Optimal Network Generation for the Simulation of Proton Transfer Processes

Marco Reidelbach

Dissertation zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

Eingereicht im Fachbereich
Mathematik und Informatik der Freien Universität Berlin

Berlin, 2018

This doctoral thesis was created within the Institute of theoretical Physics of the Freie Universität Berlin and the Computational Molecular Design group of the Zuse Institute Berlin.

Erstgutachter (Betreuer)
PD Dr. Marcus Weber
Zuse Institut Berlin
Computational Molecular Design
Takustraße 7
14195 Berlin, Deutschland

Zweitgutachter
PD Dr. Konstantin Fackeldey
Technische Universität Berlin
Institut für Mathematik
Straße des 17. Juni 136
10623 Berlin, Deutschland

Drittgutachter
Prof. Dr. Petra Imhof
Freie Universität Berlin
Institut für theoretische Physik
Arnimallee 14
14195 Berlin, Deutschland

Tag der Disputation: 30.04.2019

Abstract

The *oxidative phosphorylation* is the most important step of the aerobic respiration. Here, electrons are transferred along several membrane-embedded enzyme complexes to finally reduce molecular oxygen to water in *Cytochrome c Oxidase*. The energy released by the electron flow and the oxygen reduction is used to translocate protons across the membrane. Thereby, an electro-chemical gradient is established which subsequently drives the synthesis of the biological energy carrier *Adenosine Triphosphate*. Numerous diseases, e.g. *Alzheimer* or *Parkinson*, are partially assigned to malfunctions of the *oxidative phosphorylation*, rendering a detailed understanding of this step indispensable.

Common computational investigations, employing *Molecular Dynamics* simulations, enhanced sampling techniques, or transition pathway finding algorithms, are limited in their description of quantum effects and often only provide a limited, or biased, description of complex reactions. The alternative *Transition Network* approach, divides a complex reaction of interest into numerous simpler transitions, connecting various local potential energy minima of the *potential energy surface* by *minimum energy pathways*, yielding a *Transition Network*, or *simple graph*, which can be analyzed in terms of optimal transition pathways by standard graph theoretical algorithms. A major drawback of the *Transition Network* approach is the exponential increase of stationary points of the *potential energy surface* with increasing numbers of *degrees of freedom* to sample, rendering the *Transition Network* approach infeasible for most complex reactions.

Within the framework of this thesis two methods optimizing the *Transition Network* approach are developed and extensively tested in small proton transfer model systems. These are: 1) the *TN-MD method* coupling the discrete sampling of states separated by substantial energy barriers with *Molecular Dynamics* simulations for the sampling of states separated by minor energy barriers, e.g. amino acid side chain dihedral angle rotations or water molecule translations, respectively, and 2) the *TN prediction method* using a known, initial *Transition Network* and an excessive two-step coarse-graining procedure for the determination of an unknown *Transition Network* in a perturbed environment, e.g. different protonation states. Both methods provide significant cost reductions, while important properties of the proton transfer reactions, e.g. rate-determining, maximal transition barriers and the variability of transition pathways or mechanisms, are maintained.

Furthermore, the *TN-MD method* is used to investigate the proton transfer through the D-channel of *Cytochrome c Oxidase* in a minimal model system, providing various proton transfer pathways with rate-determining, maximal transition barriers which are in agreement to computational results from *Liang et al* using a much more involved model and simulation setup. Other decisive aspects of previous proton transfer investigations along the D-channel, e.g. the identity of the rate-determining, maximal transition state and the behavior of the proposed asparagine gate, are reproduced and extended.

Overall, this thesis provides a methodical leap forward in terms of the *Transition Network* approach as well as another piece of analysis required for the understanding of the proton translocation along the *oxidative phosphorylation*.

Publications

Related publications:

1. **M Reidelbach**, F Betz, RM Mäusle, and P Imhof. Proton transfer pathways in an aspartate-water cluster sampled by a network of discrete states. Chemical Physics Letters 659: 169-175, 2016.

2. M Bagherpoor Helabad, T Ghane, **M Reidelbach**, AL Woelke, EW Knapp, and P Imhof. Protonation-State-Dependent Communication in Cytochrome c Oxidase. Biophysical Journal 113: 817-828, 2017.

3. T Ghane RF Gorriz, S Wrzalek, S Volkenandt, F Dalatieh, **M Reidelbach**, and P Imhof. Hydrogen-Bonded Network and Water Dynamics in the D-channel of Cytochrome c Oxidase. Journal of Membrane Biology 251: 299-314, 2018.

4. **M Reidelbach**, M Weber, and P Imhof. Prediction of perturbed proton transfer networks. PLOS ONE 13: e0207718.

Other publications:

1. V Spahn, G Del Vecchio, D Labuz, A Rodriguez-Gaztelumendi, N Massali, J Temp, V Durmaz, P Sabri, **M Reidelbach**, H Machelska, M Weber, and C Stein. A nontoxic pain killer designed by modeling of pathological receptor conformations. Science 355: 966-969, 2017.

2. V Spahn, G Del Vecchio, A Rodriguez-Gaztelumendi, J Temp, D Labuz, M Kloner, **M Reidelbach**, H Machelska, M Weber, and C Stein. Opioid receptor signaling, analgesic and side effects induced by a computationally designed pH-dependent agonist. Scientific Reports 8: 8965, 2018.

Contents

Chapter 1

Introduction

Life, as we know it, shows itself in a variety of genetically interwoven ways, due to a hypothesized common ancestor [1], rendering any attempt of classification difficile. In 1990 a three-domain classification, the *Woese-classification*, was introduced, in which all living species are classified according to their *ribosomal ribonucleic acid* (*rRNA*) as *Archea*, *Bacteria*, or *Eucarya* [2]. Despite the wide acceptance of this classification other propositions, including only two or up to five domains, are still discussed prominently [3, 4, 5, 6, 7, 8, 9]. The classification gets even more problematic when trying to define subdomains within either domain [10, 11]. Pointing out structural similarities and common operating features among all species, on the other hand, is much simpler. In fact, the cells of all living organisms: 1) possess a plasmamembrane enclosing their cytoplasm, 2) store their hereditary information on double-stranded *deoxyribonucleic acid* (*DNA*) molecules, 3) replicate and express their *DNA* similar, 4) use proteins as catalysts, and 5) require energy to maintain their vitality. Basis for all these similarities is the presence of a common set of simple molecular building blocks, e.g. sugars, amino acids, and nucleotides, as well as the ability to manipulate them [12].

Adenosine Triphosphate (*ATP*) is one of the molecules present in any living cell [1]. Discovered in 1929 [13, 14], *ATP* is involved in a variety of processes, e.g. serving as a building block in *RNA* and *DNA* synthesis [12], activating amino acids in the protein synthesis [15], and participating in intra- and extracellular signaling [16, 17]. Its most important role, however, is the storage of energy within the cells. A number of reactions, occurring under cellular conditions, is energetically unfavorable, i.e. increasing the Gibbs free energy ($\Delta G > 0$) [18]. By coupling these energetically unfavorable reactions to the energetically favorable hydrolysis of the high-energy bonds within *ATP* ($\Delta G^{\circ\prime} = -7.3$ kcal/mol or $\Delta G = -12$ kcal/mol for

the hydrolysis of ATP to *Adenosine Diphosphate* (ADP) and phosphate (P_i) under standard or cellular conditions, respectively) they become energetically favorable ($\Delta G < 0$) [1]. Therefore, an ATP concentration of $1 - 10$ mmol/L is usually maintained within either cell [19].

$$A + B \Rightarrow C \qquad \Delta G > 0 \tag{1.1}$$

$$ATP \Rightarrow ADP + P_i \qquad \Delta G < 0 \tag{1.2}$$

$$A + B + ATP \Rightarrow C + ADP + P_i \qquad \Delta G < 0 \tag{1.3}$$

Lipids, proteins and polysaccharides are the source of most cellular energy. Following their digestion, the respective monomers are gradually oxidized to synthesize ATP [20]. The generation of ATP from the monomeric sugar *glucose,* for example, involves three distinct steps in aerobic organisms, the *glycolysis*, the *citric acid cycle* and the *oxidative phosphorylation*. During the *glycolysis*, present in all cells today, a *glucose* molecule is converted into two *pyruvate* molecules with a net gain of two ATP molecules. Furthermore, two molecules of the coenzyme *Nicotinamide Adenine Dinucleotide* (NAD, oxidized NAD^+, reduced $NADH$) are reduced. With the help of coenzyme A (CoA) a carbon atom is removed from either *pyruvate* molecule and released as CO_2 molecule, while the remaining carbon atoms are added to CoA to form two molecules of *acetyl-CoA*. Once again, two molecules of NAD^+ are reduced. Both *acetyl-CoA* molecules then enter the *citric acid cycle*, completing the oxidation of the *glucose* molecule to six CO_2 molecules. Along the *citric acid cycle* one *Guanosine Triphosphate* (GTP) molecule is synthesized and three NAD^+ molecules and one *flavin adenine dinucleotide* (FAD, oxidized FAD, reduced $FADH_2$) molecule are reduced per *acetyl-CoA* molecule. The *oxidative phosphorylation* concludes the generation of ATP. Here, the reduced coenzymes $NADH$ and $FADH_2$ are reoxidized, with their electrons being transferred along an *electron transport chain*, including several enzyme complexes, to reduce O_2 to 2 H_2O. Per $FADH_2$ or $NADH$ molecule two or three ATP molecules are synthesized, respectively, giving an overall yield of 38 ATP molecules per *glucose* molecule [1, 20]. The breakdown of other monomers, e.g. fatty acids along the β-oxidation, and anaerobic pathways, e.g. the sulfate respiration, can be found in [21, 22].

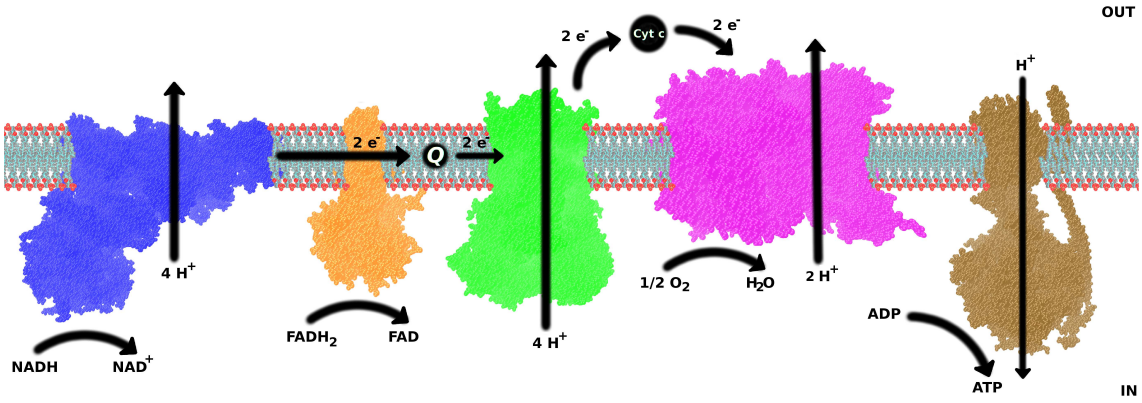


Figure 1.1: Schematic *electron transport chain*, involving *Complex I* (blue), *Complex II* (orange), *Complex III* (green), and *Complex IV* (magenta), plus *Complex V* (brown) along an arbitrary membrane, separating an inner and outer side. The soluble proteins coenzyme *Q* and *cytochrome c* are depicted by black spheres, labelled *Q* and *cyt c*, respectively. Black arrows indicate the individual reactions of the *electron transport chain*.

Approximately 90 % of the energy present in the chemical bonds of a *glucose* molecule are stored in *NADH* and $FADH_2$ molecules and subsequently used to synthesize *ATP* [23]. Hence, a more detailed view on the *oxidative phosphorylation* is expedient (cf. Fig. 1.1). The transfer of each electron pair from *NADH* to O_2 (two electron pairs are required for the reduction of O_2 to *2* H_2O) yields a free energy change of $\Delta G^{\circ\prime} = -52.5$ kcal/mol [24]. To ensure that most of the energy is stored, and not lost due to heat dissipation, the energy is only released in small increments. Four enzymatic transmembrane complexes form, together with some soluble proteins, the *electron transport chain*, either at the inner mitochondrial membrane (*eukaryotes*) or the plasmamembrane (many aerobic *bacteria*), which is in line with the *endosymbiotic theory* [25]. The four transmembrane complexes are: 1) *Complex I* or *NADH dehydrogenase*, 2) *Complex II* or *succinate dehydrogenase*, 3) *Complex III* or *CoQH2-cytochrome c reductase*, and 4) *Complex IV* or *cytochrome c oxidase* (*CcO*) [23]. At first, a hydride ion is removed from the *NADH* molecule and split into a proton and two high-energy electrons. The electrons are then transferred individually through *Complex I* to reduce coenzyme *Q* (*CoQ*) to $CoQH_2$. The energy released ($\Delta G^{\circ\prime} = -16.6$ kcal/mol [24]) is used to translocate four protons along distinct, water-filled cavities, so called *proton transfer channels*, across the membrane. Hence, *Complex I* is a true proton pump [26]. In a next step $CoQH_2$ donates its two electrons to *Complex III*, which transfers the electrons further to two *cytochrome c* molecules, thereby reducing them. Again the energy released ($\Delta G^{\circ\prime} = -10.1$ kcal/mol [24]) is used to translocate four protons across

the membrane. However, since the translocation of protons through *Complex III* is facilitated by a *redox loop*, the complex is not considered a true proton pump [26]. From *cytochrome c* the electrons are transferred individually to *Complex IV* to finally reduce O_2 ($\Delta G^{\circ\prime} = -25.8$ kcal/mol [24]). This time, only two protons are translocated across the membrane per electron pair, the remaining two are consumed during the O_2 reduction. Still, all four protons are translocated through water-filled *proton transfer channels*, rendering *Complex IV* a true proton pump [26]. For $FADH_2$ as electron donor *Complex II* is used instead of *Complex I* to reduce *CoQ* at the beginning of the *electron transport chain*. However, the energy released by this reaction is not sufficient to translocate protons across the membrane [24]. Thus, six or ten protons are translocated overall per electron pair coming from $FADH_2$ or *NADH*, respectively. This charge separation establishes a pH gradient ($\Delta pH \approx 1$ [27]) and a membrane potential ($\Delta E \approx 140$ mV [27]) across the membrane, usually referred to as electrochemical gradient following the *chemiosmotic hypothesis* proposed by Peter Mitchel in 1961 [28]. The subsequent backflow of protons through *Complex V* (*ATP-synthase*) drives the synthesis of more than 100 *ATP* molecules per second and complex. Thereby, a backflow of three to four protons is required for the synthesis of a single *ATP* molecule [23, 24, 27].

The *oxidative phosphorylation* is, in all its complexity, a truly marvelous process. At the same time, however, the complexity is its biggest drawback, always bearing the risk of an individual sub-step going wrong and thus disrupting the entire process. Among others, potential sources of error may arise from gene mutations, toxic compounds and simply aging [29]. Nearly sixty years ago, the first human disease, today known as *mitochondrial myopathy*, was assigned to a malfunction of the *oxidative phosphorylation* [30, 31]. Since then, numerous human diseases ranging from dry and brittle hair [29], to Diabetes [32], Cancer [33], Alzheimer [34], and Parkinson [34] were partially assigned to various defects along the *oxidative phosphorylation*, e.g. Parkinson and *Complex I* [35] or Alzheimer and *Complex IV* [36]. Due to the complexity of the process, similar human diseases may arise from different defects, while similar defects may lead to different human diseases [29]. Therefore, an extensive structure-function analysis of each complex along the *oxidative phosphorylation* is required, individually and in combination with others [33], to elucidate the pathogenic diseases, guide their diagnosis, and develop therapies [34]. On the other hand, the enhanced knowledge of the *oxidative phosphorylation* process provides new target pathways for anti-bacterial compounds, e.g. *diarylquinolines* against the multidrug-resistant *Staphylococcus aureus* [37].

Traditionally, the structure-function analysis of proteins is based on sets of experimentally derived structures [38, 39, 40, 41]. However, these *experimental ensembles* provide in many cases only a limited, and thereby biased, view on the dynamics of the proteins and the role of internal water molecules [42]. Other experimental techniques are often constrained in their structural and temporal resolution [43], e.g. when considering the translocation of protons through highly connected hydrogen bond networks inside *proton transfer channel*s. Theoretical investigations, employing simulations, permit a broader and more detailed view on the dynamics [42]. The first molecular dynamics (*MD*) simulation of a biologically relevant protein, the *bovine pancreatic trypsin inhibitor*, was performed in 1977 [44]. Although the simulation was only 10 ps long and performed without modeling the solvent environment of the protein it replaced the idea of rigid structure proteins [45]. Today, *MD* simulations with several hundred thousand atoms, e.g. solvated [46, 47] or membrane embedded proteins [48, 49] and others [50, 51], are common, reaching timescales up to milliseconds [52]. Apart from investigations of the dynamics, *MD* simulations are used to study thermodynamic properties, e.g. entropy and free energy, in a highly efficient, and therefore widely used, manner [53]. Despite its enormous development, the basic idea behind the *MD* simulation is maintained. Then as now, the *MD* simulation of macromolecules relies on an initial structure, derived from experiments [54] or (nowadays) computational modeling [55], and a *force field* [42], e.g. AMBER [56], GROMOS [57], or CHARMM [58]. The *force field* (or potential energy function) describes the dependence of a system's energy on the position of its atoms. The function is complex, but easy to solve [42], describing the stretching and bending of bonds and angles between two and three atoms, respectively, using harmonic springs, the torsion of dihedral angles, spanned by four atoms, using periodic functions and the pairwise interaction of non-bonded atoms by the *Lennard-Jones* potential for *van der Waals* interactions and the *Coulomb* potential for electrostatic interactions [42, 59] (cf. Fig 1.2). Parameters are derived from experiments or quantum mechanical calculations [60]. The potential energy function allows the determination of the potential energy of a specific conformation. Its first derivative, with respect to either atom position, provides the force acting on each atom. The numerical integration of Newton's 2nd law of motion then allows the propagation of each atom position and momenta in time [59].

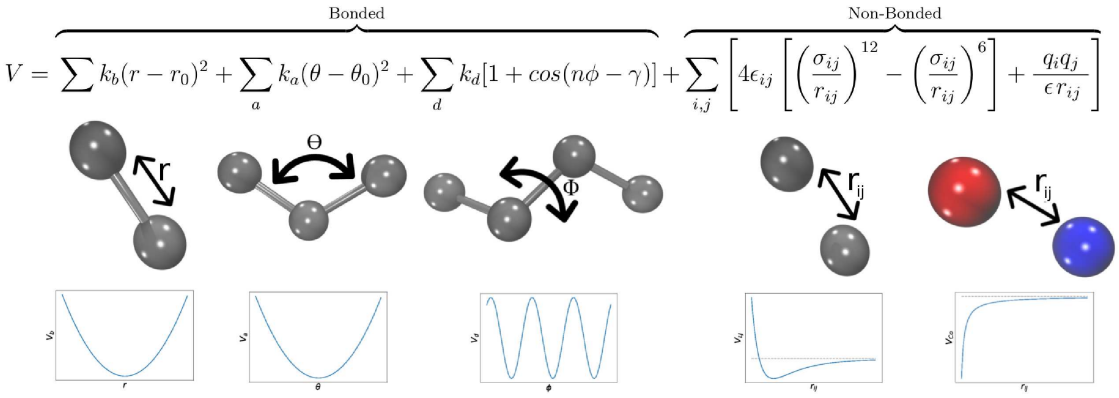


Figure 1.2: Exemplary potential energy function V and individual representations of the bond potential V_b, the angle potential V_a, the dihedral angle potential V_d, the Lennard-Jones potential V_{LJ}, and the Coulomb potential V_{Co}.

Two challenges remain when performing MD simulations. These are: 1) the accurate description of quantum effects and 2) the problem of an adequate sampling of rare transitions in the conformation space [61, 62]. Although, continually improved [63, 64, 65, 66, 67, 68], *force fields* are poorly suited for systems, in which quantum effects play an essential role, e.g. when covalent bonds need to be broken to facilitate the transfer of protons. To compensate these limitations, it is possible to treat the region of interest, e.g. the *proton transfer channel*, quantum-mechanically (*QM*) employing *ab initio* [69, 70] or *semi-empiric* [71, 72] methods, while the rest of the system is treated with usual molecular-mechanics (*MM*), in *QM/MM MD* simulations [73]. The rare transition sampling problem, on the other hand, is tackled by various enhanced or accelerated sampling techniques, which may be grouped into four different categories: *coarse-graining*, *biasing*, *multi-scaling*, and *jumping* [74, 75]. *Coarse-graining* techniques, e.g. the *Martini force field* [76], represent individual groups of atoms by single beads, thereby reducing the number of degrees of freedom (*DOFs*) to investigate and thus speeding up the simulations substantially, while still preserving the relevant physics underlying the transitions of interest. On the downside, details and chemical specificity is lost [75, 77]. *Biasing* techniques, e.g. *Umbrella sampling* [78], *conformational flooding* [79], *steered MD simulations* [80], and others [81], add an artificial potential or force to the simulation, which is acting on a set of *DOFs*. In *metadynamics*, for example, a potential is added which penalizes already visited conformations of the system, thereby accelerating the sampling of rare transitions, while recovering the equilibrium distribution by re-weighting. On the downside, the true dynamics of the system is lost and the identification of a proper set of *DOFs* to bias is by no means trivial [75, 82]. *Multi-scaling* techniques

combine the best aspects of different models with individual precision level. Examples are the aforementioned *QM/MM MD* simulations or combinations of atomistic and coarse-grained *MD* simulations. Thereby, the simulations are sped up substantially, while an accurate description of the transitions of interest is maintained. On the downside, the model boundaries require substantial work to be handled properly [75]. The *replica-exchange method* is a representative for *jumping* techniques. Here, several simulations of the system (replicas) are performed at the same time under different conditions, e.g. different temperatures [83], protonation states [84], or others [85]. From time to time, swaps between replicas, performed under neighboring conditions, are tried out and accepted using individual weights, thereby reducing the transition barriers and/or biasing the simulation towards interesting regions. On the downside, the dynamical continuity is lost [75]. Obviously, the optimal choice of an enhanced or accelerated sampling technique, or a combination thereof, always depends on the problem to be investigated.

