreasing the number of electrons in the system is clearly reproduced. This increase indicates that interactions among sites and therefore the interplay of the various possible charge states within the transfer system are reliably reproduced in the microstate formalism presented here. Thus, theoretical investigations on the basis of the theory outlined in this article promise to shed light on the function of more complicated reactions such as coupled transfer of protons and electrons.

Given the good agreement with experimental data, we further analyze the simulation by asking specific questions such as which microstates participate in the transfer process and which interactions are responsible for the observed changes in the kinetic behavior. In the present example, the simplicity of the system allows for straightforward answers of both questions within the present model. As was already suggested on the basis of electrostatic calculations in combination with experimental midpoint potentials, the change in redox kinetics upon reducing additional heme groups can be understood in terms of electrostatic interactions between the heme groups.<sup>28</sup> Reduction of heme c<sub>552</sub> and heme c<sub>556</sub> destabilizes the reduced form of heme c<sub>559</sub> and thus increases the transfer rate constants between heme c<sub>559</sub> and the SP. Table 5 lists microscopic rates related to direct electron transfer from heme c<sub>559</sub> to the SP. These rate constants only differ in the interactions between heme groups being in different redox forms. Since the transfer from heme c<sub>559</sub> to the SP is rate-determining for the reduction kinetics of the SP, they clearly indicate that the purely electrostatic interactions in the present model account for the experimentally observed dependence of lifetimes on the number of electrons in the system.

**Flux Analysis.** A picture of the individual transfer steps that contribute to the overall kinetic behavior can be obtained with a flux analysis as described in the Theory section. Figure 5

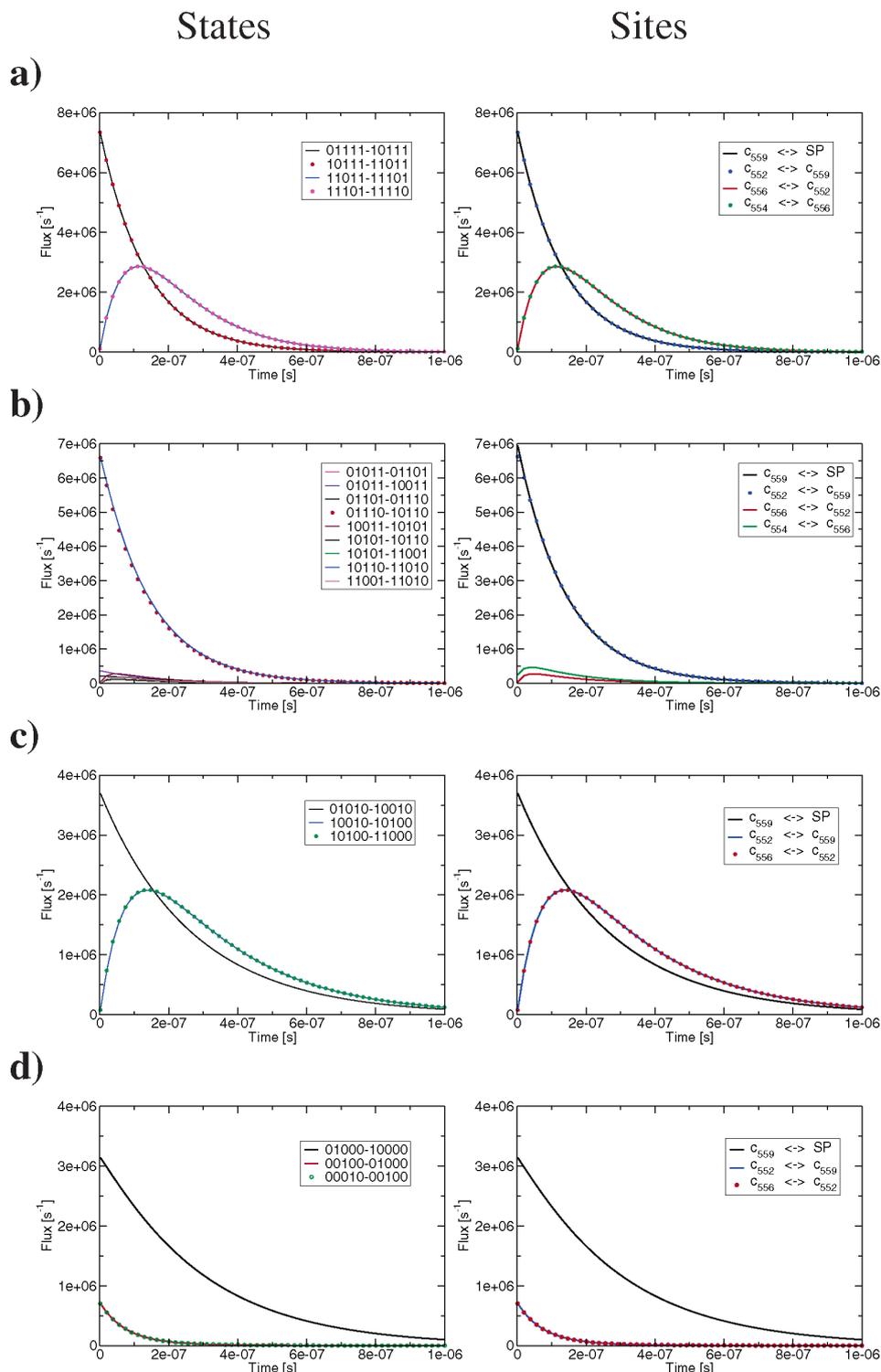
**TABLE 5: Transfer Rate Constants and Associated Microscopic Half-Lives for State Transitions Representing Electron Transfer from Heme c<sub>559</sub> to the SP**

states		$k_{ij}$ (s <sup>-1</sup> )	$t_{1/2}^{\text{micro}}$ (s)	$t_{1/2}$ (s)
donor	acceptor			
01111	10111	$7.4 \times 10^6$	$93 \times 10^{-9}$	$93 \times 10^{-9}$
01110	10110	$7.4 \times 10^6$	$93 \times 10^{-9}$	$99 \times 10^{-9}$
01010	10010	$3.7 \times 10^6$	$187 \times 10^{-9}$	$187 \times 10^{-9}$
01000	10000	$3.5 \times 10^6$	$198 \times 10^{-9}$	$220 \times 10^{-9}$

depicts the time-dependent fluxes for the simulations shown in Figure 4. Interstate fluxes shown in the left column are calculated from eq 8, and the intersite fluxes in the right column are calculated from eq 9. Only a limited number of interstate fluxes contribute significantly to the derived intersite fluxes, since only transfer events between next-neighbor sites contribute to the electron transfer between the four hemes and the SP on the sub-microsecond time scale. The next-neighbor constraint, in turn, is a consequence of the strong dependence of the electron-transfer rate constants on the cofactor distances as expressed in eq 11. This strong dependence on the cofactor distances led to the suggestion that these distances are the primary design factor in electron-transfer proteins.<sup>8,9</sup>

Details of the transfer steps present in our simulations are depicted in Figure 6. For each of the four simulations, all possible microstates are depicted (Figures 4a–d). Filled and open circles denote the reduced and oxidized forms of a redox site, respectively. The order of the microstates is given from top to bottom as heme c<sub>554</sub>, heme c<sub>556</sub>, heme c<sub>552</sub>, heme c<sub>559</sub>, and the SP. The microstates depicted in the uppermost rows of each graph a–d represent the starting configurations of the simulations. The microstates shown in the bottom rows represent the populations of microstates present at the end of the simulations. Only microstates contributing more than 0.1% to the starting or end configurations are considered. The row(s) in between the top and the bottom rows show all other possible microstates of the simulation. For the starting and final microstates the starting and final probabilities (in %) are given in parentheses, respectively. For intermediate states, the values in parentheses denote the maximal probability observed during the simulation. Fluxes contributing significantly are indicated by arrows, and their maximum values are given. Each arrow indicates a net transition from one microstate to another, i.e., from one charge configuration of the system to another. As explained in detail, such a transition represents a transfer of an electron from one redox site to another and thus contributes to an intersite flux. This contribution is indicated by the color of the arrows, where black indicates electron transfer between the SP and heme c<sub>559</sub>, blue indicates transfer between heme c<sub>559</sub> and heme c<sub>552</sub>, red indicates transfer between heme c<sub>552</sub> and heme c<sub>556</sub>, and green indicates transfer between heme c<sub>556</sub> and heme c<sub>554</sub>.