A different approach to sample transitions is the calculation of minimum energy pathways (*MEPs*) connecting two stable conformations, local potential energy minima, of the potential energy surface (*PES*) [86]. Along the *MEP* all potential energy maxima are representing first order saddle points, transition states [87], of the *PES*. The transition state with the highest potential energy provides the activation energy of the transition [88] and following the transition state theory [89, 90] also its rate. Over time, a wide range of techniques, e.g. the Nudged Elastic Band method [91], the Conjugate Peak Refinement method [92], the Ridge method [93], the DHS method [94], and the Dimer method [95], has been developed and successfully used to study various transitions, e.g. optimal proton transfer pathways [96, 97, 98, 99]. In contrast to the accelerated or enhanced sampling techniques introduced before, *MEP* calculations are, in principle, bias-free optimizations, relying (most often) only on the potential energy function and its first derivative. For complex transitions, however, *MEP* calculations provide an inappropriate picture, determining only the *MEP* closest to some initial guess pathway [100]. Furthermore, a single transition pathway is, in most cases, not sufficient to fully describe a transition [101]. These drawbacks led to the formulation of the Transition Network (*TN*) approach [102, 103], in which the conformation space is represented in a simplified and discrete manner [102]. Here, a complex transition, between an initial and final state, is described by plenty of interwoven sub-transitions, connecting various intermediate states, derived by discrete samplings of relevant *DOFs* [102, 103] or other methods [104, 105, 106], with each other and both end states. All transition pairs are close in

conformation space, thus eliminating the initial guess pathway problem of the *MEP* calculation. Thereby, a reliable potential energy barrier can be assigned to each sub-transition. Furthermore, multiple transition pathways, connecting the initial and final state, can be identified easily using simple graph theoretical algorithms [102, 103]. So far, the *TN* approach [102, 103] and other discrete path sampling techniques [107, 108] provided inside into various types of reactions, ranging from small proton transfer reactions [109], to conformational changes in peptides [110] and proteins [101] as well as the identification of complex reaction mechanisms [111].

Apart from all the benefits of a *TN* calculation a decisive difficulty remains, namely, the exponential increase of stationary points of the *PES* with increasing numbers of *DOFs* to sample [112]. This leads to numbers of *MEP* calculations, which are not processable within a finite amount of time. Therefore, an efficient *TN* calculation relies on a limited number of *DOFs* to sample. Deplorably, most processes of biological relevance represent the complete opposite of this requirement, including various highly connected *DOFs* on the most different time and length scales. Consequently, meaningful *TN* calculations need to include, in most cases, more *DOFs* than efficiently processable. One ansatz to attenuate this drawback, is the indirect inclusion of individual *DOFs*, meaning that several *TN* calculations are performed with the same feasible set of *DOFs*, while certain unsampled *DOFs* are altered in between the individual *TN* calculations. In other words, the set of unsampled *DOFs* is perturbed in between the individual *TN* calculations. Thereby, a comprehensive description of biological relevant processes would be possible at the cost of a new *TN* calculation for any *DOF* sampled indirectly. Within the framework of this thesis, this major drawback of the *TN* approach is addressed from two perspectives:

1. Combining the *TN* approach [102, 103] with *MD*-based samplings of specific *DOFs* which generate large numbers of states separated by relatively low transition barriers.

2. Using the information from a known, initial (or unperturbed) *TN* and an excessive, two-step coarse-graining procedure to improve the generation of an unknown, perturbed *TN*.

This leads to the development of the *TN-MD method* and the *TN prediction method*. Thereby, the efficient application of the *TN* approach for more complex reactions is enabled. Consequently, the optimized *TN* approach is used to elucidate the translocation of protons through one of the *proton transfer channels* of *CcO,* the

D-channel, to provide another piece of the structure-function analysis, so eagerly anticipated to cure *oxidative phosphorylation* related diseases.

The remaining thesis is organized as follows: Chapter 2 provides the biophysical and biochemical background necessary to understand the concept of proton transfer and its meticulous regulation in *CcO*. Therefore, the most important aspects of 200 and 40 years of research, for the proton transfer and *CcO*, respectively, are summarized and open questions, related to this work, are formulated. Chapter 3 provides the methodical background of this thesis by introducing the most important aspects of quantum mechanics necessary to understand the *semi-empiric* quantum method *OM2* applied throughout this thesis to handle quantum effects. Furthermore, different concepts and algorithms for the investigation of a *PES* are introduced. Finally, this chapter provides the graph theoretical framework, necessary to understand the concept of the *TN* approach and its optimization. Chapter 4 presents the refinement of the *TN-MD* and *TN prediction method* along with extensive tests in model systems for proton transfer reactions. In chapter 5 the optimized *TN-MD method* is applied to the proton transfer through the D-channel of *CcO* to identify ideal proton transfer pathways and sequences of events. Chapter 6 concludes this thesis and raises follow-up questions.

Chapter 2

Biophysical and Biochemical Background

The following chapter establishes the biophysical and biochemical background of this thesis, providing a detailed introduction of the transfer of protons in water or ice, and proteins, along with the necessary properties of the studied systems, as well as a detailed introduction of the membrane-embedded enzyme complex *CcO* from the respiratory chain. Due to the ever increasing importance of proton transfer reactions in biophysics and biochemistry, and the thus ever increasing number of results and theories for individual aspects of the proton transfer stemming from more than 200 years of research, a complete summary of all results is far beyond the scope of this thesis. Instead, only the most important aspects of the proton transfer are introduced and aspects under discussion are stated. The same holds for the 40 years of research concerning various aspects of prokaryotic and eukaryotic *CcO*.

2.1 Proton Transfer Reactions

Proton transfer reactions are part of almost all chemical reactions catalyzed by enzymes [113], underlining their importance in biophysics and biochemistry. Systems, in which proton transfer reactions readily occur are liquid water and solid ice. Hence, the first part of this subchapter provides an introduction of various properties of both these states of water, as well as an introduction of the proton transfer and proton-related structures in either of these states. Furthermore, the most prominent noncovalent interaction form [114], the hydrogen-bond, is introduced in this part. The second part of this subchapter focuses on the proton transfer in proteins, thus increasing the set of proton-donating and proton-accepting species to molecules

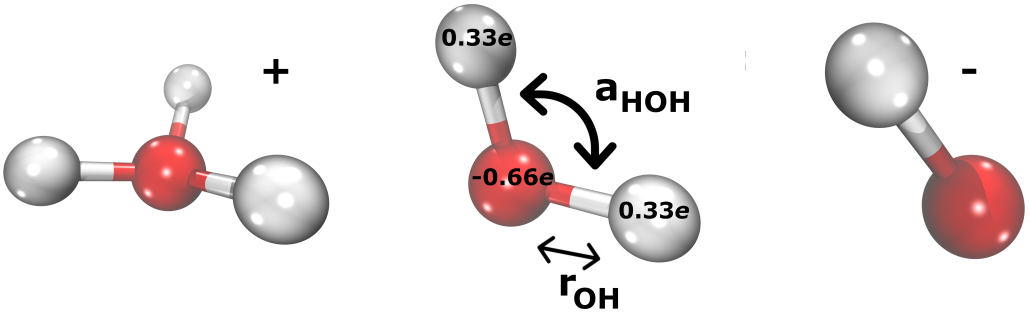


Figure 2.1: Left: Water molecule displaying an ionic defect (added proton) called hydronium ion. Middle: Water molecule with bond lengths r_{OH}, bond angle a_{HOH}, and partial atomic charges with e the elementary charge. Right: Water molecule displaying an ionic defect (missing proton) called hydroxide ion.

other than water, and introducing protein-specific properties.

2.1.1 Water, Ice, and Proton Transfer

Liquid water and solid ice represent different arrangements of individual water molecules with varying fluidity. Actually, the combination of hydrogen and oxygen in water represents a perfect match, combining the most and third most abundant elements of the universe [115]. The ground-state structure of a water molecule involves $O-H$ distances of $r_{OH} = 0.9572$ Å and an $H-O-H$ angle of $a_{HOH} = 104.52°$. Thus, the water molecule is bent with distinct centers of positive and negative partial charge, giving rise to a dipole moment of 1.855 D, causing its polar nature [116] (cf Fig 2.1).

In principle, the structure of a water molecule can be described by an isosceles triangle. This representation, however, is misleading and ignoring the two *lone electron pairs* of the oxygen. Instead, a water molecule should be described by a tetrahedron with the *lone electron pairs* and hydrogen atoms forming the four corners and the oxygen atom residing in the center. This tetrahedral structure allows the formation of a periodic, three-dimensional network, only adjoined by the corners of the tetrahedrons with unequal partners [117]. Such networks are called defect-free. Consequently, networks containing orientational defects, e.g. defects in which two oxygen atoms or two hydrogen atoms are adjoined (called L- and D-Bjerrum defects [118]) and other subspecies, or ionic defects forming charged ions, e.g. the hydronium ion, H_3O^+, or the hydroxide ion, OH^- (cf Fig 2.1), at individual lattice sites, violate the three-dimensional network structure [117].

In the different crystalline configurations of ice, all tetrahedrons, representing

individual water molecules, are adjoined to four other tetrahedrons, i.e. every corner of the tetrahedrons is adjoined, while the tetrahedrons of liquid water are only adjoined on average to four other tetrahedrons with controversially discussed fractions of water molecules adjoined by two, three, and four corners [119]. Hence, the three-dimensional network of liquid water is, due to random distortions, looser than the three-dimensional networks formed by the different crystalline configurations of ice.

The corners of the individual tetrahedrons are adjoined by hydrogen bonds, a non-covalent interaction form among specific atoms [117], schematically depicted for two water molecules (1 and 2) by

$$O^1 \cdots H^2 - O^2 \quad \text{with } \cdots \text{ a hydrogen and } - \text{ a covalent bond.} \tag{2.1}$$

In eq. 2.1 the hydrogen atom, H^2, is exposed to strong attractive forces from two oxygen atoms. These are the hydrogen bond donor, O^2, and the hydrogen bond acceptor, O^1. Therefore, the hydrogen atom could be seen as a bond connecting donor and acceptor [120]. With a strength of about 5 kcal/mol the hydrogen bond in eq. 2.1 is much weaker than the covalent bond. Thereby, it is possible to form and break hydrogen bonds in water at ambient conditions on the basis of thermal fluctuations alone [117].

A set of different geometrical criteria might be used to determine the existence of hydrogen bonds. These criteria [121] are:

1. The maximal distance between the oxygen atoms r_{OO}^{max}.
2. The maximal length of the hydrogen bond $r_{O\cdots H}^{max}$.
3. The minimal bond angle $a_{O\cdots H-O}^{min}$.

Apart from a geometrical definition, hydrogen bonds might be defined by an energetic criterion. Here, two water molecules are involved in a hydrogen bond if their pair interaction energy is lower than a maximal value E_{WW}^{max} [121]. Furthermore, combinations of geometric and energetic criteria might be used [122]. Reasonable values for r_{OO}^{max} and $r_{O\cdots H}^{max}$ are, for example, 3.2 Å and 2.4 Å, respectively [119]. Other criteria, e.g. involving the change of the electronic structure [123], also exist.

The ability of water molecules to form three-dimensional hydrogen bond networks turns a collection of rather simple molecules, and actually the smallest molecules with more than one atom [124], into a very anomalous substance, e.g. phase anomalies like an unusually high melting, boiling, or critical point, physical anomalies

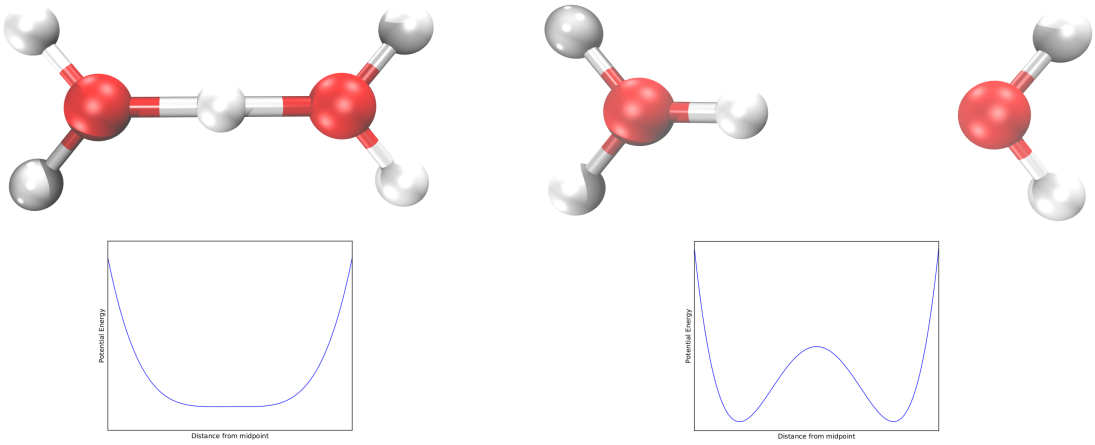


Figure 2.2: Left top and bottom: Two water molecules with optimal r_{OO} distance. The excess proton is shared by both water molecules, giving rise to a barrier-free potential. Right top and bottom: Two water molecules with r_{OO} larger than the optimal distance. The excess proton is located on either of the water molecules, as hydronium ion, giving rise to a double-well potential. Adapted from [117].

like an unusually high viscosity, thermodynamic anomalies like an unusually high specific heat capacity, and other anomalies like an increased density at the surface compared to the bulk, are observed [125]. For the most part, however, liquid water simply displays a typical behavior for the class of liquids it belongs to, i.e. liquids with three-dimensional hydrogen bond networks and/or liquids with tetrahedral shaped molecules [119]. Hence, the curiosity of water is not a single anomaly, but simply the shear mass of anomalies compared to other substances. The same holds for water in the form of solid ice [119].

In the previous paragraphs two basic properties of the hydrogen bond were introduced. These were: 1) the geometric property and 2) the energetic property. Next to these two basic properties, however, there is a third important property, namely, hydrogen bonds enable the transfer of protons [126]. Imagine for example a hydronium ion and a water molecule interacting in close proximity. In the energetically most favorable state the excess proton will be located directly in the middle of both water molecules with degenerated hydrogen bonds connecting the proton to either oxygen atom (cf. Fig 2.2, left). Such degenerated hydrogen bonds are also found in Ice X [127], the only ice not containing water molecules [119]. If the distance r_{OO} is constrained to the optimal value, a move of the proton out of equilibrium and towards either oxygen atom increases the energy. If, on the other hand, r_{OO} is increased and the proton is allowed to move on the axis formed by the oxygen atoms, the original potential energy minimum turns into a maximum (cf.

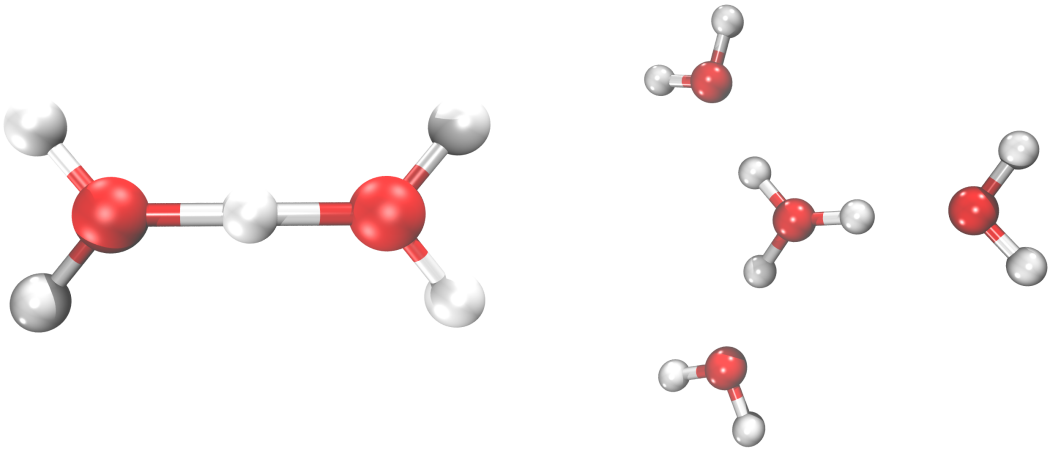


Figure 2.3: Left: Zundel ion containing two water molecules and a shared excess proton, Right: Eigen ion containing a centered hydronium ion connected to three water molecules by hydrogen bonds. Adapted from [117].

Fig 2.2, right). Consequently, in such situations the excess proton is preferentially not shared between the water molecules, but forms a hydronium ion with either of them. From Fig 2.2, it can be deduced that the proton transfer barrier, pronounced on the right and absent on the left side, is increasing with increasing r_{OO}. On the other hand, the proton transfer barrier is decreasing, and ultimately vanishing, with decreasing r_{OO}. Hence, barrier-free proton transitions are possible in liquid water and readily occur for small r_{OO}. Due to the very small (or absent) proton transfer barriers for small r_{OO}, there is also no barrier that could be tunneled by the proton [128]. These findings render the tunneling of protons in liquid water negligible. In certain configuration of ice, on the other hand, proton tunneling readily occurs [117].

But how is an excess proton stored in a hydrogen bond network and how is the actual proton transfer step taking place? Already in 1899 [129], it was suggested that hydronium ions, representing distorted tetrahedrons, should be compatible with the structure of hydrogen bond networks. More than sixty years later, Zundel [130] and Eigen [131, 132] identified the preferred solvation structures of hydrated excess protons, namely the Zundel, $[H_2O \cdots H \cdots OH_2]^+$, and Eigen, $H_3O^+ \cdot (H_2O)_3$, ion (cf. Fig. 2.3).

In the Zundel ion, the excess proton is equally shared between two molecules of water, connected by degenerated hydrogen bonds. The Eigen ion, on the other hand, contains three water molecules distributed around a centered hydronium ion, such that all hydrogen atoms of the hydronium ion are involved in hydrogen bonds with

the oxygen atoms of the water molecules. Both of these ions are closely related, i.e. small shifts in the donor-acceptor distances readily convert one ion into the other [133, 134]. Furthermore, *ab initio* simulations suggested that both ion types are required to describe the proton transfer process [135, 136]. The actual proton transfer step was envisioned by Grotthus in 1806 [137], even without a proper atomistic knowledge of the water molecule. Here, an excess proton is translocated from one water molecule to the other by the exchange of covalent and hydrogen bonds. Such a shift, or Grotthus diffusion, occurs spontaneously at 300 K [117]. In ice the proton mobility is about one third of the proton mobility in water [138].

Putting all these information together, one arrives at the commonly accepted proton transfer mechanism. In this mechanism the excess proton is initially stored in an Eigen ion. Due to thermal fluctuations around the Eigen ion, a hydrogen bond connecting the first and second solvation shell is broken, reducing the number of hydrogen bonded partners of a water molecule in the first solvation shell from four to three. This water molecule is now able to accept the excess proton, following a Grotthus diffusion step with the Zundel ion as intermediate. Finally, the excess proton is stored in an Eigen ion centered around the water molecule which accepted the proton [117]. The rate-limiting step of this mechanism is the re-orientation of the water molecules, giving rise to proton transfer timescales of one picosecond [135].

The mechanism of proton hole transfer, i.e. the transfer of hydroxide ions, is directly emerging from the proton transfer mechanism, when considering similar preferred solvation structures and a similar hydrogen bond breaking between the first and second solvation shell. Here, the preferred solvated structures are a Zundel ion analogue, $[HO \cdots H \cdots OH]^-$, and an Eigen ion analogue, $OH^- \cdot (H_2O)_3$. Hence, the proton hole is stored in an Eigen ion analogue and undergoes a Grotthus diffusion step, with a Zundel ion analogue as intermediate, upon the breakage of a hydrogen bond connecting the first and second solvation shell due to thermal fluctuations. At the end the proton hole is, once again, stored in an Eigen ion analogue centered around the water molecule which accepted the proton hole. In contrast to the transfer of an excess proton, however, the mechanism of the proton hole transfer is still discussed controversially [139, 140], and mechanisms other than the mirror image mechanism are proposed and backed up by experiments like neutron scattering [141, 142, 117].

2.1.2 Proton Transfer in Proteins

Hydrogen bonds, as introduced in the previous subchapter, are not limited to liquid water and solid ice, e.g. the structure and stability of two of the most important types of molecules in life, the *DNA* and proteins, depends to great extent on hydrogen bonds [143, 144]. Furthermore, the hydrogen bond is not limited to the case presented in eq. 2.1. Other types of hydrogen bonds are [145]:

$$N \cdots H - O \quad O \cdots H - N \quad N \cdots H - N. \tag{2.2}$$

The hydrogen bonds in eqs. 2.1 - 2.2 represent a collection of hydrogen bonds predominantly occurring in biological macro-molecules. Other, less abundant, hydrogen bonds are formed with sulfur as hydrogen bond acceptor [146]. Hydrogen bonds with $C-H$ groups as hydrogen bond donor are much weaker [145]. However, in most proteins the ratio of hydrogen bonds from eqs. 2.1 - 2.2 to hydrogen bonds containing $C-H$ groups as hydrogen bond donor is 1:1.4 [147], underlining their importance by their ubiquitous presence. Consequently, the transfer of protons, following the Grotthus mechanism, is not limited to liquid water and ice.