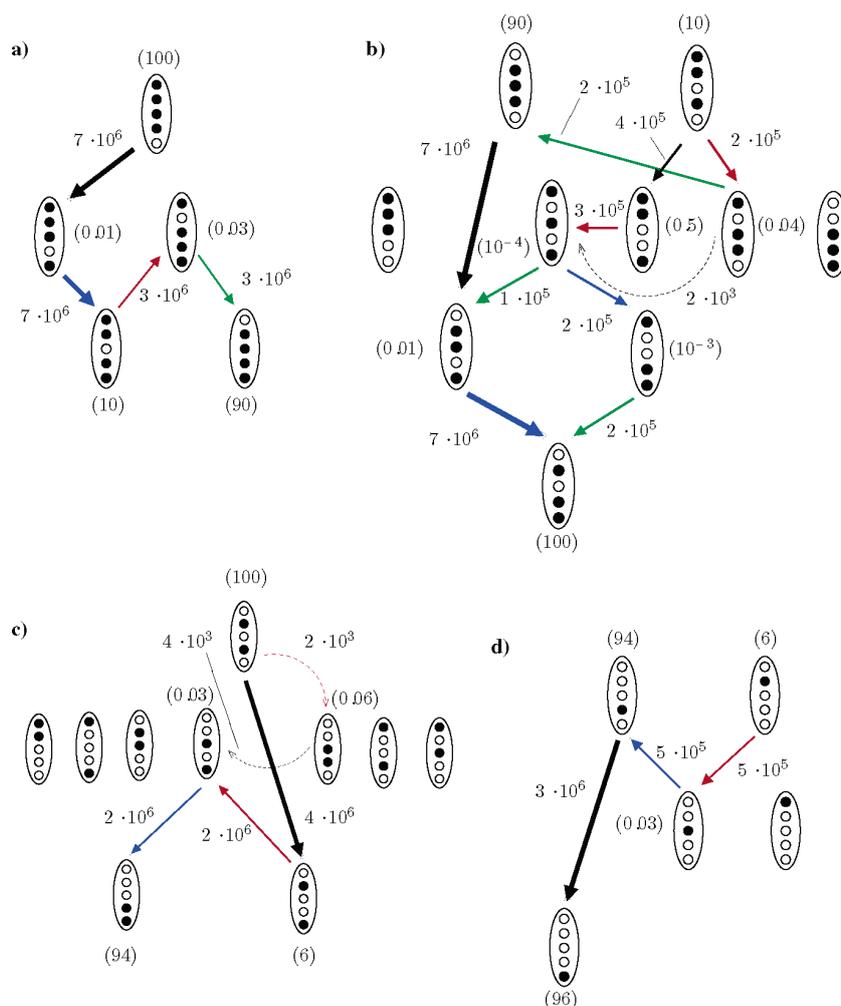
The first simulation, Figure 6a, already gives an example of the next-neighbor character of electron transfer between the four hemes and the SP. Initially, all four hemes are reduced, and the SP is oxidized. Re-reduction of the SP, as known from experiments, occurs via an electron transfer from heme c<sub>559</sub>, i.e., via the microstate transition (0,1,1,1,1) → (1,0,1,1,1). This initial transfer is rapidly followed by an electron transfer from heme c<sub>552</sub> to heme c<sub>559</sub> ((1,0,1,1,1) → (1,1,0,1,1)). The necessary intermediate state (1,0,1,1,1) is only transiently populated (maximum ≈ 0.01%) and could thus not be observed experimentally. Further relaxation of the system toward its equilibrium distribution again follows the next-neighbor rule. A direct electron transfer from heme c<sub>554</sub> to heme c<sub>552</sub>, although



**Figure 5.** Probability fluxes for the kinetic simulations of Figure 4. Interstate fluxes shown in the left column were calculated according to eq 8. Intersite fluxes calculated from eq 9 are shown in the right column. The state vector is given in the order (SP,c<sub>559</sub>,c<sub>552</sub>,c<sub>556</sub>,c<sub>554</sub>).

energetically favorable, does not occur due to the large distance between these cofactors ( $\sim 37$  Å). As a consequence, the transition to the lowest-energy state (1,1,1,0) occurs via an uphill transfer step from heme c<sub>556</sub> to heme c<sub>552</sub> ((1,1,0,1,1)  $\rightarrow$  (1,1,1,0,1)) followed by the energetically favorable transfer from heme c<sub>554</sub> to heme c<sub>556</sub> ((1,1,1,0,1)  $\rightarrow$  (1,1,1,1,0)). Again, the intermediate state is only transiently populated and thus not accessible to experimental observations.

The second simulation as seen in Figure 4b presents a similar picture for the time dependence of the population of accessible microstates. Starting from a population of the two microstates (0,1,1,1,0) (90%) and (0,1,0,1,1) (10%) the system relaxes toward an equilibrium distribution that is mainly given by one microstate (1,1,0,1,0). The underlying transfer dynamics of the system as depicted in Figure 6b, however, are considerably more complex. The highly populated initial state (0,1,1,1,0) can



**Figure 6.** Reaction scheme for the charge transfer in the RC as deduced from the flux analysis of our simulations. Each oval represents a microstate of the system. The circles symbolize the redox cofactors in the order heme  $c_{554}$ , heme  $c_{556}$ , heme  $c_{552}$ , heme  $c_{559}$ , and SP from the top to the bottom. Filled and open circles denote the reduced and oxidized forms of the sites, respectively. Panels a–d describe the reaction scheme after the first to fourth flash, respectively. Each has a particular order: Initial states are on top, intermediate states are in the middle, and final states are at the bottom. The initial, final, and maximum probabilities are given in parenthesis for initial, final, and intermediate microstates, respectively. Interstate fluxes significantly contributing to the kinetic behavior are indicated by arrows. The associated intersite transfer is indicated by color: Black indicates transfer between heme  $c_{559}$  and the SP, blue between heme  $c_{552}$  and heme  $c_{559}$ , red between heme  $c_{556}$  and heme  $c_{552}$ , and green between heme  $c_{554}$  and heme  $c_{556}$ .

rapidly decay into the final state via just one intermediate, (1,0,1,1,0). In contrast, the initial state (0,1,0,1,1) has to relax toward the final state via a succession of several intermediates due to the next-neighbor restriction discussed already. Again, these intermediate states are only transiently populated. Each interstate flux into one of these intermediates is accompanied by an equally high flux out of these intermediates. For example, the transition from the initial state to the intermediate (1,0,0,1,1) is rapidly followed by a transition to a second intermediate state (1,0,1,0,1). This intermediate state in turn decays either into state (1,0,1,1,0) via an electron transfer from heme  $c_{554}$  to heme  $c_{556}$  or alternatively to state (1,1,0,0,1) via electron transfer from heme  $c_{552}$  to heme  $c_{559}$ . As can be seen from the arrow colors in Figure 6b, electron transfer between two given sites is represented by more than one interstate transition. This multiplicity is the main obstacle in defining rate constants for transfer reactions between two given sites. The presented microstate formalism, however, naturally incorporates this multiplicity and thus avoids the common ambiguities.

The experimental setup resembled in the last two simulations c and d, i.e., transfer with two or one reduced cofactors present

in the system, has been widely discussed in the context of uphill electron transfer.<sup>10,28,63</sup> In Figure 4c, for example, it can be seen that the reduction of the SP is accompanied by the (partial) oxidation of heme  $c_{559}$  and heme  $c_{556}$  whereas heme  $c_{552}$  remains basically oxidized throughout. This behavior, however, must not be taken as evidence for electron transfer between heme  $c_{556}$  and heme  $c_{559}$  or even between heme  $c_{556}$  and the SP. Again, the distances between heme  $c_{556}$  and heme  $c_{559}$  ( $\sim 22$  Å) and between heme  $c_{556}$  and the SP ( $\sim 37$  Å) render electron-transfer rates based on eq 11 too slow to contribute significantly to the re-reduction kinetics of the SP. This result is in agreement with mutational studies on the RC of *B. viridis* that suggested that electron transfer always occurs along the sequence  $c_{556}$ – $c_{552}$ – $c_{559}$ –SP.<sup>28</sup> Thus, oxidation of heme  $c_{556}$  should be accompanied by reduction of heme  $c_{552}$  although heme  $c_{552}$  is observed to remain basically oxidized throughout the simulation. This apparent contradiction arising from the experimentally accessible data can be readily resolved in the microstate formalism. As discussed already, only next-neighbor electron-transfer contributes to the transitions in our simulations due to the strong distance dependence in the rate formula (eq 11). The intermedi-

ate states showing a reduced heme  $c_{552}$ , however, are only transiently populated. The maximal probabilities for the intermediates in the simulations of Figure 4c and 4d amount to  $\sim 0.02\%$  for states (1,0,1,0,0) and (0,0,1,0,0), and thus, reduction of heme  $c_{552}$  is hardly accessible to experimental observation in these kinetic experiments.