The proton transfer reactions in biological systems, e.g. enzymes, generally fall into two different categories. These are:

1. Catalytic proton transfer reactions

2. Proton translocations

The former category describes processes of chemical conversion, while the latter category describes the proton transfer over long distances following permanent or transient proton transfer pathways, e.g. along *proton transfer channels* [148]. The remainder of this subchapter focuses on the latter category.

Onsager [149], Nagle, and Morrowitz [150] coined the term *hydrogen bond chain* (*HBC*) to describe the transfer of protons over long distances involving a series of hydrogen bond donors and acceptors. These *HBCs* contain individual water molecules, side chains of various amino acids (the buildings blocks of a protein), or a combination of both of them. Amino acid residues suggested to participate in *HBCs* are: serine, threonine, tyrosine, glutamic acid, aspartic acid, glutamine, asparagine, lysine, arginine, and histidine [151]. The transfer of protons along the *HBC* is believed to occur via the "hop-turn" mechanism [150].

In the "hop-turn" mechanism a proton enters the *HBC* from one side and, following a series of subsequent proton hops or a concerted [152] (collective) proton

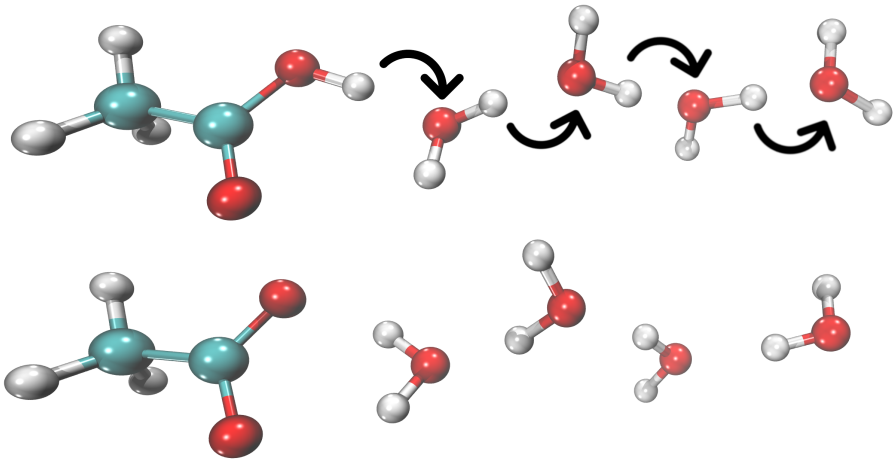


Figure 2.4: Top: HBC formed by an acetate molecule and four water molecules. Black arrows indicate the proton hops. Bottom: Reversed HBC as a result of the proton transfer shown in the top. A second proton transfer, from left to right, is only possible after a turn from all water molecules.

hop, a different proton leaves the *HBC* on the other side (cf. Fig. 2.4 up). Due to the proton transfer, the *HBC* is re-orientated, i.e. the transfer of a second proton in the same direction is not possible until the *HBC* is turned in its initial configuration (cf. Fig. 2.4 down). Interestingly, the net charge of a proton is only transferred after the hop *and* turn step [153].

One of the simplest proton transfer channels is provided by *Gramicidin*, a water-filled, cation-selective pore containing only 15 amino acids, which is readily incorporated into membranes and thus controlling the cation concentration on either side of the membrane. Electro-osmosis and streaming potential measurements allowed the determination of the water and cation flux within the channel, respectively. For the transfer of protons no water flux was measured, indicating a *Grotthus*-like proton transfer. This absence of water molecule displacements explains the exceptionally high conductivity of protons in *Gramicidin* compared to the conductivity of other cations and also the conductivity of protons in bulk water. According to various MD simulations, the water molecules inside the channel are oriented as an *HBC* with few defects due to interactions with the channel walls, suggesting a semi-collective, i.e. partially concerted and partially step-wise, proton transfer [152]. Hence, a water-filled pore is sufficient to enable proton translocations.

On the other hand, investigations of the water-filled, water-conducting pore *Aquaporin* showed that individual modifications of such pores prevent proton translocations almost completely. In *Aquaporin* the conduction of protons is prevented by three factors. These are [152]:

1. The hydrophobic channel walls.
2. Two asparagine residues in the middle of the pore disrupting the *HBC*.
3. The positive charge of the asparagine residues repelling the protons.

The proton translocations in *Gramicidin* and *Aquaporin* are limiting cases. The bulk of biologically relevant processes, however, require a middle course, i.e. regulated proton translocations. Three means of regulation may be envisioned. These are:

1. Steric regulations (or steric gates), e.g. due to amino acid side chains disrupting the *HBCs.*
2. Hydration regulations (or hydration gates), e.g. due to partially dehydrated channel regions.
3. Voltage regulations (or voltage gates), e.g due to pH changes.

The *M2 Viral Proton Channel* is a representative of the first group. Here, four histidine residues occlude the channel and thus no continuous *HBC* is formed and no proton transfer is observed. Upon the protonation of two of the histidine residues, however, the histidine gate opens up (due to electrostatic repulsion), thereby allowing the formation of a continuous *HBC* and thus proton transfer. *Bacteriorhodopsin* is another representative of this group. Here, however, the proton transfer is initialized by a light-induced steric change of an amino acid side chain. Interestingly, it seems to be evident that *Bacteriorhodopsin* is never displaying a continuous proton transfer pathway [152]. MD simulations of *Complex I* suggest a proton transfer regulation according to group two, i.e. regions of individual *proton transfer channels* are less hydrated, rendering the formation of continuous *HBCs* unlikely [154]. Quite amazingly, however, protons are able to increase the hydration of *proton transfer channels*, thereby building their own temporary *HBCs* [155]. CcO comprises regulatory features of groups one and two [156]. A detailed introduction of this enzyme complex is provided in the next subchapter. *Complex V* and the closely related *Flagellar Motor,* comprise features of group three. For more information on voltage-gated proton transfer channels the interested reader is referred to [152].

2.2 Cytochrome c Oxidase

Molecular oxygen provides a maximal amount of energy, when oxidizing hydrocarbons. Based on this property, a long list of aerobic prokaryotes and eukaryotes developed along the evolution utilizing this abundant molecule. Overall, more than 90 %

of the molecular oxygen consumed by these organisms is catalyzed by membrane-embedded *heme-copper oxidases* (*HCOs*) [157], comprising such illustrious names like *CcO* or *quinol oxidase*. In general, the large family of *HCOs* is divided in three different types, A-, B-, and C-type, which vary in terms of structure and appearance. Both enzymes mentioned before are A-type *HCOs*, which limits their appearance to eukaryotes and some bacteria. *HCOs* of B- and C-type, on the other hand, are exclusive to bacteria and archea. All three types of *HCOs*, however, fulfill the same function, namely, generating a proton-motive force Δp across the membrane they are embedded in, i.e. the inner mitochondrial membrane in eukaryotes or the cell membrane in prokaryotes [158]. Thereby, Δp is defined as

$$\Delta p = F\Delta\Psi + RTln(\frac{[H^+_{in}]}{[H^+_{out}]}) = F\Delta\Psi - ln(10)RT\Delta pH, \tag{2.3}$$

where the former and latter term describe the electrical and chemical contribution, respectively, F represents the Faraday constant, $\Delta\Psi$ represents the electrical potential across the membrane, R represents the gas constant, T represents the temperature, and ΔpH represents the pH difference across the membrane [159]. The generation of Δp is achieved in two steps:

1. The uptake of electrons and protons from the positively-charged (P) and negatively-charged (N) side of the membrane.

2. The pumping of one proton per consumed electron from the N- to the P-side.

Hence, the overall proton translocation ratio of *HCOs* across the membrane is, at least in theory, two protons per electron, due to the fact that step one is thermodynamically equal to a proton transfer across the membrane. A-type *HCOs* are most efficient, almost reaching the theoretical value even for high opposing Δp's, while *HCOs* of the B- and C-type are much more inefficient [158].

A cylindrical-shaped subunit (SU) I, containing twelve slightly tilted transmembrane helices organized in three clusters, is highly conserved in all *HCOs*. Two of these clusters are ligated by histidines to three redox-active sites, namely a low spin heme and a high spin heme plus Cu_B. Furthermore, SU I contains three pores, lined by hydrophilic residues, between the clusters. The proton transfer function of two of these pores, namely the D- and K-channel, is known, while the function of the third pore, the H-channel, is discussed prominently. A second highly conserved structure is SU II. It exists in A-type *HCOs* as well as some B- and C-type *HCOs*. SU II comprises two transmembrane helices and a β-sheet, which is ligated to the

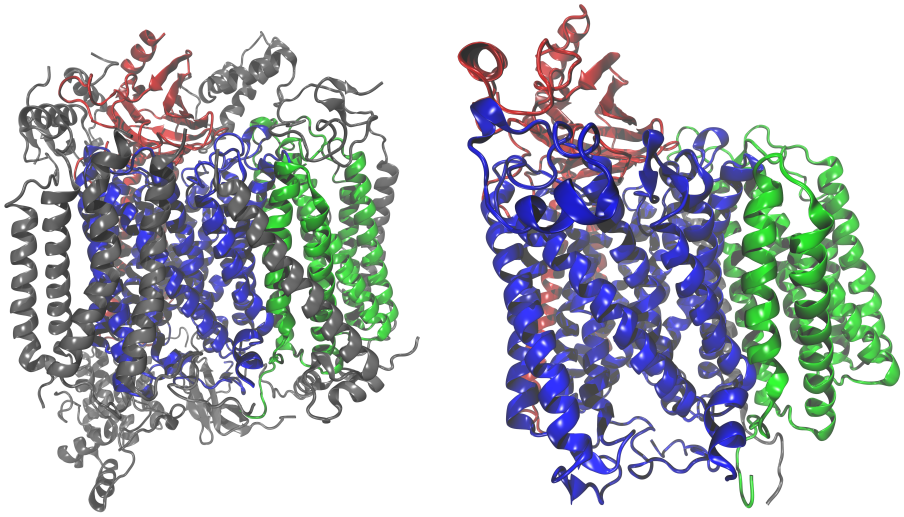


Figure 2.5: Left: Structure of mitochondrial CcO from heart cells of *Bos taurus* (PDB entry: 1OCC [161]). SU I (blue), SU II (red), SU III (green), accessory SUs IV-XIII (gray). Right: Structure of bacterial CcO from *Rhodobacter sphaeroides* (PDB entry: 1M56 [160]). SU I (blue), SU II (red), SU III (green), and accessory SU IV (gray).

redox-active site Cu_A. A third highly conserved structure is SU III. It contains seven transmembrane helices and is only found in A-type *HCOs*. It does not contain any redox-active site nor is it involved in the transfer of protons. Its main function seems to be the stabilization of the *HCO* complexes. Furthermore, it is involved in the channeling of the molecular oxygen [160, 158].

The remainder of this subchapter focuses on the A-type *HCO*, *CcO*, containing all three SUs mentioned above plus some accessory SUs whose function is, until now, poorly understood (cf. Fig. 2.5).

To generate the proton-motive force it catalyzes the following reaction:

$$4e_P^- + 4H_{C,N}^+ + 4H_{T,N}^+ + O_2 \rightarrow 4H_{T,P}^+ + 2H_2O, \tag{2.4}$$

where C and T distinguish chemical protons used for the reduction of molecular oxygen to water and translocated protons, respectively. Thus eight electrical charges are translocated per oxygen molecule. The following subchapters will introduce the most important aspects of the *CcO* complex, starting with the active site, i.e. the site of the oxygen reduction, and the catalytic cycle by which eq. 2.4 is facilitated. Subsequently, the oxygen uptake pathway, water release pathway, and electron uptake pathway are briefly presented, while the proton uptake pathways, the D-, K-, and H-channel, are introduced in detail. Finally, the most entertained principle of the proton-coupled electron transfer is introduced along with different regulatory

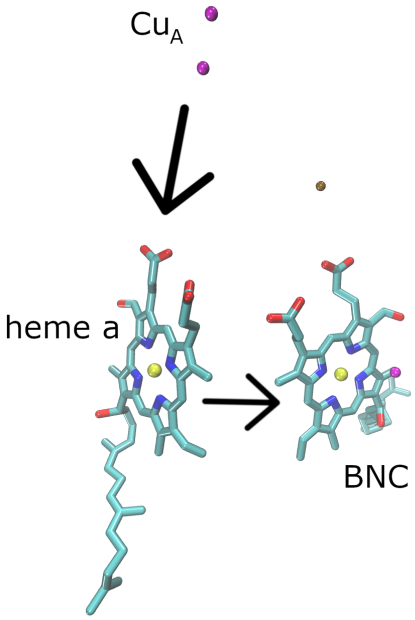


Figure 2.6: The redox-active sites of *CcO*: Cu_A, low spin *heme a*, and the *BNC* containing high spin *heme* a_3 and Cu_B. *Cu* atoms depicted in magenta, *Fe* atoms depicted in yellow, and an Mg^{2+} ion depicted in brown. Black arrows indicate the pathway of the electrons.

aspects.

2.2.1 Active Site and Catalytic Cycle

The active site of *CcO*, also called the *binuclear center* (*BNC*), catalyzes the reduction of molecular oxygen to water. It is located in SU I and contains the high spin *heme* a_3 and Cu_B in close proximity, i.e. about 5 Å in between the *Fe* and *Cu* atom (cf. Fig. 2.6). As described before the redox-active sites are linked to SU I by histidines, one for *heme* a_3 and three for Cu_B. Furthermore, the proprionate groups of *heme* a_3 interact electrostatically with conserved amino acid residues of SU I, which are important for the pumping of protons [162, 163]. Interestingly, one of the histidine residues Cu_B is ligated to, is linked to a tyrosine, forming a unique histidine-tyrosine ligand, which is highly conserved among all A-type *HCOs* [160, 158].

The reaction catalyzed by *CcO* (cf. eq. 2.4) does not occur at once, but step wise in a catalytic cycle (cf. Fig. 2.7), containing (until now) six or seven individual configurations of the *BNC*, depending on the redox-state of the low spin *heme a*, the second heme located in SU I, and Cu_A, located in SU II. Here, the individual

configurations of the *BNC* along the catalytic cycle, including the conserved tyrosine residue, are discussed [158, 164].

1. R-state

   The R-state (or reduced state) contains a five-coordinated, *ferrous* (Fe^{2+} or Fe^{II}) *heme* a_3, a *cuprous* Cu_B^+ (Cu_B^I), and the tyrosine (*tyr-OH*). This *ferrous/cuprous* configuration is the only configuration of the catalytic cycle, in which molecular oxygen will be bound. This limitation is assigned to a fourth ligand of *cupric* Cu_B^{2+} (Cu_B^{II}), which prevents the binding of the molecular oxygen [165].

2. A-state

   The A-state is reached from the R-state, once the molecular oxygen is bound "end-on" [166] to *heme* a_3, usually described as *ferrous-oxy* (Fe^{II}-O_2) *heme* a_3. Interestingly, the binding of molecular oxygen is rather reversible, which contrasts *CcO*'s high oxygen affinity. This contradiction is resolved, when considering fast electron transfers, trapping the bound oxygen [167].

3. P-states

   The P-states of the catalytic cycle were originally named after the *ferric-cupric-peroxide* structure involved [168], which turned out to be a *ferryl-heme cupric* structure with scission of the O-O bond [166]. Two of these states, with identical absorption spectra, were identified, namely, the P_M-state and the P_R-state. The P_M-state is formed spontaneously from the A-state, if the low spin *heme a* and Cu_A are oxidized. Here, the electrons required to split the O-O bond come from *heme* a_3, Cu_B, and *tyr-OH*. Thereby, *heme* a_3 is turned from *ferrous* to *ferryl* (Fe^{IV}), Cu_B is turned from *cupric* to *cuprous*, and *tyr-OH* is turned into a neutral radical (*tyr-O**). As stated before, *cupric* Cu_B^{II} gets a fourth ligand (OH^-), with the oxygen atom stemming from the molecular oxygen and the proton stemming from *tyr-OH*. Hence, the overall structure of the P_M-state is Fe^{IV}=O^{2-}, Cu^{II}-OH^-, and *tyr-O** [169]. The P_R-state, on the other hand, is formed from the A-state, if the low spin *heme a* and Cu_A are reduced. In principle, the electrons required to split the O-O bond come from the same residues as before, except for the one coming from *tyr-OH*. This fourth electron comes now from the low spin *heme a*. Still, *tyr-OH* provides its proton, thus turning into *tyr-O*$^-$. Hence, the overall structure of the P_R-state is the same as for the P_M-state, only exchanging *tyr-O** by *tyr-O*$^-$ [169].

Interestingly, the P_R-state is the only stable state of the *BNC*, which is not neutral, displaying a charge of -1. This charge allows the first proton to be transferred to the *proton-loading site* (*PLS*), i.e. the site at which a pumped proton transiently resides before being expelled to the P-site [170, 171]. The P_R-state is reached from the P_M-state by electron transfer [158, 164].

4. F-state

   The F-state is reached following the protonation of the P_R-state. Thereby, the OH^- ligand of Cu_B is likely protonated to water and thus leaving the *BNC*. Hence, the overall structure of the *BNC* in the F-state is $Fe^{|V}{=}O^{2-}$, $Cu^{||}$, and *tyr*-O^- [172].

5. O-states

   The O_H-state (or oxidized state) is reached upon reduction and protonation of the F-state, turning *heme* a_3 from *ferryl* into *ferric* ($Fe^{|||}$), and leaving *cupric* Cu_B and *tyr*-O^- unchanged. Furthermore, a strong hydrogen bond is suggested to connect the *ferric* OH^- ligand and Cu_B [173]. Recent *DFT* simulations, however, suggest an O_H-state as equilibrium of the state mentioned before and a state with *cuprous* Cu_B and *tyr*-O^* [174, 175, 176]. Another oxidized state, the O-state, is reached by relaxation from the O_H-state. This state, however, is not displaying any catalytic activity [158, 164].

6. E-state

   The E-state is reached upon reduction and protonation of the O_H-state, turning *cupric* Cu_B into *cuprous* Cu_B and *tyr*-O^- into *tyr*-*OH*. Finally, the reduction and protonation of the E-state yields the R-state again [158, 164].

Proton and electron transfer steps are observed, in quite similar fashion, along the A $\rightarrow$ F, F $\rightarrow$ O_H, O_H $\rightarrow$ E, and E $\rightarrow$ R transition.Overall, the intermediate states R, A, P_M, P_R, and F are well defined, while the intermediate states O_H and E require further investigations [158, 164].

2.2.2 Oxygen, Water and Electron Pathways

The oxygen reduction catalyzed by *CcO* at the *BNC*, involves three reacting species, namely, molecular oxygen, protons, and electrons, and a single product species, namely, water. These four compounds differ substantially in their properties, rendering different pathways to and from the *BNC* a necessity [158, 164]. According

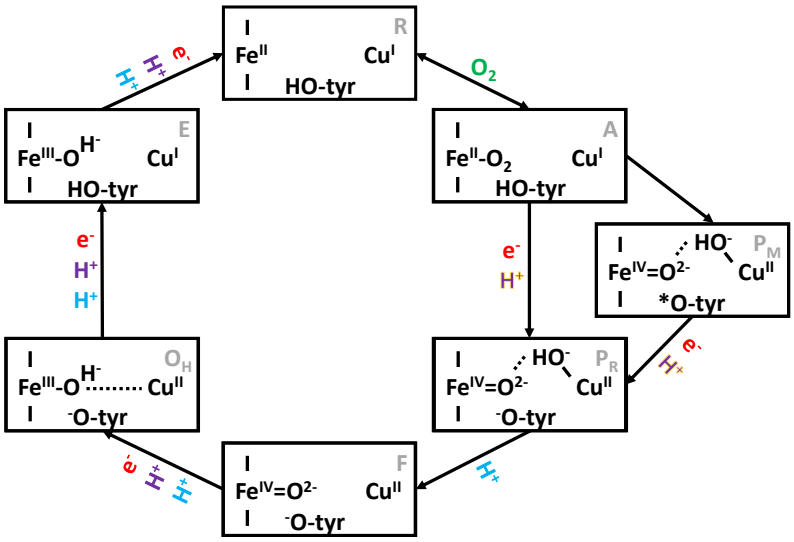


Figure 2.7: Catalytic Cycle of *CcO* with seven individual states, R, A, P_M, P_R, F, O_H, and E (gray). The rectangular boxes depict the conformation of the *BNC*, more precisely the *Fe* atom of high spin *heme* a_3, Cu_B, and the tyrosine. Arrows indicate directed transitions between individual states, associated with oxygen uptake (green), electron uptake (red), proton uptake to the *PLS* (purple) with subsequent expulsion to the P-side, and proton uptake to the *BNC* (blue). The proton uptake to the *PLS* in the transitions A $\rightarrow$ P_R and P_M $\rightarrow$ P_R is highlighted (see text). Adapted from [164].

to eq. 2.4, however, a second product, namely, translocated protons, exists, rendering a proton exit pathway, connecting the *PLS* and the P-side, a further necessity. So far, however, this pathway is not identified. The oxygen, electron, and water pathways are briefly introduced in the following, while the three putative proton transfer channels, the D-, K-, and H-channel, are introduced in detail in subsequent subchapters.

- Oxygen Uptake Pathway

  Biological membranes, containing bilayers of *phospholipids*, are hydrophilic at the borders and hydrophobic at the center. Hence, the hydrophobic molecular oxygen accumulates preferably in the middle of the membranes [177, 178]. In SU III of A-type *HCOs* several transmembrane helices surround a hydrophobic channel, connecting the inner of the membranes with the active site of the *HCOs*. Furthermore, several *phospholipids* are bound to SU III in this region [179]. This channel was identified by various techniques as oxygen uptake pathway to the *BNC* [180, 181, 158].

- Water Release Pathway

Following eq. 2.4 two water molecules are produced per molecular oxygen molecule reduced. Spectroscopic experiments [182], as well as simulations [183], suggest a rapid water diffusion from the *BNC* towards an Mg^{2+} ion located above the *BNC* (cf. Fig 2.6), indicating a water release pathway. Such a hydration requires, obviously, a proper regulation to avoid a decoupling of the proton pump activity from the oxidase activity. A possible form of regulation is for example the redox-dependent dynamics of the water molecules [158]. Simulations, furthermore, suggested a redox dependent split of a *propionate-arginine* ion pair [183, 158].

- Electron Uptake Pathway

  CcO receives its electrons, individually, from reduced *cytochrome c* on the P-side of the membrane [184]. The first redox-active site to accept the electrons is Cu_A (cf. Fig 2.6). From Cu_A the electrons are transferred to the low-spin *heme a* (≈12 Å), because its the closest of all other redox-active sites. Finally, the electron is transferred from *heme a* to the *BNC* (≈7 Å) [158]. All steps involve pure electron tunneling.

2.2.3 D-Channel Proton Pathway

Protons require, other than electrons, a medium for their transport over long distances on biologically relevant timescales, e.g. water-filled *proton transfer channels* to facilitate a Grotthus diffusion.

One of these channels in *CcO* is the D-channel, a ≈25 Å long cavity [156], located in SU I, connecting a highly conserved aspartate (D132, numbering according to *CcO* from *Rhodobacter Sphaeroides*), presumably acting as a proton antenna [185], on the N-side with a highly conserved glutamate (E286) in the middle of the *CcO* complex, ≈12 Å from the *BNC* [156]. It is lined by several hydrophilic amino acids, i.e. aspartate (D132), histidine (H26), asparagines (N121, N139, and N126), tyrosine (Y33), serines (S142, S197, S200, and S201), and glutamate (E286), and nine to twelve water molecules according to individual crystal structures [161, 186] (up to sixteen according to computational studies [187]) and is responsible for the translocation of six or seven protons per oxygen molecule. In detail, all four pumped protons and two or three chemical protons are believed to be translocated through the D-channel [158, 164].

The first quarter of the D-channel is rather narrow, containing two or three water molecules with more or less fixed positions, if the side chain of D132 points

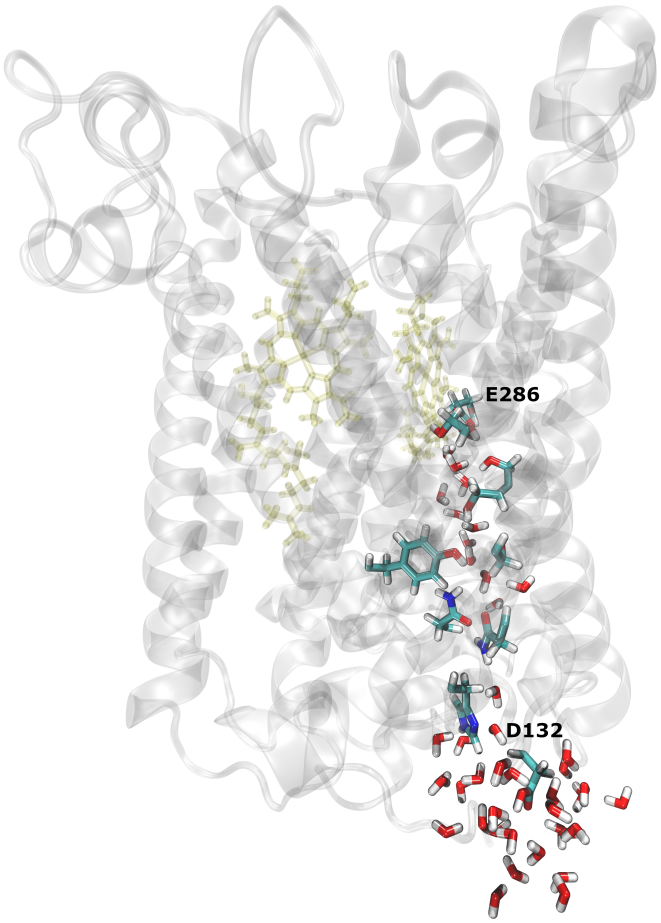


Figure 2.8: D-channel in SU 1 (gray) of *CcO* from *R. sphaeroides.* Entrance and exit labeled by D132 and E286, respectively. Unlabeled, residues along the D-channel are H26, N121, N139, N126, Y33, S142, S197, S200, S201, and all water molecules within 5 Å to these residues. Low spin *heme a* and high spin *heme* a_3 are depicted in yellow.

in- or outside the channel, respectively. In the latter case, the third water molecule occupies the position of the side chain of D132. The narrowest part of the D-channel is reached at the highly conserved N139, whose side chain is pointing directly inside the D-channel axis [161, 186, 188, 189]. Following N139, the D-channel gets wider, offering a highly connected and highly intermixing network of water molecules until E286, at the end of the D-channel, is reached [156].

Due to the side chain of N139, no continuous *HBC*, connecting the entrance and exit of the D-channel, can be formed. Hence, the side chain of N139 is representing a steric gate for the proton transfer through the D-channel. Computational studies of the D-channel, however, suggest the existence of a metastable open conformation of N139 (≈4 kcal/mol higher in free energy). Here, the side chain of N139 is rotated out of the D-channel axis, thereby allowing, in principle, the formation of

a continuous *HBC*. In numbers: closed conformation of N139, (χ_1, χ_2)=$(-15°, 41°)$, and open conformation of N139, (χ_1, χ_2)=$(-75°, -70°)$. Still, no continuous *HBC*, connecting the entrance and exit of the D-channel, has been observed in simulations, furthermore suggesting a hydration gate surrounding N139 and the opposing N121 [156].

Mutational experiments underlined the importance of N139, revealing a decoupling of the proton pump and oxidase activity, if N139 is mutated to various other amino acids [190, 191, 192, 193, 194]. *MD* simulations in which N139 is replaced by an alanine (N139A), for example, show a stable, continuous *HBC* along the entire D-channel, possibly allowing an uncontrolled proton transfer through the D-channel, leading to a decoupling of the proton pump and oxidase activity [156]. Two further mutations decoupling the proton pump and oxidase activity are N139C and N139T. Computational studies suggest a decoupling of the proton pump and oxidase activity due to an increased backflow of protons. Therefore, it seems that N139 is not just regulating the proton transfer from D132 to E286, but also the transfer from E286 to D132 [195]. Other mutations, e.g. N139L and N139V, abolish the transfer of protons through the D-channel completely. These mutants display neither oxidase activity nor proton pumping [156, 195]. All mutations mentioned so far were uncharged. An example for a charged mutation would be N139D which also decouples the proton pump and oxidase activity. Here, the additional charge is likely to disrupt the accurate regulation of the proton transfer at E286. The same assumption holds for other charged mutations [192, 196]. A final mutation of interest is D132N. Here, no uptake of protons from the N-side to the inside of the D-channel is possible. Still, a proton transfer along the D-channel has been observed by electrometry [197]. This proton transfer was assigned to a deprotonation of Y33 located shortly above N139. The authors of this study furthermore suggest, that the depronotation of Y33 is not limited to D132N mutants, but also occurring in the wildtype *CcO* complex. Here, however, the re-protonation of Y33 from D132 is too fast to be measured [197].

E286 represents the end of the D-channel. It is essential, although surprisingly not indispensable [198], in shuttling the protons to the *BNC* or *PLS* [199], respectively, especially when considering the absence of any water molecules in the hydrophobic cavity above E286. The valve-like function of E286, corresponding to an up and down conformation of its side chain, and the functional role of dynamic water molecules in the hydrophobic cavity above E286 are discussed prominently [200, 201, 202, 195].

Over time, a wide range of computational studies focused on the proton transfer

through the D-channel, either from N139 to E286 [203, 204] or from the entrance to the exit [205, 206], including several mutants [205, 195]. The latest investigations, concerning the transfer of protons in the $A \rightarrow F$ transition, revealed transition barriers of ≈9 kcal/mol and corresponding transition states in which the excess protons are located in between N139 and N121 [206, 195]. Furthermore, the investigations proposed that the excess protons are responsible for the opening and closing of N139 and that the excess protons are capable of building their own *HBCs*, while being transferred [206, 195]. The latter behavior has also been observed before for the transfer of protons through carbon nanotubes [155].

2.2.4 K-Channel Proton Pathway

The D-channel is responsible for the translocation of six or seven protons per oxygen molecule. This includes all translocated protons. Hence, one or two further protons, need to be translocated to the *BNC* to complete the oxygen reduction. The K-channel, connecting E101 in SU II with Y288 in SU I close to the *BNC* [207, 208], is believed to supply the remaining protons in the reductive phase of the *BNC* [158, 164]. Other than the D-channel, the K-channel is much less hydrated and sparse in hydrophilic amino acids [160, 209]. A notable representative of the hydrophilic amino acid residues along the K-channel is K362, which is highly conserved in all A-type *HCOs* and suggested to participate in the transfer of protons along the K-channel by redox-state dependent rotations [210] of its side chain [211, 212]. Furthermore, the mutation of K362 to any hydrophobic amino acid residue produces mutants which display neither oxidase activity nor proton pumping [213].

More general questions, however, are still unanswered, namely, why is the K-channel required next to the D-channel, why is the K-channel only responsible for a few and not all proton translocations to the *BNC*, and how does the K-channel interact with the D-channel [158, 164].

2.2.5 H-Channel Proton Pathway

The H-channel is located in SU I. It is separated from the D- and K-channel, spanning the entire membrane, connecting D407 (numbering according to *CcO* from *Bos taurus*) on the N-side and D51 on the P-side. In doing so, the H-channel leads to and passes the low-spin *heme a*. Consequently, the distance between the H-channel and the *BNC* is rather large. Along the H-channel, numerous hydrophilic residues are located, i.e. D407, the partially conserved H431, S461, S458, T424, S454, R38,

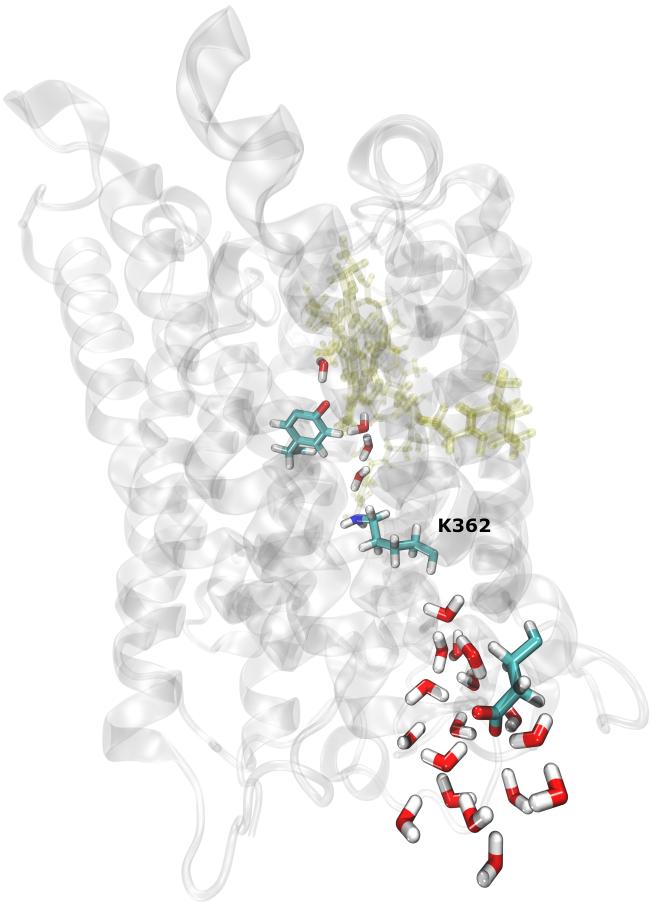


Figure 2.9: K-channel in SU 1 (gray) of *CcO* from *R. sphaeroides.* Highly conserved residue K362 is labeled. Unlabeled, residues along the K-channel are E101 (in SU II), Y288, and all water molecules within 5 Å to these residues. Low spin *heme a* and high spin *heme* a_3 are depicted in yellow.

Y371, Y54, Y440, S441, D51, and S205 in SU II (cf. Fig. 2.10). Computational studies, furthermore, suggest the presence of 20 to 22 water molecules [214]. Overall, the H-channel may be divided into three distinct regions. These are [214]:

1. The upper region, comprising Y54, Y371, Y440, S441, D51, and S205, contains most of the water molecules. Deprotonation of D51 and subsequent hydrogen bond breaking leads to further hydration of the upper region.

2. The middle region, comprising S458, S454, R38, and T424, is located next to the low spin *heme a*. It contains four water molecules, which is sufficient to translocate protons.

3. The lower region, comprising D407, H413, and S461, is less hydrated, which leads to connectivity gaps in the *HBC*. The number of water molecules in the lower region is significantly increased for protonated H413. The calculated pk_a

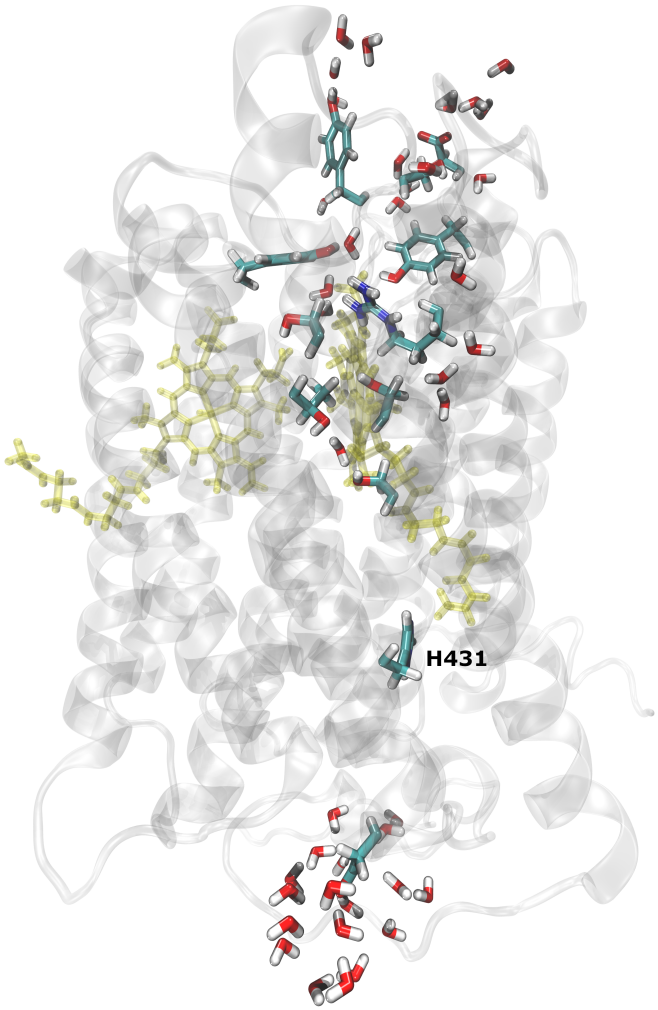


Figure 2.10: H-channel in SU I (gray) of *CcO* from *B. taurus*. The partially conserved residue H431 is labeled. Unlabeled, residues along the H-channel are D407, H431, S461, S458, T424, S454, R38, Y371, Y54, Y440, S441, D51, S205 (in SU II), and all water molecules within 5 Å to these residues. Low spin *heme a* and high spin *heme* a_3 are depicted in yellow.

value of H413, however, suggests a deprotonated H413.

The H-channel, as sole the pathway for all four pumped protons in mammalian, mitochondrial *CcO*, is emphasized by *Yoshikawa et al* [215, 188, 216, 217, 160]. Despite the limitation to mammalian, mitochondrial *CcO*, most structural elements of the H-channel are shared by bacterial A-type *HCOs* [158]. However, while mutations along the H-channel abolish the proton pump activity of mammalian, mitochondrial *CcO* [215], no such effects have been observed for bacteria [218, 219], and even the mitochondrial *CcO* from yeast, which displays a clear H-channel [220, 221, 222]. Overall, a possible proton transfer through the H-channel in mammalian, mitochondrial *CcO* requires further investigations [214, 158]. Instead of being a proton

transfer channel, the H-channel could also serve as a dielectric well [223, 214].

Putative proton translocations through the H-channel also require a different coupling to the oxidase activity. In the following subchapter, the generally excepted proton translocation mechanism, the *directly coupled proton pump* employing the D- and K-channel [224], is introduced. For information on modifications and alterations necessary for the proton translocation mechanism proposed by *Yoshikawa et al* the interested reader is referred to the literature [217, 225, 226].

2.2.6 Proton-Coupled Electron Transfer

For *CcO*, and any other *HCO*, to function as a proton pump, the internal electron and proton transfer steps, along the pathways introduced before, need to be coupled meticulously. The most entertained coupling mechanism for *CcO* is the *directly coupled proton pump* (*DCP*) mechanism [224], in which the electron and proton transfer steps are coupled directly via electrostatics, thus obviating the need for large conformational changes and structural interactions over long distances [158].

The crucial steps of the *DCP* mechanism in *CcO* are the same for all transitions of the catalytic cycle involving the transfer of electrons and protons, i.e. the transitions from A $\rightarrow$ F, F $\rightarrow$ O_H, O_H $\rightarrow$ E, and E $\rightarrow$ R. Only the details vary. The proton-coupled electron transfer starts with a fast, $\approx$10 μs [227], electron transfer from Cu_A to the low spin *heme a*. This transfer is pure electron tunneling without a linkage to proton movements [158]. In a second phase of $\approx$150 μs a proton is taken up from the N-side of the membrane via the D-channel and translocated to the *PLS* [227]. Initially, it was proposed that the uptake of a proton towards the *PLS* is coupled to the oxidation of *heme a* [228], other suggestions and experimental results indicate a coupling of the proton transfer step to the reduction of *heme a* [229, 230, 227]. But what is the *PLS*? So far, there is no definite answer to this question. Different propositions include the histidine ligands of Cu_B [231, 228, 232], the *D* or *A propionates* of *heme* a_3 [233, 234, 235], a water cluster in their vicinity [236], as well as a combination of several sites [158]. All sites proposed, however, share one common feature, namely, they are close to the *BNC*, thus ensuring strong, coulombic interactions between the proton and the electron, stabilizing the proton position [231, 170, 237]. With the end of the second phase, the electron is stored in *heme* a_3 of the *BNC* [158]. In a third phase of $\approx$800 μs [227], a second proton is translocated, this time from the N-side of the membrane to the *BNC*, either through the D-channel (in the A $\rightarrow$ F, F $\rightarrow$ O_H, and eventually E $\rightarrow$ R transition) or K-channel (in the O_H $\rightarrow$ E, and eventually E $\rightarrow$ R transition). The transfer of the

second proton towards the *BNC* is expected to lower the pK_a value of the *PLS* and thus to destabilize the *PLS*. Thereby, the proton located at the *PLS* is expelled to the P-side, involving a time constant of ≈2.6 ms. This setup ensures that one proton is pumped per electron consumed [238, 239, 158].

A fundamental property of all proton pumps is the prevention of leaks. In *CcO* several steps of the *DCP* mechanism bear a leak potential abolishing the pumping of protons, for example [158]:

1. An electron transfer to the *BNC* preceding the proton transfer to the *PLS*.
2. A transfer of the first proton to the *BNC*.
3. A backflow of protons following the *PLS* destabilization.

The former two aspects are accounted for by the *water-gated mechanism*, which relies on the transiently filled cavity above the D-channel. Here, the water molecules are suggested to arrange for the proton transfer to the *PLS* or *BNC* [240]. Up to today, this gating mechanism is actively discussed [200, 206, 202]. For the latter aspect, a kinetic gating, including the orientation of the side chain of E286 or the asparagine region in the D-channel, is discussed [241, 242, 243]. So far, however, no definitive gating mechanism is established [158].

Chapter 3

Methodical Background

The following chapter provides the methodical background of this thesis by introducing and discussing the methods and algorithms used. Furthermore, related methods and algorithms are mentioned to provide a proper overview. The chapter starts with an introduction of the *semi-empiric* quantum method *OM2* [244, 245] used in all calculations to account for quantum effects. To get a proper feeling of the advantages and limitations of this *semi-empiric* quantum method, the method is derived from the "holy-grail" of quantum physics, the (time-dependent) Schrödinger equation, by the application of several approximations. The second part of this chapter focuses on the analysis of the *PES*, i.e. identifying and characterizing different kinds of stationary points as well as *MEPs* connecting these points. Additionally, the advantages and drawbacks of algorithms used to identify individual states or pathways are discussed, which leads, finally, to the detailed introduction of the *TN* approach. In principle, *TNs* are nothing else than weighted, undirected graphs, heavily used in graph theory, an area of discrete mathematics. Hence, the last part of this chapter focuses on graph theory, introducing the proper terminology as well as important properties along with algorithms for their determination.

3.1 Semi-Empiric Quantum Methods

In quantum chemistry three different kinds of methods are distinguished [246]. These are:

1. *ab initio* methods
2. *density functional theory* methods
3. *semi-empiric* methods

While *ab initio* methods provide the most accurate results, their application is limited to rather small systems due to high calculation costs. *Semi-empiric* methods, on the other hand, are less accurate than *ab initio* or *density functional theory* methods, but rather cheap in comparison, which allows the investigation of larger systems. In the following subchapters, the *semi-empiric* quantum method *OM2* (Orthogonalization Model 2) [244, 245] is introduced, starting from the time-dependent Schrödinger equation. Along the derivation, several approximations are introduced, e.g. the Born-Oppenheimer approximation [247, 248], the Hartree-Fock approximation [249, 250], and others [251, 252, 253].