## Conclusions

In the present article, we present a microstate formalism to simulate charge-transfer systems. The system is described by state vectors, indicating the redox form of each site possibly taking part in the transfer process. In this approach, charge transfer is not primarily described as a transfer of a charge from one site to another but as a change in the state vector of the complete system. Thus, there are as many possible transfer reactions between two given sites as there are possible configurations of all other sites. A microstate formalism naturally takes into account the dependence of transfer rates between sites on their interaction with the surrounding charges. For the interconversion between two microstates, a unique charge-transfer rate can be defined. This formalism allows us to describe the kinetics of a charge-transfer system by a master equation describing the time evolution of microstates. The possibly complex kinetic behavior of single sites of the system can be deduced from the time-dependent probability distribution of microstates. Charge flow within the system can be investigated by calculating elementary fluxes between microstates and relating them to the accompanied charge transfer between sites. Thus, the transfer network can be built up by elementary fluxes. Poisson–Boltzmann electrostatics together with existing heuristic rate laws<sup>6,8</sup> can be used to describe biological electron-transfer systems. The overall kinetics of the system are simulated using coupled differential equations. In this work, the re-reduction of the SP by the C-subunit of the RC serves as a model system to compare the simulations with experiments. All parameters required for calculating electron-transfer rates can be obtained from calculations using structural data. It is therefore possible to simulate the kinetics of an electron-transfer system solely based on structural information without referring to additional experimental data. Our simulations reproduce the experimentally observed re-reduction kinetics of the SP. The calculated lifetimes of the photo-oxidized state of the SP agree well with the experimentally determined values.

All calculations put together in this article to obtain a coherent scheme for simulating the kinetics of an electron-transfer system use standard computational techniques that have been demonstrated to reproduce experimental results for a broad range of biological electron-transfer systems. The formalism presented here should, thus, be readily applicable to a large class of related charge-transfer systems.

**Acknowledgment.** This work was supported by the Deutsche Forschungsgemeinschaft (Grant No. UL 174/7-1). We thank Donald Bashford for providing his program MEAD and useful discussions. R.T.U. thanks the Universität Bayern e.V. for a fellowship.