3.1.1 Time-Independent Schrödinger Equation

In classical mechanics, Newton's second law of motion

$$\boldsymbol{F} = m \cdot \boldsymbol{a} \tag{3.1}$$

allows the propagation of macroscopic particles in time, given an initial position $\boldsymbol{x_{init}}$ and momentum $\boldsymbol{p_{init}} = m \cdot \boldsymbol{v_{init}}$. A similar propagation would be desirable for microscopic particles, e.g. nuclei and electrons of atoms. However, according to Heisenberg's uncertainty principle [254] it is not possible to simultaneously determine the positions and momenta of such particles, rendering their classical propagation impracticable. Instead, a quantum mechanical propagation is required.

In quantum mechanics, the state of a system is described by a complex-valued wave function Ψ, containing all possible information about the system, e.g. $\Psi = \Psi(x,t)$ with position x and time t for one particle in one dimension. The probability for finding the particle at different positions of x is given by the probability density $|\Psi(x,t)|^2$. Consequently, for the propagation of a microscopic particle in time, the propagation of Ψ in time is required. In 1926, Erwin Schrödinger formulated such an equation [255], today known as the (non-relativistic) time-dependent Schrödinger equation. For one particle in one dimension the equation reads as

$$-\frac{\hbar}{i}\frac{\partial\Psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x,t)}{\partial x^2} + V(x,t)\Psi(x,t) \tag{3.2}$$

with $\hbar = \frac{h}{2\pi}$, the Planck constant h, $i = \sqrt{-1}$, the mass of the particle m, and the potential energy function $V(x,t)$. Using eq. 3.2 it is possible to determine the wave function at any time, provided the knowledge of an initial wave function. In many applications, however, the simpler time-independent Schrödinger equation is used to solve for stationary states, i.e. states with time-independent probability

densities and time-independent expectation values of observables. To derive the time-independent Schrödinger equation, starting from eq. 3.2, the potential energy function is first restricted to a position dependence only, i.e. $V(x,t) \rightarrow V(x)$, which is true as long as the system is not experiencing time-dependent external forces. Thereby, the solution of eq. 3.2, $\Psi(x,t)$, can be determined using the *method of separation of variables* with the ansatz function

$$\Psi(x,t) = \Psi(x)f(t). \tag{3.3}$$

Inserting eq. 3.3 in eq. 3.2 gives

$$-\frac{\hbar}{i}\frac{1}{f(t)}\frac{df(t)}{dt} = -\frac{\hbar^2}{2m}\frac{1}{\Psi(x)}\frac{d^2\Psi(x)}{dx^2} + V(x). \tag{3.4}$$

In eq. 3.4, the left side of the equation only depends on t, while the right side only depends on x. Thus, either side has to be equal to a constant, here named E. Hence, it holds for the left side that

$$-\frac{\hbar}{i}\frac{1}{f(t)}\frac{df(t)}{dt} = E \quad \text{with} \quad f(t) = e^{-\frac{iEt}{\hbar}} \tag{3.5}$$

and for the right side that

$$\underbrace{-\frac{\hbar^2}{2m}\frac{d^2\Psi(x)}{dx^2} + V(x)\Psi(x)}_{H\Psi(x)} = \Psi(x)E \quad \text{with the Hamiltonian} \quad H. \tag{3.6}$$

Eq. 3.6 represents the time-independent Schrödinger equation for one particle in one dimension. From eq. 3.6 it is evident that $V(x)$ and E have the same dimension. Thus, E is an energy and called the energy of the system. Furthermore, the combination of eq. 3.3 and 3.5 yields the wave function of a *stationary state* with constant energy E and time-independent probability density [256, 257].

3.1.2 Born-Oppenheimer Approximation

In 1927 *Jens Born* and *J. Robert Oppenheimer* published an approximate description of the true molecular wave function, $\Psi(\boldsymbol{x}, \boldsymbol{X})$, with electron coordinates $\boldsymbol{x}$ and nuclei coordinates $\boldsymbol{X}$, today known as the Born-Oppenheimer approximation [247], which is up to now indispensable in quantum chemistry

$$\Psi(\boldsymbol{x}, \boldsymbol{X}) = \Psi_{el}(\boldsymbol{x}; \boldsymbol{X})\Psi_N(\boldsymbol{X}) \quad \text{if} \quad \left(\frac{m_e}{m_N}\right)^{1/4} << 1. \tag{3.7}$$

In words eq. 3.7 states, that the true molecular wave function, $\Psi(\boldsymbol{x}, \boldsymbol{X})$, can be approximated, in an adequate way, by the electronic wave function, $\Psi_{el}(\boldsymbol{x}; \boldsymbol{X})$, a function of $\boldsymbol{x}$ depending only parametrically on $\boldsymbol{X}$, and the nuclear wave function, $\Psi_N(\boldsymbol{X})$, a function of $\boldsymbol{X}$, if the electron mass, m_e, is much smaller than the mass of the nuclei, m_N, e.g. $\frac{m_e}{m_N} = \frac{1}{1840}$ for the hydrogen atom [248]. Due to the substantial mass difference of electrons and nuclei, the nuclei are more or less stationary while the electrons move (1). On the other hand, once the nuclei move the electrons immediately adjust their positions (2). Therefore, it is possible to define separate Schrödinger equations for the electrons and nuclei, respectively [258, 248].

Consider the Hamiltonian H from eq. 3.6. For a system with $i = 1, 2, ..., N_{el}$ electrons and $j = 1, 2, ..., N_N$ nuclei it is defined as

$$H(\boldsymbol{x}, \boldsymbol{X}) = T_N(\boldsymbol{X}) + T_{el}(\boldsymbol{x}) + V_{N,N}(\boldsymbol{X}) + V_{el,el}(\boldsymbol{x}) + V_{el,N}(\boldsymbol{x}, \boldsymbol{X}), \tag{3.8}$$

where $T_{el}(\boldsymbol{x})$ and $T_N(\boldsymbol{X})$ represent the kinetic energy of the electrons and nuclei, respectively,

$$T_{el}(\boldsymbol{x}) = -\sum_{i=1}^{N_{el}} \frac{\hbar^2}{2m_{el}} \frac{\partial^2}{\partial \boldsymbol{x_i}^2} \quad T_N(\boldsymbol{X}) = -\sum_{j=1}^{N_N} \frac{\hbar^2}{2m_N} \frac{\partial^2}{\partial \boldsymbol{X_j}^2}, \tag{3.9}$$

while $V_{el,el}(\boldsymbol{x})$ and $V_{N,N}(\boldsymbol{X})$ represent the potential energy of electronic and nuclear repulsion, respectively,

$$V_{el,el}(\boldsymbol{x}) = \sum_{i<k}^{N_{el}} \frac{e^2}{4\pi\epsilon_0 |\boldsymbol{x_i} - \boldsymbol{x_k}|} \quad V_{N,N}(\boldsymbol{X}) = \sum_{j<l}^{N_N} \frac{Z_j Z_l e^2}{4\pi\epsilon_0 |\boldsymbol{X_j} - \boldsymbol{X_l}|}, \tag{3.10}$$

and $V_{el,N}(\boldsymbol{x}, \boldsymbol{X})$ represents the potential energy of electron-nuclei attraction

$$V_{el,N}(\boldsymbol{x}, \boldsymbol{X}) = -\sum_{j=1}^{N_N} \sum_{i=1}^{N_{el}} \frac{Z_j e^2}{4\pi\epsilon_0 |\boldsymbol{x_i} - \boldsymbol{X_j}|}. \tag{3.11}$$

In eqs. 3.8 till 3.10 Z_j and Z_l represent the atomic numbers of nuclei j and l, respectively, e represents the elementary charge, m_N represents the mass of the nuclei, and ϵ_0 represents the vacuum permitivity. Recalling statement (1), the nuclei are almost stationary, while the electrons move. Therefore, it is possible to remove the kinetic energy of the nuclei ($\mathbf{T_N}(\boldsymbol{X}) \approx 0$) and the potential energy of nuclear repulsion ($V_{N,N}(\boldsymbol{X}) \approx$ const) from the Hamiltonian, when investigating the motion of the electrons. This simplification provides the *purely electronic* Hamiltonian

$$H_{el}(\boldsymbol{x};\boldsymbol{X}) = T_{el}(\boldsymbol{x}) + V_{el,el}(\boldsymbol{x}) + V_{el,N}(\boldsymbol{x};\boldsymbol{X}), \tag{3.12}$$

a function of the electron positions, which depends only parametrically on the positions of the nuclei. The corresponding *electronic* Schrödinger equation reads as

$$H_{el}\Psi_{el} = E_{el}\Psi_{el} \tag{3.13}$$

with E_{el} the *purely electronic* energy in the field of the stationary nuclei. Eq. 3.13 allows the determination of E_{el}. The *electronic energy containing internuclear repulsion* U is then simply defined as

$$U = E_{el} + V_{N,N}. \tag{3.14}$$

Once the *electronic* Schrödinger equation is solved, the nuclear motion is considered. Recalling statement (2) the electrons readily adjust to changes in the positions of the nuclei, i.e. a change $\boldsymbol{X_1} \rightarrow \boldsymbol{X_2}$ readily changes Ψ_{el} and U. In other words, $U = U(\boldsymbol{X})$, becomes the potential energy for the motion of the nuclei. The corresponding *nuclear* Schrödinger equation reads as

$$H_N\Psi_N = E\Psi_N \quad \text{with} \quad H_N(\boldsymbol{X}) = -\sum_j^M \frac{\hbar^2}{2m_N}\frac{\partial^2}{\partial \boldsymbol{X_j^2}}, \tag{3.15}$$

where E represents the total energy of the molecule. Alternatively, the motion of the nuclei is treated classically. Overall, the error introduced by the Born-Oppenheimer approximation is small [258, 248].

The following subchapters provide approximate solutions of the *electronic* Schrödinger equation (cf. eq. 3.13). For clarity the index *el* is dropped, unless stated otherwise.

3.1.3 Hartree Approximation

In 1928, *Douglas R. Hartree* formulated an approximation [251], today known as the Hartree approximation, which allows the determination of $\Psi(\boldsymbol{x})$ for an atom with N interacting electrons and a single nucleus. A desirable approach to solve the *electronic* Schrödinger equation for such a system, would be a separate treatment of either electron as already done for electrons and nuclei in the Born-Oppenheimer approximation. However, due to the electronic repulsion term in the *purely electronic* Hamiltonian (cf. eqs. 3.10 and 3.12), such a separation is not possible, because

the potential energy of N interacting electrons cannot be written as the sum of the potential energy of either electron, rendering more approximate separations a necessity [259, 260].

The simplest of these approximations is the Hartree approximation. Here, $\Psi(\boldsymbol{x})$ for N interacting electrons is treated as

$$\Psi(\boldsymbol{x}) = \psi_1(\boldsymbol{x_1})\psi_2(\boldsymbol{x_2})...\psi_N(\boldsymbol{x_N}), \tag{3.16}$$

where $\psi_1(\boldsymbol{x_1}), \psi_2(\boldsymbol{x_2}), ..., \psi_N(\boldsymbol{x_N})$ represent the spatial, one-electron wave functions, the spatial atomic orbitals, for electron $1, 2, ..., N$, respectively. Hence, the electrons are independent, only interacting with each other by a mean-field Coulomb potential. Consequently, the electronic repulsion term (cf eq. 3.10) changes to

$$V_{el,el,i}(\boldsymbol{x}) = \sum_{k \neq i}^{N} \frac{e^2}{4\pi\epsilon_0} \int \boldsymbol{dx_k} \frac{|\psi_k(\boldsymbol{x_k})|^2}{|\boldsymbol{x_i} - \boldsymbol{x_k}|}. \tag{3.17}$$

Finally, $i = 1, 2, ..., N$ one-electron Hartree equations

$$\left(-\frac{\hbar^2}{2m_{el}} \frac{\partial^2}{\partial \boldsymbol{x_i^2}} - \frac{Ze^2}{4\pi\epsilon_0 |\boldsymbol{x_i} - \boldsymbol{X}|} + \sum_{k \neq i}^{N} \frac{e^2}{4\pi\epsilon_0} \int \boldsymbol{dx_k} \frac{|\psi_k(\boldsymbol{x_k})|^2}{|\boldsymbol{x_i} - \boldsymbol{x_k}|} \right) \psi_i(\boldsymbol{x_i}) = \epsilon_i \psi_i(\boldsymbol{x_i}) \tag{3.18}$$

need to be solved in a self-consistent field [259, 260]. A major drawback of the Hartree method is its failure to obey the anti-symmetry requirements of the total wave function $\Psi(\boldsymbol{x})$ demanded by the *Pauli exclusion principle* [261].

3.1.4 Hartree-Fock Approximation

In 1930, *Vladimir A. Fock* [249] and *John C. Slater* [250] emphasized the explicit inclusion of the electronic spin in the Hartree approximation, i.e. using spin atomic orbitals ϕ_i, instead of spatial atomic orbitals ψ_i. Both types of orbitals are related by

$$\phi_i(i) = \phi_i(\boldsymbol{x_i}, m_s) = \psi_i(\boldsymbol{x_i})\sigma(m_s), \tag{3.19}$$

where the spin function σ takes one of two values,

$$\sigma(m_s) = \begin{cases} \alpha & \text{if } m_s = \frac{1}{2} \\ \beta & \text{if } m_s = -\frac{1}{2} \end{cases} \tag{3.20}$$

for the up and down spin of the electrons, respectively. Furthermore, to satisfy the

Pauli exclusion principle the spin atomic orbitals should be anti-symmetrized, using a Slater determinant of spin atomic orbitals

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & ... & \phi_{N-1}(1) & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & ... & \phi_{N-1}(2) & \phi_N(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi_1(N-1) & \phi_2(N-1) & ... & \phi_{N-1}(N-1) & \phi_N(N-1) \\ \phi_1(N) & \phi_2(N) & ... & \phi_{N-1}(N) & \phi_N(N) \end{vmatrix}. \tag{3.21}$$

Finally, $i = 1, 2, ..., N$ one-electron Hartree-Fock equations

$$\begin{aligned} \left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial \boldsymbol{x_i}^2} - \frac{Ze^2}{4\pi\epsilon_0|\boldsymbol{x_i} - \boldsymbol{X}|} + \sum_{k\neq i}^{N} \frac{e^2}{4\pi\epsilon_0} \int \boldsymbol{dx_k} \frac{|\phi_k(\boldsymbol{x_k})|^2}{|\boldsymbol{x_i} - \boldsymbol{x_k}|} \right) \phi_i(\boldsymbol{x_i}) \\ - \sum_{k\neq i}^{N} \delta_{\sigma_i,\sigma_j} \frac{e^2}{4\pi\epsilon_0} \int \boldsymbol{dx_k} \frac{\phi_k^*(\boldsymbol{x_k})\phi_i(\boldsymbol{x_k})}{|\boldsymbol{x_i} - \boldsymbol{x_k}|} \phi_k(\boldsymbol{x_i}) = \epsilon_i \phi_i(\boldsymbol{x_i}) \end{aligned} \tag{3.22}$$

with $\psi_k^*(\boldsymbol{x_k})$ the conjugate complex of $\psi_k(\boldsymbol{x_k})$, need to be solved in a self-consistent field. The last term on the left hand side of eq. 3.22 is called the exchange term. Due to $\delta_{\sigma_i,\sigma_j}$ the term is only non-zero if electrons i and j display an equal spin, i.e. electrons with equal spin tend to avoid each other. The exchange term represents, together with the coulomb term, the two-electron integrals of the Hartree-Fock Hamiltonian, while the kinetic and potential energy term represent the one-electron integrals. Usually, the Hartree-Fock Hamiltonian [259, 246, 262] is abbreviated by $\hat{\mathbf{F}}$, reducing eq. 3.22 to

$$\hat{\mathbf{F}}\phi_i = \epsilon_i \phi_i. \tag{3.23}$$

3.1.5 Linear Combination of Atomic Orbitals Approximation

So far, the equations presented in the previous subchapters were derived for single atoms, i.e. $i = 1, 2, ..., N$ electrons and a single nucleus. However, an extension to molecules, i.e. $i = 1, 2, ..., N$ electrons and several nuclei, is readily possible. The most important difference between atomic spin orbitals ϕ_i and molecular spin orbitals ϕ_j^M is the number of nuclei they are located around. Atomic spin orbitals have a single center, while molecular spin orbitals are polycentric. The *Linear Combination of Atomic Orbitals* approximation provides the $j = 1, 2, ..., N$ molecular spin

orbitals as linear combinations of $i = 1, 2, ..., N$ atomic spin orbitals

$$\phi_j^M = \sum_{i=1}^{N} c_{i,j}\phi_i, \tag{3.24}$$

where $c_{i,j}$ are coefficients, which determine the electron-nuclei membership. This approximation leads, in the end, to the Roothan-Hall equation system [263, 264]

$$\mathbf{FC} = \mathbf{SCE}, \tag{3.25}$$

where **F** is the Fock matrix, **C** is a diagonal matrix containing the coefficients, **E** is a diagonal matrix containing the energies, and **S** is an overlap matrix, which can be solved for the coefficients in a self-consistent field [265, 246].

3.1.6 Differential Overlap Approximations

The calculation of molecular spin orbitals is computationally expensive, scaling with $\mathcal{O}(N^4)$, when screening is not regarded, due to the calculation of numerous two-electron repulsion integrals [266]. *Semi-empiric* quantum methods reduce these costs significantly. Among other things, this is achieved by only considering valence electrons explicitly, while the remaining core electrons are set off against the charge of the nuclei [267]. A further, central approximation of most of the *semi-empiric* quantum methods is the zero differential overlap approximation [252], a counterexample would be the *extended Hückel theory* [268]. Here, the product of two atomic spin orbitals, located on two distinct nuclei, is set to zero, if and only if the atomic spin orbitals are different. Thereby:

1. The overlap matrix **S** in eq. 3.25 is reduced to the unit matrix.
2. One-electron integrals with three centers are set to zero.
3. Two-electron integrals with three or four centers are neglected.

This setup reduces the calculation costs to a $\mathcal{O}(N^2)$ scale. Three distinct treatments for the remaining one or two electron integrals with one or two centers evolved [267, 246, 269]:

1. *Neglect of diatomic differential overlap* (*NDDO*) [270] - No further neglect of electron integrals [267].
2. *Intermediate neglect of differential overlap* (*INDO*) [271] - Neglect of non-coulomb type two-electron integrals with two centers [267].

3. *Complete neglect of differential overlap* (*CNDO*) [253] - Neglect of all electron integrals, which are not coulomb type two-electron integrals with one or two centers [267].

To compensate these approximations, parameters are introduced for some or all integrals. Hence, the remaining electron integrals are either calculated from the atomic spin orbitals, represented by parameters based on few experimental data, or represented by parameters fitted to many experimental data [267]. The three traditional approximations, *NDDO*, *INDO*, and *CNDO*, presented before are based on the former two aspects, while other approximations, e.g. *modified intermediate neglect of differential overlap* [272], *modified neglect of differential overlap* [273, 274], and others [275, 276, 277, 278, 279, 280, 281], are based on the latter aspect.

3.1.7 Orthogonalization Methods

The reduction of eq. 3.25 by the differential overlap approximation is also possible via an orthogonalization of the atomic spin orbitals. Hence, the *semi-empiric* integrals, briefly acknowledged in the previous subchapter, should represent the theoretical integrals with an orthogonal set of atomic spin orbitals [246], thereby justifying the neglect of two-electron integrals with three or four centers [282, 283, 246]. The *semi-empiric* one-electron integrals, on the other hand, are rather different from the theoretical integrals with an orthogonal set of atomic spin orbitals, giving rise to several deficiencies concerning bonding and antibonding orbitals or repulsion, which leads to inaccuracies in conformation-dependent properties [244, 245]. Consequently, the *semi-empiric* one-electron integrals need to be parametrized to account for the aforementioned orthogonalization effects. These corrections are included in the *semi-empiric orthogonalization methods* (*OM*) *1* [284, 285], *2* [244, 245], and *3* [246], which are all based on the *NDDO* approximation. In *OM1* only one-electron integrals with one center are corrected, while *OM2* is also correcting one-electron integrals wit two centers. Latter corrections are important for conformation-dependent properties. *OM3* is based on the same formalism as *OM2* with a few simplifications [246].

3.2 Potential Energy Surface

The following subchapter provides a closer look at the *PES* as a central concept in the computational study of molecular structures and the basis for the under-

standing of chemical reactions. The *PES* relates the energy of a molecule, a collection of electrons and nuclei, to its geometry, or to be more precise, to the position of its nuclei only, ignoring the coordinates of the electrons. This simplification is based on the *Born-Oppenheimer approximation* [247], which states that the nuclei of a molecule are stationary compared to its electrons. Hence, the permanent geometric parameters of a molecule, e.g. bond lengths or angles, are defined by the nuclear coordinates alone. The dimensionality D of a *PES* is $3N$, where N is the number of atoms in the system. Thus, the *PES* of a mono-atomic system is three-dimensional ($x_1, y_1,$ and z_1), the *PES* of a di-atomic system is six-dimensional ($x_1, x_2, y_1, y_2, z_1,$ and z_2) and so on. Therefore, the *PES* is often called a D-dimensional hypersurface [286]. Several question are frequently asked with regard to the *PES*:

1. Which conformations correspond to local potential energy minima of the *PES*?
2. Which conformations correspond to transition states of the *PES*?
3. Which pathway provides an optimal connection of two local potential energy minima?
4. How to derive reasonable transition pathways on D-dimensional hypersurfaces?

The following subchapters address these questions. Therefore, definitions for individual stationary points of the *PES* are provided and the main ideas behind specific minimization and transition state finding algorithms are presented. Furthermore, the concept of *MEPs* is discussed and the *TN* approach, as a technique capable of deriving reasonable transition pathways on D-dimensional hypersurfaces, is introduced in detail. This subchapter is concluded by a short introduction of *MD* simulations.

3.2.1 Stationary Points

Local potential energy minima (and maxima) or transition states are called stationary points of the *PES*. Here, the first order partial derivative of the potential energy function $V = V(\boldsymbol{x})$, the gradient $\boldsymbol{g} = \boldsymbol{g}(\boldsymbol{x})$, is zero with respect to the x-, y-, and z-coordinates of all N atoms $\boldsymbol{x} = (x_0, x_1, ..., x_{D-1})$

$$\boldsymbol{g} = \frac{\partial V}{\partial \boldsymbol{x}} = \boldsymbol{0}. \tag{3.26}$$

A further classification of the stationary points is possible when looking at the composition of the eigenvalues, λ_i with $i = 0, 1, .., D-1$, of the Hessian matrix $\mathbf{H} = \mathbf{H}(\boldsymbol{x})$, containing the second order partial derivatives of V, at the stationary point

$$\mathbf{H} = \frac{\partial^2 V}{\partial \boldsymbol{x}^2} \quad \text{with} \quad \mathbf{H_{ij}} = \frac{\partial^2 V}{\partial x_i \partial x_j} \quad \text{and} \quad i, j = 0, 1, ..., D-1. \tag{3.27}$$

Four different compositions of λ_i need to be distinguished [287]:

1. $\lambda_i > 0$ for $i = 0, 1, .., D-1 \Rightarrow$ The stationary point corresponds to a local potential energy minimum.

2. $\lambda_i < 0$ for $i = 0, 1, .., D-1 \Rightarrow$ The stationary point corresponds to a local potential energy maximum.

3. $\lambda_i \lessgtr 0$ for $i = 0, 1, .., D-1 \Rightarrow$ The stationary point corresponds to a saddle point.

4. $\lambda_k = 0$ for k in $i \Rightarrow$ Further information are necessary to classify the stationary point.

Local potential energy minima (and maxima) of the *PES* are clearly defined. On the other hand, saddle points of the *PES* exist in variety of forms. However, only first order saddle points, i.e. saddle points with only one negative eigenvalue, are considered to be transition states. Therefore, it holds that at a certain transition state, there is one direction in which the transition state is a local potential energy maximum, whereas it is a local potential energy minimum in all other directions [288].

Minimization algorithms, for the determination of local potential energy minima, are at the heart of most computational investigations of molecular structures. Their frequent use renders two distinct aspects a necessity: 1) the algorithms should converge towards the local potential energy minimum within a reasonable amount of steps (or time), and 2) the storage requirements should be kept at a minimum. In the last decades, a variety of numerical methods used to identify local potential energy minima has been formulated, e.g. gradient descent minimizations [289, 290, 291], and Newton minimizations [292, 293]. Here, the main ideas underlying three minimization algorithms are summarized. These are:

1. Steepest descent (*SD*) minimization by *Cauchy* [291, 294]

2. Conjugate gradients (*CG*) minimization by *Fletcher* and *Reeves* [289]

3. Adopted basis Newton-Raphson (*ABNR*) minimization by *States et al* [293]

All three minimization algorithms follow the same scheme, i.e. the local potential energy minimum $\boldsymbol{x_m}$ is determined as the limit of a sequence of states $\boldsymbol{x_i}$ with $i = 0, 1, ..., m$ and some initial approximation $\boldsymbol{x_0}$. The directions along which the local potential energy minima are searched, are derived along the minimizations and differ in between the minimization algorithms. This setup is much more efficient than minimization algorithms with pre-defined search directions [289], e.g. the alternating directions method.

In the *SD* and *CG* minimization algorithms, succeeding positions $\boldsymbol{x_{i+1}}$ are defined as

$$\boldsymbol{x_{i+1}} = \boldsymbol{x_i} - \alpha_i \boldsymbol{d_i} \quad \text{with} \quad i = 0, 1, ..., m-1, \tag{3.28}$$

where $\boldsymbol{d_i}$ and α_i denote the search direction and step size, respectively. For the *SD* minimization algorithm $\boldsymbol{d_i}$ is defined as the negative gradient $-\boldsymbol{g_i}$ while it is the **H**-conjugate gradient $\boldsymbol{s_i}$ in the *CG* minimization algorithm. In both algorithms the step size is defined by a line search

$$\alpha_i = \arg\min_{\alpha} V(\boldsymbol{x_i} - \alpha_i \boldsymbol{g_i}). \tag{3.29}$$

The *ABNR* algorithm combines the Newton-Raphson (*NR*) minimization algorithm with the *SD* minimization algorithm. Thus, succeeding positions are defined as

$$\boldsymbol{x_{i+1}} = \boldsymbol{x_i} + \Delta \boldsymbol{x_i^{NR}} + \Delta \boldsymbol{x_i^{SD}} \quad \text{with} \quad i = 0, 1, ..., m-1, \tag{3.30}$$

where $\Delta \boldsymbol{x_i^{NR}}$ and $\Delta \boldsymbol{x_i^{SD}}$ define the displacement due to the *NR* and *SD* minimization algorithms, respectively. In the NR minimization algorithm succeeding positions are defined as

$$\boldsymbol{x_{i+1}} = \boldsymbol{x_i} - \mathbf{H_i^{-1}} \boldsymbol{g_i} \quad \text{with} \quad i = 0, 1, ..., m-1, \tag{3.31}$$

where $\mathbf{H_i^{-1}}$ defines the inverse of $\mathbf{H_i}$. Eq. 3.31 provides a fast convergence towards $\boldsymbol{x_m}$ in terms of iteration steps, however, the calculation and handling of **H** in every iteration step is rather time consuming [292]. Therefore, the *ABNR* minimization algorithm applies the *NR* minimization algorithm to a subspace of the system only. This subspace is derived by the *SD* minimization algorithm.

Overall, the storage requirements for all three minimization algorithms are rather

low, while the efficiency increases from the *SD* minimization algorithm, to the *CG* minimization algorithm, to the *ABNR* minimization algorithm. The latter algorithm is used for all potential energy minimizations within the framework of this thesis, while the former algorithm is crucial for the determination of *MEPs*. The *CG* minimization algorithm, on the other hand, is neither used nor crucial, its main ideas, however, underlie the conjugate peak refinement (*CPR*) algorithm used within the framework of this thesis for the determination of transition states. A detailed description of all three minimization algorithms can be found in Appendix A.

Apart from the identification of conformations representing local potential energy minima of the *PES*, the rates k for individual transitions among these conformations are of great interest. Harmonic transition state theory provides an accurate estimate of the transition rate k

$$k = \frac{\prod_i^D \nu_{R,i}}{\prod_i^{D-1} \nu_{TS,i}} e^{-\frac{V_{TS}-V_R}{k_B T}} \tag{3.32}$$

with ν_i the normal mode frequencies, R the reactant state, and TS the transition state. Thereby, the determination of the transition rate is boiled down to the search of a transition state with minimal potential energy V_{TS}, connecting two distinct local potential energy minima. The identification of the relevant transition state, however, is challenging. Once, the transition state is identified the transition pathway (or *MEP*), connecting the local potential energy minima via the transition state, can be determined by following the steepest descent from the transition state towards either local potential energy minimum [295]. Thereby, the tangent direction of the *MEP* is represented, at any point, by $\boldsymbol{g}$, rendering the *MEP* stationary for any perpendicular *DOF* [296]. Although the concept of the *MEP* is widely used, its interpretation is questionable, due to the absence of kinetic energy [297, 298]. A different *MEP* concept, following the direction of least curvature from a local potential energy minimum to the transition state, exists [299, 300], but is not elaborated on further in the framework of this thesis.

In the last decades, a variety of algorithms to identify transition states has been formulated [301, 296, 302, 303, 304, 305, 306, 95, 307, 308, 309, 310, 311, 312, 313]. Furthermore, a variety of algorithms to identify transition pathways has been formulated [301, 296, 91, 314, 315, 316, 317, 318, 319, 320, 321, 322, 92, 323, 324]. Here, the main ideas underlying two transition pathway search algorithms are presented. These are:

1. *CPR* algorithm [92]

2. Nudged elastic band (*NEB*) algorithm [91, 314, 315]

The *CPR* algorithm allows the determination of transition states on a D-dimensional *PES* forming a continuous transition pathway connecting two pre-defined local potential energy minima. The main idea of the algorithm is, that in the vicinity of a transition state any set of **H**-conjugate vectors has one direction, $\boldsymbol{s_0}$, along which there is a local potential energy maximum and $D-1$ directions, $\boldsymbol{s_1}$ till $\boldsymbol{s_{D-1}}$, along which there are local potential energy minima. Thereby, an iterative determination of transition states is possible, performing one arbitrary line maximization and $D-1$ conjugate line minimizations per iteration cycle. At the end a pathway is determined along which every local potential energy maximum represents a transition state. The connection of the individual transition states by the *SD* minimization algorithm provides the *MEP* [288, 92].

The *NEB* algorithm allows the determination of transition pathways, and corresponding transition states, on a D-dimensonal *PES* connecting pre-defined local potential energy minima. The main idea of the algorithm is, that the *MEP* represents a pathway which is stationary for any perpendicular DOF. Thereby, an iterative determination of the *MEP* is possible by optimizing all states along an initial pathway, represented by states connected by springs, until they represent potential energy minima in all directions perpendicular to the pathway. To achieve a proper description of the transition state region the perpendicular and parallel component of the spring force and gradient, acting on each state, needs to be projected out, respectively. At the end a pathway is determined which represents the *MEP* along with all transition states.

The comparison of the *CPR* and *NEB* algorithm suggests that both algorithms are able to determine the transition states along the *MEP* connecting pre-defined local potential energy minima. To determine the actual *MEP*, however, the *CPR* algorithm relies on subsequent *SD* minimizations, while it is directly provided by the *NEB* algorithm. Still, the *CPR* algorithm is used within the framework of this thesis for all transition pathway calculations within the *TN* approach, since it is the maximal transition state along the pathway which is relevant and not the exact *MEP*. A detailed introduction of both transition pathway search algorithms is presented in Appendix B along with helpful examples and improvements.

3.2.2 Transition Network Approach

A drawback of the *CPR*, *NEB* and basically all other transition pathway finding algorithms is, that the quality of the transition pathway finally derived, depends

to great extent on the pathway initially guessed, which is in most cases simply a straight line connecting the end states of the transition [92, 91]. Only the transition pathways whose transition states are closest to the initial guess will be found [100]. The emerging problem gets already apparent, when investigating rather simple, low-dimensional problems. In Figs. B.1 and B.2, for example, two distinct transition states with equal potential energies are located on the *PES*. The *CPR* and *NEB* algorithm, however, only provide one state, due to its proximity to the initial guess. Here, the determined transition rates would still be correct. In other situations, however, in which farther transition states display lower potential energies, the transition rates would be incorrect. Apart from the calculation of incorrect transition rates, an entirely different transition mechanism is overlooked, when only identifying one transition state. These problems inevitable grow with the complexity of the system, resulting in a limited and biased description of most complex transitions of interest.

The *TN* approach attenuates these problems, dividing the complex transition of interest into a series of less complex sub-transitions, connecting various intermediate states. For these sub-transitions, the crude straight line initial guess used by the transition pathway finding algorithms is sufficient to receive optimal sub-transition pathways. Piecing together all the sub-transitions, provides the *TN* describing the complex transition of interest. The identification of the optimal transition pathways is then achieved by standard graph theoretical methods [102, 101, 325].

The determination of *TNs* starts with the definition of the *DOFs* to sample. To prevent a bias stemming from the selection of the *DOFs*, most (if not all) *DOFs* relevant for the transition of interest need to be sampled. The sampling is discrete and uniform [325] or discrete with random selections from the uniform distribution [326]. States displaying clashes of *heavy* atoms, i.e. pairs of non-hydrogen atoms closer than 1 Å, are readily rejected. Four different types of *DOFs* may be considered:

1. Bond length sampling

   The bond length is the most important *DOF* to sample, when covalent bonds need to be formed or broken during the transition of interest, e.g. the *bromide-alkyl chloride-alkyl exchange* reaction [327] (cf Fig. 3.1).

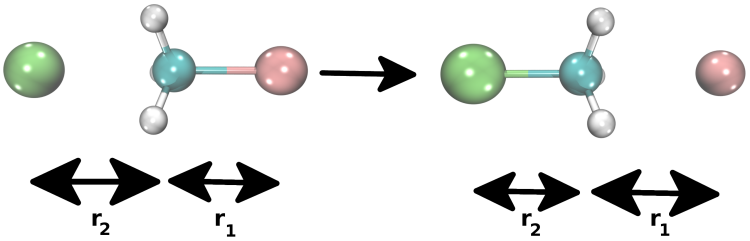


Figure 3.1: Initial and final state of the *bromide-alkyl chloride-alkyl exchange* reaction, hydrogens (white), carbon (turquoise), bromide (ocher), and chloride (lime). The distance between the carbon and bromide atom is labeled r_1, the distance between the carbon and chloride atom is labeled r_2.

Here, the distance between the carbon and bromide atom r_1 is stepwise increased, while the distance between the carbon and chloride atom r_2 is stepwise decreased. The sampling of bond lengths is irrespective of the atom types involved. A manual definition is required.

2. Water and metal ion position sampling

   The motion of water molecules plays a crucial role in protein dynamics [328] and the transfer of protons [329]. Their sampling is stepwise in x-, y-, and z-direction (cf Fig. 3.2, left).

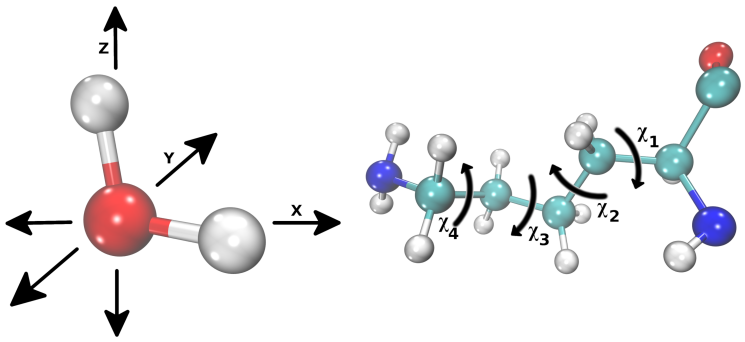


Figure 3.2: Left: Schematic sampling of different water positions in x-, y-, and z-direction. Right: The side chain dihedral angles χ_1 (rotation around the C_α-C_β bond), χ_2 (rotation around the C_β-C_γ bond), χ_3 (rotation around the C_γ-C_δ bond), and χ_4 (rotation around the C_δ-C_ϵ bond) of lysine.

A manual definition of the sampling interval, and of all water molecules to be sampled, is required. The same applies to the sampling of metal ions.

3. Side chain dihedral angle sampling

   Side chain dihedral angle rotations play an important role in various types of reactions, e.g. by opening/occluding cavities [330] or stabilizing, disrupting,

and re-forming hydrogen bond networks [200]. Their sampling includes stepwise rotations around the initial conformation of χ_1, χ_2, χ_3, χ_4, and χ_5 (cf. Fig. 3.2, right). In Table 3.1 the atoms involved in the side chain dihedral angles, χ_1 till χ_5, of all 20 standard, proteinogenic amino acids are listed. The two non-standard, proteinogenic amino acids, selenocysteine and pyrrolysine, are not considered.

Table 3.1: Atoms involved in the side chain dihedral angles, χ_1 till χ_5, of all 20 standard, proteinogenic amino acids (AA, one letter code).

AA	χ_1	χ_2	χ_3	χ_4	χ_5
A	-	-	-	-	-
R	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma C_\delta$	$C_\beta C_\gamma C_\delta N_\epsilon$	$C_\gamma C_\delta N_\epsilon C_\zeta$	$C_\delta N_\epsilon C_\zeta N_\eta$
N	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma N_{\delta,1}$	-	-	-
D	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma O_{\delta,1}$	-	-	-
C	$NC_\alpha C_\beta S_\gamma$	-	-	-	-
Q	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma C_\delta$	$C_\beta C_\gamma C_\delta O_{\epsilon,1}$	-	-
E	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma C_\delta$	$C_\beta C_\gamma C_\delta O_{\epsilon,1}$	-	-
G	-	-	-	-	-
H	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma N_{\delta,1}$	-	-	-
I	$NC_\alpha C_\beta C_{\gamma,1}$	$C_\alpha C_\beta C_{\gamma,1} C_\delta$	-	-	-
L	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma C_{\delta,1}$	-	-	-
K	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma C_\delta$	$C_\beta C_\gamma C_\delta C_\epsilon$	$C_\gamma C_\delta C_\epsilon N_\zeta$	-
M	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma S_\delta$	$C_\beta C_\gamma S_\delta C_\epsilon$	-	-
F	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma C_{\delta,1}$	-	-	-
P	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma C_\delta$	-	-	-
S	$NC_\alpha C_\beta O_\gamma$	-	-	-	-
T	$NC_\alpha C_\beta O_{\gamma,1}$	-	-	-	-
W	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma C_{\delta,1}$	-	-	-
Y	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma$-$C_{\delta,1}$	-	-	-
V	$NC_\alpha C_\beta C_{\gamma,1}$	-	-	-	-

A manual definition of the sampling interval, and of all side chain dihedral angles to be sampled, is required.

4. Protonation state sampling

   The protonation state is the most important *DOF* to sample, when protons need to be transferred along the transition of interest, e.g. the transfer of

protons along water wires [331] (cf Fig. 3.3).

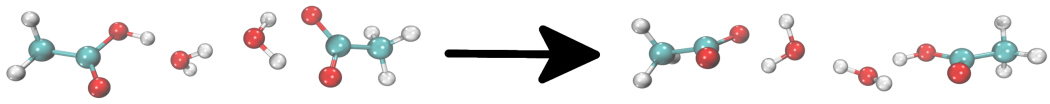


Figure 3.3: Initial and final state of the proton transfer along a wire of two water molecules.

In principle, a protonation state change is nothing but a breaking and forming of $O-H$, $N-H$, or $S-H$ bonds. Here, however, the actual protanation states are more important than the process of bond breaking or forming. In Tab. 3.2 generally considered protonation states of water and individual amino acid side chains are listed.

Table 3.2: Generally considered protontation states of water and individual amino acid side chains.

Species	Protonation States	Charge		
Water	3	-1	0	+1
Aspartate	3	-1	0	0
Glutamate	3	-1	0	0
Histidine	3	0	0	1
Lysine	2	0	1	

Other formally possible protonation states, e.g. double-protonated aspartate, glutatmate and water [332], or deprotonated serine [333] and cysteine [334], are generally not considered. A manual definition of all possible protonation states is necessary. Here, the rejection criterion, stated before, is extended to hydrogen atoms [325].

Once the set of initial states is generated, they are subjected to a potential energy minimization using the *ABNR* minimization algorithm with a convergence criterion of 0.001 kcal/(molÅ). The local potential energy minima thus derived form the *nodes* of the *TN*. Due to the potential energy minimization, the initial state assignment of either node is, in most cases, no longer correct. Hence, each node is assigned a new state, represented by a vector, with respect to the initial configuration by:

1. Comparing all *DOFs* sampled initially [325].

2. Comparing all *DOFs* present in the system in terms of protonation state, water positions, and side chain dihedral angles [325].

The initial state vector contains zeros for all *DOFs*, except the initial protonation state, which is denoted by the residue number of the species protonated initially, while the entries in the state vectors of all other nodes display the necessary step numbers along either *DOF* and the residue numbers of the species protonated in either node. If several nodes are assigned an equal state vector, the node with the lowest potential energy is chosen as representative. Furthermore, a relabeling of atoms and molecules is performed, using the Hungarian algorithm [335], to avoid distinct nodes representing chemically equivalent structure, which only differ in their labels, e.g. rotated, deprotonated carboxyl groups. Finally, all nodes, which differ by a user-defined number of maximal steps only, are declared neighbors [325].

For each neighbor pair a transition pathway is determined using the *CPR* algorithm. To ensure a minimal difference between the initial and final state of either sub-transition a thorough relabeling of atoms and molecules is performed. The resulting transition pathways are called the *edges* of the *TN*. The corresponding edge weights are defined as the maximal potential energy along the transition pathways with respect to the initial state of the overall transition. Hence, the edge weights are independent of the direction and thus the *TN* resembles an undirected, weighted graph. Finally, the optimal pathways for the overall transition are determined using a variant of Dijkstra's algorithm [336].