## References and Notes

- Nicholls, D. G.; Ferguson, S. J. *Bioenergetics 3*; Academic Press: New York, 2002.
- Saraste, M. *Science* **1999**, *283*, 1488–1493.
- Lancaster, C. R. D.; Michel, H. In *Handbook of Metalloproteins*; Albrecht Messerschmidt, A., Huber, R., Poulos, T., Wieghardt, K., Eds.; Wiley: New York, 2001; pp 119–135.
- Marcus, R. A. *J. Chem. Phys.* **1956**, *24*, 966–978.
- Marcus, R. A. *J. Chem. Phys.* **1956**, *24*, 979–989.
- Moser, C. C.; Keske, J. M.; Warncke, K.; Farid, R. S.; Dutton, P. L. *Nature* **1992**, *355*, 796–802.
- Moser, C. C.; Page, C. C.; Chen, X.; Dutton, P. L. *J. Biol. Inorg. Chem.* **1997**, *2*, 393–398.
- Page, C. C.; Moser, C. C.; Chen, X.; Dutton, P. L. *Nature* **1999**, *402*, 47–52.
- Page, C. C.; Moser, C. C.; Dutton, P. L. *Curr. Opin. Chem. Biol.* **2003**, *7*, 551–556.
- Alric, J.; Lavergne, J.; Rappaport, F.; Verméglio, A.; Matsuura, K.; Shimada, K.; Nagashima, K. V. *J. Am. Chem. Soc.* **2006**, *128*, 4136–4145.
- Noy, D.; Moser, C. C.; Dutton, P. L. *Biophys. Biochem. Acta* **2006**, *1757*, 90–105.
- Bashford, D.; Gerwert, K. *J. Mol. Biol.* **1992**, *224*, 473–486.
- Ullmann, G. M.; Knapp, E.-W. *Eur. Biophys. J.* **1999**, *28*, 533–551.
- Sharp, K. E. *Biophys. J.* **1998**, *73*, 1241–1250.
- Onufriev, A.; Case, D. A.; Ullmann, G. M. *Biochemistry* **2001**, *40*, 3413–3419.
- Ullmann, G. M. *J. Phys. Chem. B* **2003**, *107*, 1263–1271.
- Alric, J.; Cuni, A.; Maki, H.; Nagashima, K. V.; Verméglio, A.; Rappaport, F. *J. Biol. Chem.* **2004**, *279*, 47849–47855.
- Sham, Y. Y.; Muegge, I.; Warshel, A. *Proteins* **1999**, *36*, 484–500.
- Nogi, T.; Hirano, Y.; Miki, K. *Photosynth. Res.* **2005**, *85*, 87–99.
- Deisenhofer, J.; Epp, O.; Miki, K.; Huber, R.; Michel, H. *J. Mol. Biol.* **1984**, *180*, 358–398.
- Deisenhofer, J.; Epp, O.; Miki, K.; Huber, R.; Michel, H. *Nature* **1985**, *318*, 618–624.
- Deisenhofer, J.; Epp, O.; Sinning, I.; Michel, H. *J. Mol. Biol.* **1995**, *246*, 429–457.
- Dracheva, S. M.; Drachev, L. A.; Konstantinov, A. A.; Semenov, A. Y.; Skulachev, V. P.; Arutjunjan, A. M.; Shuvalov, V. A.; Zaberezhnaya, S. N. *Eur. J. Biochem.* **1988**, *171*, 253–264.
- Fritzsch, G.; Buchanan, S.; Michel, H. *Biochim. Biophys. Acta* **1989**, *977*, 157–162.
- Ortega, J.; Mathis, P. *Biochemistry* **1993**, *32*, 1141–1151.
- Rappaport, F.; Béal, D.; Verméglio, A.; Joliot, P. *Photosynth. Res.* **1998**, *55*, 317–323.
- Ortega, J.; Drepper, F.; Mathis, P. *Photosynth. Res.* **1999**, *59*, 147–157.
- Chen, I.-P.; Mathis, P.; Koepke, J.; Michel, H. *Biochemistry* **2000**, *39*, 3502–3602.
- Gunner, M.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 9151–9155.
- Scharnagl, C.; Hettenger, J.; Fischer, S. F. *Int. J. Quantum Chem.* **1994**, *52*, 33–56.
- Onufriev, A.; Smondyrev, A.; Bashford, D. *J. Mol. Biol.* **2003**, *332*, 1183–1193.
- Ferreira, A.; Bashford, D. *J. Am. Chem. Soc.* **2006**, *128*, 16778–16790.
- Ullmann, G. M. *J. Phys. Chem. B* **2000**, *104*, 6293–6301.
- Marcus, R. *J. Chem. Phys.* **1963**, *38*, 1858–1862.
- Marcus, R. A.; Sutin, N. *Biochim. Biophys. Acta* **1985**, *811*, 265–322.
- Williams, R. J. P. *J. Solid State Chem.* **1999**, *145*, 488–495.
- Olsson, M. H. M.; Ryde, U.; Roos, B. O. *Protein Sci.* **1998**, *81*, 6554–6558.
- Ryde, U.; Olsson, M. H. M. *Int. J. Quantum Chem.* **2001**, *81*, 335–347.
- Higuchi, Y.; Kusunoki, M.; Matsuura, N.; Yasuoka, N.; Kakudo, M. *J. Mol. Biol.* **1984**, *172*, 109–139.
- Brünger, A. T.; Karplus, M. *Proteins* **1988**, *4*, 148–156.
- Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *1200*, 58.
- Perdew, P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671.
- Guerra, C. F.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Acta* **1998**, *391*, 391–403.
- Breneman, C. N.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11* (3), 361–373.