3.2.3 Molecular Dynamics Simulations

Standard *MD* simulations propagate the nuclei of N atomic systems, e.g. molecules, glasses, liquids, or solids, in time by solving Newton's equation of motion. Thereby, a sequence of points, a trajectory, is generated in a $6N$-dimensional phase space spanned by the positions ($3N$-dimensional) and momenta ($3N$-dimensional) of all atoms. The equation of motion for N atoms reads as

$$m_i \frac{\partial^2 \boldsymbol{x_i}(t)}{\partial t} = \boldsymbol{F_i}(t) = -\frac{\partial V}{\partial \boldsymbol{x_i}(t)} \quad \text{with} \quad i = 0, 1, ..., N, \tag{3.33}$$

where m_i, $\boldsymbol{x_i}(t)$, and $\boldsymbol{F_i}(t)$ represent the mass of atom i, its position at time t, and the force acting on it at time t, respectively. The propagation in time is then achieved by a two step procedure:

1. Determine the force acting on each atom according to eq. 3.33.

2. Propagate the atom positions in time.

Hence, the *PES* is essential for standard *MD* simulations. For the propagation of

the positions of all atoms (step two) an integration algorithm is needed. One of the simplest algorithms is the *Verlet integrator* [337], which is based on two Taylor expansions of the atom position

$$\boldsymbol{x_i}(t+\delta t) = \boldsymbol{x_i}(t) + \boldsymbol{v_i}(t)\delta t + \frac{\boldsymbol{F_i}(t)}{2m_i}\delta t^2 + \frac{1}{3!}\frac{\partial^3 \boldsymbol{x_i}(t)}{\partial t^3}\delta t^3 + \mathcal{O}(\delta t^4), \tag{3.34}$$

$$\boldsymbol{x_i}(t-\delta t) = \boldsymbol{x_i}(t) - \boldsymbol{v_i}(t)\delta t + \frac{\boldsymbol{F_i}(t)}{2m_i}\delta t^2 - \frac{1}{3!}\frac{\partial^3 \boldsymbol{x_i}(t)}{\partial t^3}\delta t^3 + \mathcal{O}(\delta t^4), \tag{3.35}$$

where $\boldsymbol{v_i}(t)$ represents the velocity of atom i at time t. Adding up eqs. 3.34 and 3.35 gives

$$\boldsymbol{x_i}(t+\delta t) \approx 2\boldsymbol{x_i}(t) - \boldsymbol{x_i}(t-\delta t) + \frac{\boldsymbol{F_i}(t)}{m_i}\delta t^2 \tag{3.36}$$

with an error of $\mathcal{O}(\delta t^4)$. More suitable integration algorithms are the *Velocity Verlet integrator* or the *Leap Frog integrator*. In all cases the initial positions and velocities of all N atoms need to be specified [338].

Newton's equation of motion is energy conserving. Consequently, standard *MD* simulations are performed under constant total energy which is not appropriate for the investigation of many biochemical systems. Alternatively, it is possible to perform *MD* simulations under constant temperature and pressure. For a detailed introduction the interested reader is referred to [338].

3.3 Graph Theory

Graph theory is one branch of the wide field of discrete mathematics. Here, discrete structures, *graphs* $G = (V, E)$, containing a set of $|V|$ *vertices* ($V = v_1, v_2..., v_{|V|}$) and $|E|$ *edges* ($E = e_1, e_2, ..., e_{|E|}$) with or without *edge weights* ($\Omega = \omega_1, \omega_2, ..., \omega_{|E|}$) connecting individual *vertices*, are used to model problems in various fields [339], including:

1. Social structure models

   Graphs in which the *vertices* represent individual persons and the *edges* represent different kinds of relationships [339, 340, 341, 342], e.g. the friendship graph [343], the acquaintance graph [344], the influence graph [345], or the collaboration graph [346]. Among other things used to analyze the spreading of diseases [347].

2. Communication structure models

   Graphs in which the *vertices* represent individual devices and the *edges* represent different kinds of communication links [339], e.g. the call graph [348]. Among other things used to analyze the spreading of mobile phone viruses [349].

3. Information structure models

   Graphs in which the *vertices* represent individual information sources and the *edges* represent different kinds of information links [339], e.g. the web graph [350], the citation graph [351], or Google's knowledge graph [352]. Among other things used to analyze the choices of consumers [353].

4. Software structure models

   Graphs in which the *vertices* represent individual modules and the *edges* represent dependencies [339], e.g. the module depending graph [354], or the precedence graph [355]. Among other things used to analyze time-cost trade-off problems [356].

5. Transport structure models

   Graphs in which the *vertices* represent individual locations and the *edges* represent different kinds of transport ways [339], e.g. the flight graph [357], the train graph [358, 359], or the road graph [360]. Among other things used to minimize transport costs [361].

6. Biological structure models

   Graphs in which the *vertices* represent individual biological entities and the *edges* represent different kinds of interactions [339], e.g. the protein interaction graph [362, 363], or the niche overlap graph [364]. Among other things used to analyze the brain efficiency under an alcohol dependency [365].

The *graphs* mentioned in points one to six display a variety of different properties. A friendship, for example, is usually a mutual relationship, while an influence is side-directed, giving rise to *graphs* with undirected and directed *edges*, respectively. On the other hand, having several friendships with the same person or a friendship with one-self is rather unusual, while individual locations are, in most cases, connected by various transport ways and software modules are able to call themselves. All these properties, i.e. *edge* direction, *edge* multiplicity, and *edge* loops, are used

to distinguish six types of *graphs* [339]:

Table 3.3: Types of *graphs* and their *edge* characteristics.

Graph Type	Direction	Multiplicity	Loops
Simple Graph	Undirected	No	No
Multigraph	Undirected	Yes	No
Pseudograph	Undirected	Yes	Yes
Simple directed Graph	Directed	No	No
Directed Multigraph	Directed	Yes	No
Mixed Graph	Dircted/Undirected	Yes	Yes

In the following subchapters the basic terminology used and fundamental properties investigated in graph theory are introduced, pathways and (spanning) trees connecting distinct vertices are discussed, the concepts of robust and sensitivity analyses are presented, and a small insight into spectral graph theory is provided. The focus of all subchapters lies on simple graphs, which include *TNs* as introduced before.

3.3.1 Basic Graph Terminology

In this subchapter, the basic terminology used in graph theory is introduced, e.g. adjacency, incident, degree, subgraph, or clique. Furthermore, fundamental properties like the isomorphism, or the connectedness are discussed.

- Basic Graph Terminology

  Two vertices in a simple graph are called adjacent (or neighbors) if they represent the endpoints of an edge. Consequently, the neighborhood of a vertex is defined as the set of all its adjacent vertices, where the edges connecting distinct vertices are said to be incident. The degree of a vertex is defined by the number of incident edges, i.e. the size of the neighborhood. The summation of all vertex degrees provides $2|E|$. All these properties may be used to represent a graph (cf. Fig. 3.4),

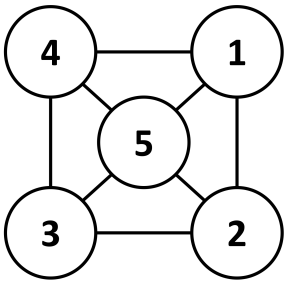


Figure 3.4: Simple graph with $|V| = 5$ vertices (V=v_1, v_2, v_3, v_4, v_5), four with a degree of three and one with a degree of four, and $|E| = 8$ edges (E=$e_{1,2}$, $e_{2,3}$, $e_{3,4}$, $e_{1,4}$, $e_{1,5}$, $e_{2,5}$, $e_{3,5}$, $e_{4,5}$).

in an adjacency list,

Table 3.4: Adjacency list of a simple graph.

Vertex	Adjacent Vertices
1	2,4,5
2	1,3,5
3	2,4,5
4	1,3,5
5	1,2,3,4

in an adjacency matrix $\mathbf{A}$,

$$\mathbf{A} = \begin{array}{c|ccccc} & v_1 & v_2 & v_3 & v_4 & v_5 \\ v_1 & 0 & 1 & 0 & 1 & 1 \\ v_2 & 1 & 0 & 1 & 0 & 1 \\ v_3 & 0 & 1 & 0 & 1 & 1 \\ v_4 & 1 & 0 & 1 & 0 & 1 \\ v_5 & 1 & 1 & 1 & 1 & 0 \end{array} \tag{3.37}$$

$$a_{i,j} = \begin{cases} 1 & \text{if } e_{i,j} \text{ is an edge of G} \\ 0 & \text{if } e_{i,j} \text{ is not an edge of G} \end{cases},$$

in an incidents matrix $\mathbf{M}$,

$$\mathbf{M} = \begin{array}{c|cccccccc} & e_{1,2} & e_{1,4} & e_{2,3} & e_{3,4} & e_{1,5} & e_{2,5} & e_{3,5} & e_{4,5} \\ v_1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ v_2 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ v_3 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ v_4 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ v_5 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{array} \tag{3.38}$$

$$m_{ij} = \begin{cases} 1 & \text{if } e_{i,j} \text{ is incident to } v_k \\ 0 & \text{if } e_{i,j} \text{ is not incident to } v_k \end{cases},$$

or in a degree matrix $\mathbf{D}$,

$$\mathbf{D} = \begin{array}{c|ccccc} & v_1 & v_2 & v_3 & v_4 & v_5 \\ v_1 & 3 & 0 & 0 & 0 & 0 \\ v_2 & 0 & 3 & 0 & 0 & 0 \\ v_3 & 0 & 0 & 3 & 0 & 0 \\ v_4 & 0 & 0 & 0 & 3 & 0 \\ v_5 & 0 & 0 & 0 & 0 & 4 \end{array}. \tag{3.39}$$

Obviously, the most convenient representation depends on the problem to be investigated, e.g. adjacency lists for sparse graphs and adjacency matrices for dense graphs [339].

- *Isomorphism* of Graphs

  The *isomorphism* of graphs is an important property in a variety of fields, e.g. image processing [366, 367], protein structure analysis [368, 369, 370], computer aided design [371, 372, 373], chemical structure analysis [374], or road maps [375]. Two graphs are called *isomorphic* if there is a one-to-one correspondence between their vertices and thus a preserved adjacency relationship (cf. Fig. 3.5). The determination of graph *isomorphism* is rather difficult, especially for large graphs, since there are $|V|!$ one-to-one vertex correspondences to check. Determining, on the other hand, if two graphs are *non-isomorphic* is, at least in some cases, much simpler, when considering properties that need to be conserved across the graphs, so called *graph invariants*, e.g. for simple graphs, the number of vertices $|V|$, the number of edges $|E|$, and the degree of the vertices $deg(V)$ needs to be constant [339].

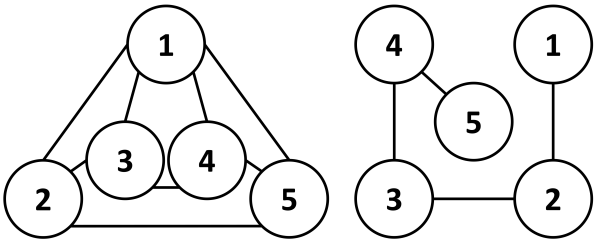


Figure 3.5: Left: Simple graph with $|V| = 5$ vertices, four with a degree of three and one with a degree of four, and $|E| = 8$ edges. Right: Simple graph with $|V| = 5$ vertices, three with a degree of two and two with a degree of one, and $|E| = 4$ edges. The left graph is isomorphic to the graph from Fig 3.4.

- *Connectivity* of Graphs

  The *connectivity* of graphs is another property which is important in a variety of fields, e.g. communication networks [376], protein interaction networks [377], brain networks [378], or traffic control [379]. A graph is said to be *connected* if every vertex can be reached from any other vertex following the edges of the graph. If that is not the case, the graph is said to be *disconnected*, containing a set of maximally connected subgraphs. Hence, a *disconnected* graph represents the union of maximally connected subgraphs. The connectivity of a graph is measured by the vertex- or edge-connectivity, i.e. the number of vertices (and corresponding edges) or edges, so called cut vertices or cut edges, which need to be removed from the graph to produce a disconnected subgraph. Simple, complete graphs, $K_{|V|}$, containing $|E| = \frac{|V|^2-|V|}{2}$ edges, and other graphs lacking cut vertices are called inseparable graphs. The more vertices or edges need to be removed to obtain a disconnected subgraph, the higher the connectivity of the graph. In general it holds, that the number of vertices to remove to disconnect a graph, is smaller or equal than the number of edges to remove. The minimal vertex degree represents the upper bound for both types of connectivity [339]. For the graphs in Fig. 3.5, for example, three, or one, vertices, or edges, need to be removed to disconnect the left and right graph, respectively.

- *Cliques* in Graphs

  The identification of *cliques* in graphs is essential in a variety of fields, e.g. the clustering of gene expression data [380], the modeling of ecological niches [381], the protein structure prediction [382], or the modeling of chemical bounds [383]. A c*lique* is a subset of vertices of a graph, which represent a complete subgraph. In other words, all subset vertices are adjacent. In many cases *maximal cliques*, i.e. *cliques* which cannot be extended by adding another

vertex, or *maximum cliques*, i.e. *cliques* with maximal vertex number, need to be identified. The opposite of a clique is the *independent set*, which is a set of non-adjacent vertices. Consequently, the *maximal independent set* cannot be extended by adding another vertex and a *maximum independent set* contains a maximal number of vertices. The *clique* and *independent set* problem is *NP-complete* [384], rendering approximate algorithms, in most cases, a necessity for their determination [339].

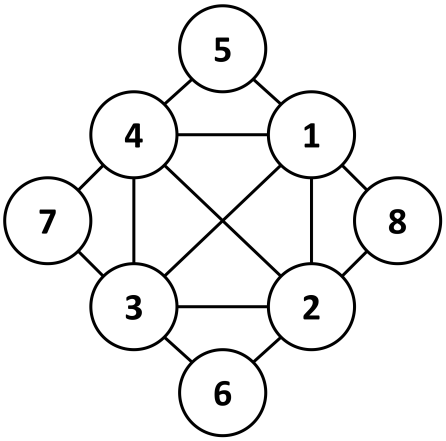


Figure 3.6: Simple graph with $|V| = 8$ vertices and $|E| = 14$ edges. The graph contains eight *one-vertex cliques* (the vertices), fourteen *two-vertex cliques* (the edges), eight *three-vertex cliques* (1-2-5, 2-4-8, 3-4-6, 1-3-7, 1-2-3, 1-2-4, 2-3-4, 1-3-4), and one *four-vertex clique* (1-2-3-4). The four *three-vertex cliques* (1-2-5, 2-4-8, 3-4-6, 1-3-7) represent *maximal cliques*, while the *four-vertex clique* represents the *maximum clique*.

3.3.2 Pathways

A pathway on a simple graph is defined as a sequence of edges traveling through the vertex set, starting at an initial and ending at a final vertex. If the pathway starts and ends at the same vertex, it is called a circuit. A pathway or circuit containing every edge only once is called simple [339]. In the remainder of this subchapter the Euler [385] and Hamilton pathway [386, 387] are introduced. Furthermore, three types of optimal pathways, the shortest pathway (SP), the widest pathway (WP), and the minimax best pathway (MBP), are presented along with appropriate algorithms, the Dijkstra [336] and the Floyd-Warshall algorithm [388, 389], for their determination.

The Euler pathway denotes a pathway traversing through a graph including all available edges only once. Hence, the Euler pathway is a simple pathway containing all available edges. If the pathway starts and ends at the same vertex it is called an

Euler circuit, while the graph containing either of these structures is called *eularian* [339, 390]. Euler pathways and circuits are of great interest in a variety of fields, e.g. the Chinese postman problem [390], the usage of communication links [339], and the assembly of DNA fragments [391, 392]. The concept of Euler circuits (and pathways) emerged in 1736, when Euler investigated the seven bridges in Königsberg problem [385], which is seen today as the foundation of graph theory [390]. Königsberg, today known as Kaliningrad, was located at the Pregel river, which divided the city into four different sections. In 1736, seven bridges connected the individual sections of Königsberg with each other. An abstract representation of the city is achieved by a multigraph in which the vertices represent the individual sections and the edges represent the bridges (cf. Fig 3.7).

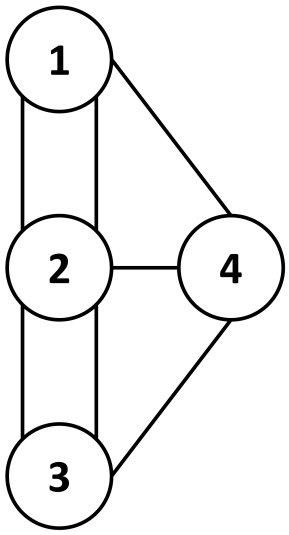


Figure 3.7: Multigraph representing the city of Königsberg in 1736 with four individual sections, or $|V| = 4$ vertices, and seven bridges, or $|E| = 7$ edges.

According to the legend, the inhabitants of Königsberg wondered if it is possible to start in any section of Königsberg, cross all bridges once, and thus return to the initial section. Following Euler, this is not possible, because for a graph to contain an Euler circuit all vertices need to be even. Vice versa, a graph that contains an Euler circuit involves even degrees only (cf. Fig. 3.8 left). Hence, both statements are equivalent. Another equivalent statement is the possibility to partition the edge set of the graph into cycles [385, 393, 394]. The condition for an Euler pathway, on the other hand, is that the graph contains only two vertices with odd degree [390] (cf. Fig. 3.8 right).

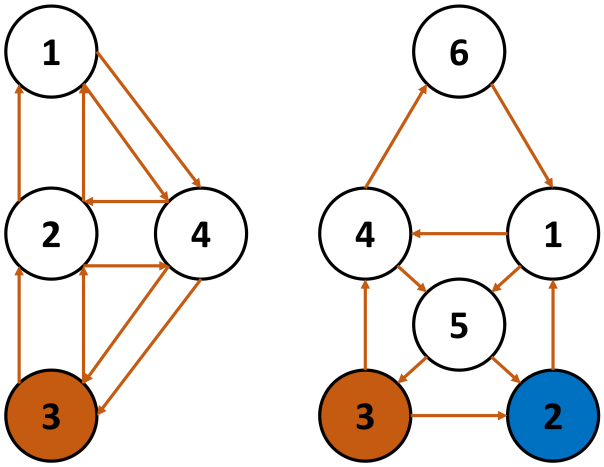


Figure 3.8: Left: Multigraph with $|V| = 4$ vertices, all with even degree, and $|E| = 10$ edges. The orange vertex represents the start and end vertex of an Euler circuit following the orange arrows. Right: Simple graph with $|V| = 5$ vertices, three with even and two with odd degree, and $|E| = 8$ edges. The orange and blue vertex represent the start and end vertex of an Euler pathway following the orange arrows.

The Hamilton pathway denotes a pathway traversing through a graph including all available vertices only once. If the pathway starts and ends at the same vertex it is called a Hamilton circuit, while the graph containing either of these structures is called *hamiltonian* [339]. The problem of finding Hamilton pathways and circuits is *NP-complete*. Hamilton pathways and circuits are of great interest, e.g. the traveling salesman problem [395]. The concept of Hamilton circuits (and pathways) emerged in 1857, when Hamilton invented the *Icosian puzzle*, a game played on a dodecahedron with 20 vertices. To win one had to start at one of the vertices and travel through all the other once and return to the start. Other than for the Euler pathway no simple conditions for a Hamilton pathway or circuit are known today. However, the more edges a graph possesses the more likely it is to find a Hamilton pathway or circuit. On the other hand, it is not possible to find a Hamilton circuit in a graph with a vertex with a degree of one [339].

The Euler and Hamilton pathway, introduced before, focused on unweighted, simple graphs. A variety of problems, however, is optimally modeled by weighted graphs. Here, each edge is assigned a weight, e.g. traveling times, traveling distances, communication costs or rate-determining, maximal transition barriers in *TNs*. The search for an optimal pathway connecting two distinct vertices evolves naturally from the introduction of weights [339]. In general, three different kinds of optimal pathways are distinguished (cf. Fig. 3.9):

1. *SPs* - Pathways minimizing the sum of weights.

2. *WPs* - Pathways maximizing the minimal weight.

3. *MBPs* - Pathways minimizing the maximal weight.

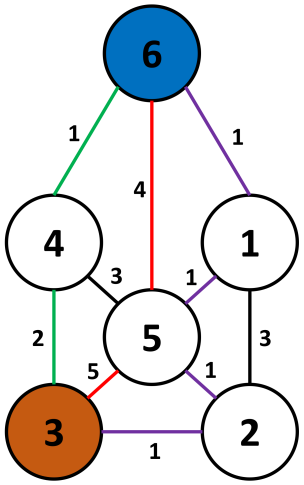


Figure 3.9: Weighted, simple graph with $|V| = 6$ vertices and $|E| = 8$ edges. The *SP*, connecting a start and end vertex depicted in orange and blue, respectively, is green, the *WP* is red, and the *MBP* is purple.

All three types of optimal pathways may be used for the same problem. However, the focus is shifted. In transportation graphs, for example, the *SP* provides the fastest transportation route connecting two distinct vertices [396], the *WP* provides a transportation route with maximal capacity [397], and the *MBP* provides a pathway with a minimal distance to individual service units [398].

A wide range of algorithms exists to determine optimal pathways [399, 400, 401, 402, 403, 404, 405, 406, 407]. Here, the main idea behind the Dijkstra [336] and Floyd-Warshall algorithm [388, 389] is introduced along with advantages and disadvantages of either algorithm. Detailed descriptions of both algorithms, with helpful examples, can be found in Appendix D.

The Dijkstra algorithm [336] provides a solution to the single-source *SP* problem. It starts with an initial pathway of length 0, the source vertex, and subsequently extends a pathway to any vertex. Therefore, in each cycle of the algorithm the vertex with minimal distance to the source vertex is identified and the distance of its neighbors to the source is updated. This is done until every vertex has been visited. The two main ideas underlying the Dijkstra algorithm are the concept of the optimal substructure, meaning that a sub-pathway of an *SP* is an *SP* itself, and the triangle inequality, meaning that the length of the pathway connecting vertices u and v via x is always larger (or equal) to the length of the *SP* connecting vertices

u and v without x. The former idea renders the algorithm *non-interative*, i.e. each vertex needs to be visited only once, resulting in time complexities of $\mathcal{O}(|V|^2)$ to $\mathcal{O}(|E|+|V|log(|V|))$, depending on the structure of the vertex queue. Furthermore, the algorithm is *uninformed*, removing the necessity of a prior knowledge of the target vertex, which increases the costs due to a blind search, the algorithm cannot handle negative edge weights, and the algorithm provides the length of the *SP* and the corresponding pathway [339, 408, 409].