- (47) Chen, J. L.; Noodleman, L.; Case, D.; Bashford, D. *J. Phys. Chem.* **1994**, *98*, 11059–11068.
- (48) Rabenstein, B.; Ullmann, G. M.; Knapp, E. W. *Biochemistry* **1998**, *37*, 2488–2495.
- (49) Rabenstein, B.; Ullmann, G. M.; Knapp, E. W. *Biochemistry* **2000**, *39*, 10487–10496.
- (50) Taly, A.; Sebban, P.; Smith, J. C.; Ullmann, G. M. *Biophys. J.* **2003**, *84*, 2090–2098.
- (51) Bashford, D.; Case, D. A.; Dalvit, C.; Tennant, L.; Wright, P. E. *Biochemistry* **1993**, *32*, 8045–8056.
- (52) Ullmann, G. M.; Knapp, E. W. *Eur. Biophys. J.* **1999**, *28*, 533–551.
- (53) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (54) Ryde, U.; Olsson, M. H. M.; Roos, B. O.; Borin, C. A. *Theor. Chem. Acc.* **2001**, *105*, 452–462.
- (55) Oppenheim, I.; Shuler, K. E.; Weiss, G. H. *Stochastic Processes in Chemical Physics: The Master Equation*; MIT Press: Cambridge, MA, 1977.
- (56) van Kampen, N. *Stochastic Processes in Physics and Chemistry*; North-Holland: New York, 1981.
- (57) Anderson, E.; Bai, Z.; Bischof, C.; Blackford, S.; Demmel, J.; Dongarra, J.; Du Croz, J.; Greenbaum, A.; Hammarling, S.; McKenney, A.; Sorensen, D. *LAPACK Users' Guide*, 3rd ed.; Society for Industrial and Applied Mathematics: Philadelphia, PA, 1999.
- (58) Press, W.; Teukolsky, S.; Vetterling, W.; Flannery, B. *Numerical Recipes in C*, 2nd ed.; Cambridge University Press: New York, 1992.
- (59) Voigt, P.; Knapp, E.-W. *J. Biol. Chem.* **2003**, *278*, 51993–52001.
- (60) Muegge, I.; Qi, P.; Wand, A.; Chu, Z.; Warshel, A. *J. Phys. Chem. B* **1997**, *101* (5), 825–836.
- (61) Parson, W.; Chu, Z.; Warshel, A. *Biophys. J.* **1998**, *74* (1), 182–191.
- (62) Mertz, E. L.; Krishtalik, L. I. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2081–2086.
- (63) Nogi, T.; Miki, K. *J. Biochem.* **2001**, *130*, 319–329.



# List of Abbreviations

ADF	Amsterdam Density Functional
ADP	adenosine diphosphate
AfAmt-1	ammonium transporter Amt-1 from <i>Archaeoglobus fulgidus</i>
AGPL	Affero General Public License (a free software license from GNU)
Amt	ammonium transporter
ATP	adenosine triphosphate
bRC	bacterial photosynthetic reaction center
CHARMM	Chemistry at HARvard Molecular Mechanics
DFT	density functional theory
DTI	discrete thermodynamic integration
EcAmtB	ammonium transporter AmtB from <i>Escherichia coli</i>
FEP	free energy perturbation
GCEM	Generalized Continuum Electrostatics Model
GMCT	Generalized Monte Carlo Titration
GNU	GNU is Not Unix (reverse acronym), GNU is a free software project that provides a Unix-like operating system
HIV	human immunodeficiency virus

LPBE	linearized Poisson-Boltzmann equation
MC	Monte Carlo
MDC	multipole derived charge analysis
MEAD	Macroscopic Electrostatics with Atomic Detail
NAD	oxidized nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NeRh50	ammonium/ammonia transporter Rh50 from <i>Nitrosomonas europae</i>
NEW	non-equilibrium work
NMR	nuclear magnetic resonance
PaAz	azurin from <i>Pseudomonas aeruginosa</i>
PDB	Brookhaven Protein Data Bank ( <a href="http://www.rcsb.org/pdb">www.rcsb.org/pdb</a> )
pmf	proton-motive force
Rh	Rhesus (in Rhesus proteins)
TI	thermodynamic integration
UQ	ubiquinone
UQH <sub>2</sub>	ubiquinol
VMD	Visual Molecular Dynamics

# Originalitätserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet habe.

Ferner erkläre ich, dass ich nicht bereits anderweitig mit oder ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

Bayreuth, den 3. Dezember 2012,

R. Thomas Ullmann