The Floyd-Warshall algorithm [388, 389] provides a solution to the all-pairs *SP* problem. It starts with $|E|$ pathways, representing the edges of the graph, and pathway lengths according to the edge weights. In $|V|$ cycles, pathways going through the vertices v_1, v_1 and v_2, v_1 and v_2 and v_3, and so on, are determined as assembly of shorter pathways from previous cycles. If the new pathway, connecting vertices u and v via x, displays a shorter length than the old pathway, connecting vertices u and v, it is replaced, providing, at the end, the *SPs* between all vertices. The main idea underlying the Floyd-Warshall algorithm is the concept of the optimal substructure, already introduced for the Dijkstra algorithm. Thereby, a time complexity of $\mathcal{O}(|V|^3)$ is reached. Furthermore, the algorithm is simple to implement, the algorithm can handle positive and negative edge weights, but no negative cycles, and the algorithm only provides the length of the *SPs* but not the corresponding pathways [410, 411].

The *TN* approach is used, in most cases, to determine the rate-determining, maximal transition barrier of the *MBP*, and the corresponding pathway, connecting a pre-defined source and target vertex in a *TN* possibly containing negative edge weights. Both algorithms introduced before can be used with little modifications for the determination of *MBPs* and the corresponding pathways. These modifications would also remove the non-negative edge weights and non-negative cycle limitations. Overall, however, the Dijkstra algorithm satisfies the need of the *TN* approach more appropriately, especially when considering that the blind search limitation is removed by the definition of a target vertex.

3.3.3 Trees

A connected, undirected graph without any cycles is called a *tree*. Due to the absence of multiple edges and cycles, every *tree* is a simple graph. Vice versa, every simple graph, containing unique, simple pathways between all pairs of vertices, is a *tree*. Consequently, a tree with $|V| = N$ vertices contains $|E| = N - 1$ edges. Any *tree* containing less than $N - 1$ edges, i.e. a disconnected *tree*, is called a *forest* [412].

A special class of *trees* are *rooted trees*. Here, an individual vertex of the *tree* is defined as root, while all edges are assigned a direction increasing the distance to the root. Hence, different roots will generate different *rooted trees*. In a *rooted tree* any vertex v, except the root, has a unique parent vertex u, with a directed edge from vertex u to vertex v. Therefore, vertex v is called a child of vertex u. If two or more vertices possess the same parent, they are called siblings. The vertices along the pathway from the root to v (excluding v) are ancestors of v, while all vertices having v as ancestor are descendants of v. A vertex without a child is called a leaf, a vertex with a child is called an internal vertex [412].

A connected, simple graph with an arbitrary edge number can be transferred into a tree by the removal of all cycles. Such subgraph trees, containing all vertices of the original simple graph, are called *spanning trees* (*ST*s). The depth-first search [413] or breadth-first search [414] algorithms allow an efficient determination of *STs*. If the simple graph is weighted, it is possible to determine a *minimum spanning tree* (*MST*), i.e. an *ST* with minimal edge weight sum, using various algorithms [415, 416, 417, 418, 419, 420]. In principle, the *MST* represents the *MBPs* between all vertices in a simple graph [412].

Trees, *rooted trees*, *STs*, and *MSTs* are applied in a wide range of fields, e.g. 1) in computer science for searching (binary search trees), sorting (decision trees), and coding, to represent data structures and for the data networking [412], 2) in chemistry for the enumeration of saturated hydrocarbons [421], 3) in geology for the description of environmental transport, i.e. the flux of water, sediments, nutrients, and others [422], and 4) in psychology for the analysis of psychopathic traits [423].

In the remainder of this subchapter, two algorithms for the determination of the *MST*, Boruvkas's algorithm [417] and Kruskal's algorithm [416], are introduced along with advantages and disadvantages of either algorithm. Detailed descriptions of both algorithms, with helpful examples, can be found in Appendix E.

Boruvka's algorithm [417] contracts a graph along the edges with minimal edge weight incident to either connected component and removes multiple edges, except the one with minimal edge weight, and self-loops in the contracted graph. Each contracted edge is added to the *MST* until no further edges are left to contract. Thereby, the algorithm reaches a time complexity of $\mathcal{O}(|E|log(|V|))$. Furthermore, the algorithm is easily parallelized, requires less work per edge than other algorithms, and displays the least locality of all *MST* algorithms [424].

Kruskal's algorithm [416] separates the graph into vertices and edges sorted from low to high edge weights. Subsequently, edges are added to the vertices, in sorted

order, forming the *MST*. If an added edge produces a cycle it is removed. Thereby, the algorithm reaches a time complexity of $\mathcal{O}(|E|log(|V|))$. Furthermore, the algorithm displays a good performance for large $|V|$ and $|E|$ and considers each edge only once [424].

3.3.4 Robust Analysis

The determination of optimal pathways and *MSTs* constitutes classical optimization problems, which means that all edge weights, e.g. costs, lengths, and others, are precisely known and thereby fixed. For many real world applications, however, fixed edge weights represent a severe restriction, e.g. considering telecommunication graphs or traffic graphs. The simplest way to account for fluctuating edge weights is the introduction of independent, closed edge weight intervals [425].

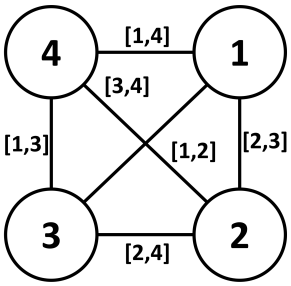


Figure 3.10: Complete graph with $|V| = 4$ vertices and $|E| = 6$ edges. The edge weights are represented by independent, closed edge weight intervals.

The *robust approach* [426] is used to model optimization problems with fluctuating edge weights. Here, the worst case scenario of a solution is minimized over all possible sets of edge weights using the maximal regret, i.e. the maximal difference between the solution and an optimal solution, as robust criterion. Other robust criteria are possible [427]. The solutions thus obtained are most stable and resistant towards edge weight fluctuations. Unfortunately, most robust analyses are *NP-hard*, rendering their investigation problematic [426]. Nevertheless, some algorithms have been developed to determine *robust spanning trees* (*RSTs*) [428, 429, 430], *robust shortest pathways* (*RSPs*) [431, 432, 433], and others for the investigation of various problems [434]. In the following, the general idea behind the *robust approach* is introduced.

A graph with $|E|$ edges, $e_1, e_2, ..., e_{|E|}$, gives rise to a set Φ of (at most) $2^{|E|}$ edge subsets, representing all feasible solutions of the optimization problem, e.g. all *STs* for the determination of the *RST* or all pathways connecting two distinct vertices

for the determination of the *RSP* connecting the vertices in question. For every edge $e \in E$ an independent, closed edge weight interval is defined as $\omega_e = [\omega_e^-, \omega_e^+]$. A particular assignment of edge weights to either edge is called a scenario S for which it holds that $S = \omega_e^S \in [\omega_e^-, \omega_e^+] : e \in E$, while Γ represents the set of all scenarios. The costs F for a given solution X under a fixed scenario are then defined as

$$F(X, S) = \sum_{e \in E} \omega_e^S, \tag{3.40}$$

while the costs of the optimal solution F^* under scenario S are defined as

$$F^*(S) = \min_{X \in \Phi} F(X, S). \tag{3.41}$$

The maximal regret Z is then defined as the maximal difference between the given and optimal solution, when considering all scenarios from Γ

$$Z(X) = \max_{S \in \Phi} \left(F(X, S) - F^*(S) \right), \tag{3.42}$$

thus representing the worst case scenario for solution X. Finally, the maximal regret is minimized over all solutions Φ

$$R = \min_{X \in \Phi} Z(X). \tag{3.43}$$

The solution corresponding to R is the robust edge set, e.g. the *RST* or *RSP*.

Apart from the algorithms stated before, approximate algorithms for the determination of *RSTs* or *RSP*s, e.g. the midpoint approximation algorithm (AM), the upper bound approximation algorithm (AU), and the combined midpoint and upper bound approximation algorithm (AMU), are highly efficient and introduce only minor deviations compared to the optimal solution [425]. The AM algorithm reduces the independent, closed edge weight intervals to the midpoint of the interval, while the AU algorithm reduces the independent, closed edge weight intervals to the upper bound of the interval. The worst case performance ratio of the AM algorithm is two, while the worst case performance of the AU algorithm is unbounded. Still, the AU algorithm provides, more often than not, a better solution than the AM algorithm. Therefore, the AMU algorithm employs both algorithms and chooses the better solution [412]. An exemplary determination of an *RST* for the graph depicted in Fig. 3.10 using the AM algorithm is shown in Fig 3.11.

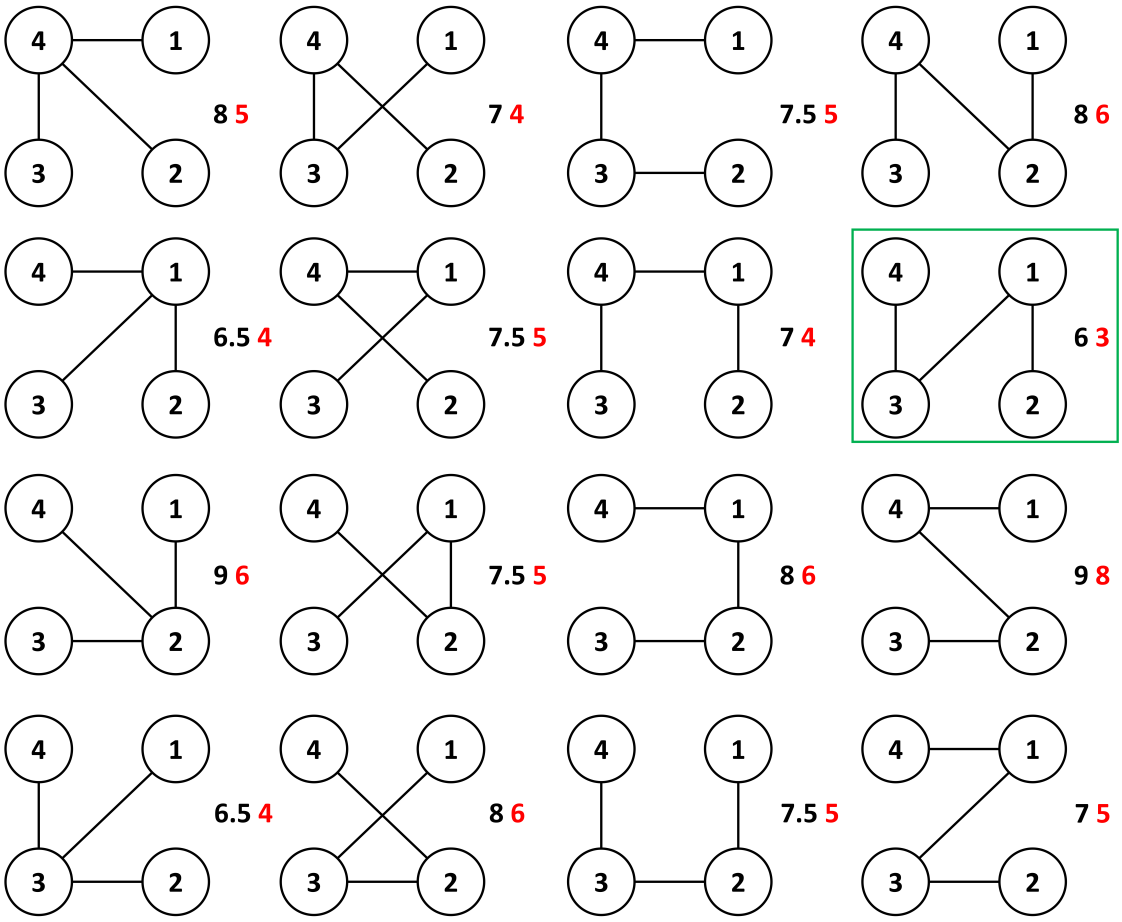


Figure 3.11: All 16 *STs* of the complete graph depicted in Fig. 3.10. The cost of either tree is depicted in black, when using the midpoint of all independent, closed edge weights intervals according to the *AM* algorithm. The maximal regret of either *ST* is depicted in red. The green box indicates, the *ST* with minimal maximal regret, representing the *RST*.

3.3.5 Sensitivity Analysis

A different approach to account for fluctuating edge weights is the sensitivity analysis of *MSTs* and optimal pathways towards individual edge weight changes. Therefore, a basic question is asked, namely "By how much can the weights of all edges in a graph change individually without affecting the *MST* or the optimal pathway?" Four different cases need to be distinguished:

1. Weight increase for non-*MST* or non optimal pathway edges.
2. Weight decrease for *MST* or optimal pathway edges.
3. Weight decrease for non-*MST* or non optimal pathway edges.
4. Weight increase for *MST* or optimal pathway edges.

The first two points are trivial, because the weight of a non-*MST* or non optimal pathway edge may be increased to infinity without affecting either structure. Likewise, a weight decrease for *MST* or optimal pathway edges down to minus infinity has no influence on either structure. In this respect, points three and four are far more interesting, due to actual implications of edge weight decreases or increases. As described before, an *MST* contains no cycles. Consequently, the addition of a non-*MST* edge to the *MST* introduces a cycle formed by at least two *MST* edges and the non-*MST* edge. To maintain the initial *MST*, the weight of the non-*MST* edge needs to stay larger than the weight of the *MST* edges. On the other hand, the weight of the *MST* edges needs to stay smaller than the weight of the non-*MST* edge. Hence, the sensitivity value of non-*MST* or *MST* edges is defined as

$$Sens(e) = \begin{cases} \omega_f - \omega_e & \text{if } e \text{ and } f \text{ in cycle and } e \notin \text{MST},\ f \in \text{MST} \\ \omega_e - \omega_f & \text{if } e \text{ and } f \text{ in cycle and } e \in \text{MST},\ f \notin \text{MST} \end{cases}, \tag{3.44}$$

For the optimal pathways the same sensitivity criterion can be applied. Here, however, it needs to be considered that the addition of a non optimal pathway edge to the optimal pathway is not necessarily introducing a cycle. Therefore, an optimal pathway may be insensitive to a wide range of edges in the graph (cf. Fig. 3.12).

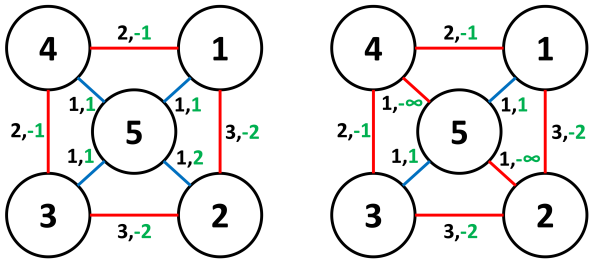


Figure 3.12: Left: Weighted, simple graph with $|V| = 5$ vertices and $|E| = 8$ edges. Blue edges indicate the *MST*, red edges indicate non *MST* edges. Black numbers represent the edge weights, green numbers represent the sensitivity of either edge. Right: Weighted, simple graph with $|V| = 5$ vertices and $|E| = 8$ edges. Blue edges indicate the *MBP* connecting vertices 1 and 3, red edges indicate non *MBP* edges. Black numbers represent the edge weights, green numbers represent the sensitivity of either edge.

Different algorithms have been developed to determine the sensitivity of *MSTs* and optimal pathways using *transmuter graphs* [435], a "divide-and-conquer" approach [436] or *MSTs* [437].

3.3.6 Spectral Graph Theory

In spectral graph theory, the eigenvalues of a graph are used to determine its structural properties. Likewise, structural properties are used to determine the nature of the eigenvalues [438]. In principle, almost all graph theoretical invariants are closely related to the eigenvalues, thus underlining their central role [439]. Spectral graph theory has a long standing history. Nowadays [439], it is applied in a wide range of fields, e.g. mathematics, chemistry, theoretical physics, and quantum mechanics. Roughly half of these investigations focus on the identification of boundary values for the eigenvalues, while the others focus on the impact and consequences of certain boundaries [439].

Matrices, which may be used for a spectral analysis are [438]:

1. the adjacency matrix $\mathbf{A}$,
2. the (combinatorial) *Laplacian* matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$ with the degree matrix $\mathbf{D}$,
3. the sign-less *Laplacian* matrix $\mathbf{L} = \mathbf{D} + \mathbf{A}$,
4. and the normalized *Laplacian* matrix $\mathbf{L_N} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$.

The eigenvalues of all these matrices allow the determination of some (but not all structural properties), e.g. the eigenvalues of the adjacency matrix indicate a bipartite graph as well as the number of edges, but not the number of connected components or the number of bipartite components [438].

The $|V| \times |V|$ *Laplacian* matrix $\mathbf{L}$ is symmetric. Therefore, it holds that the $|V|$ eigenvalues, λ_0, λ_1, λ_2, ..., $\lambda_{|V|-1}$, are real and the corresponding eigenvectors, $\boldsymbol{v_0}$, $\boldsymbol{v_1}$, , $\boldsymbol{v_2}$, ..., $\boldsymbol{v_{|V|-1}}$, form an orthonormal basis. Furthermore, it holds that $\mathbf{L1} = \mathbf{0}$, giving rise to the following sequence of n eigenvalues [440],

$$0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{|V|-1}. \tag{3.45}$$

The second smallest eigenvalue λ_1 is generally referred to as Fiedler's eigenvalue [441] or the algebraic connectivity, and is receiving more and more attention. If $\lambda_1 = 0$ the graph is disconnected. On the other hand, the larger λ_1 the more connected is the graph. For a disconnected graph, according to λ_1, the corresponding eigenvector $\boldsymbol{v_1}$ partitions the graph into two components. These components represent connected components of the graph if $\lambda_2 \neq 0$. In general, if $\lambda_i = 0$ and $\lambda_{i+1} \neq 0$ the graph possesses $i + 1$ connected components.

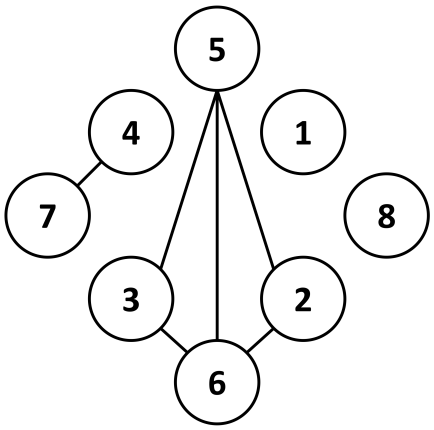


Figure 3.13: Disconnected graph with $|V| = 8$ vertices and $|E| = 6$ edges. The eigenvalues of the *Laplacian* matrix are $\lambda_0 = \lambda_1 = \lambda_2 = \lambda_3 = 0$, $\lambda_4 = \lambda_5 = 2$, and $\lambda_6 = \lambda_7 = 4$. Hence, the graph contains four connected components, $4 - 7$, $2 - 3 - 5 - 6$, 1, and 8. $\boldsymbol{v_1}$ splits the vertices $1, 2, 3, 4, 5, 6, 7, 8$ in two sets $4, 7$ and $1, 2, 3, 5, 6, 8$. $\boldsymbol{v_2}$ splits the vertices $1, 2, 3, 5, 6, 8$ in two sets 1 and $2, 3, 5, 6, 8$. $\boldsymbol{v_3}$ splits the vertices $2, 3, 5, 6, 8$ in two sets 8 and $2, 3, 5, 6$, yielding the four connected components.

For further spectral graph properties, e.g. cospectral graphs, and relations, e.g. the mixing of random walks , the interested reader is referred to the literature [440, 438, 439].

Chapter 4

Results I: Method Development

The following chapter focuses on the two central methods developed within the framework of this thesis, namely, the coupling of *TN* calculations with *MD* simulations, the *TN-MD method*, for the optimal sampling of *DOFs* generating large numbers of states separated by low transition barriers, as well as the *TN prediction method*, which uses information from a known, unperturbed or initial *TN* and an excessive, two-step coarse-graining procedure to improve the generation of an unknown, perturbed *TN*. For either of these methods, the underlying idea is introduced along with the implementation. Furthermore, model calculations are presented and discussed to work out benefits and limitations of the individual methods. Both methods are fully documented in *Reidelbach et al* [442] (*TN-MD method*) and *Reidelbach et al* [443] (*TN prediction method*) and presented here in the following subchapters.

4.1 The TN-MD Method

The *TN* approach divides a complex transition of interest into a set of simpler subtransitions. Therefore, a discrete sampling is performed along all *DOFs* relevant for the complex transition. For proton transfer reactions, the sampling is, in most cases, limited to different protonation states, different side chain dihedral angle conformations, and different positions and orientations of the water molecules. These *DOFs* may be divided into two distinct groups:

1. *DOFs* generating a low number of states, separated by high transition barriers.

2. *DOFs* generating a large number of states, separated by low transition barriers.

The side chain dihedral angle conformations are representatives for the former group of *DOFs*, while the positions and orientations of the water molecules are represen-

tatives for the latter group of *DOFs*. The protonation states, on the other hand, are hybrids generating only a low number of states with moderate transition barriers. In the following, however, they are considered as members of group one.

The main problem of *TN* calculations is the exponential increase of local minima and transition states on the PES with an increasing number of *DOFs* to be sampled [112]. Obviously, the *DOFs* of the latter group represent the main cause of this problem. Simultaneously, however, these *DOFs* are least dependent on the *TN* approach. In other words, a usual *MD* simulation, is already generating a comprehensive set of energetically favorable states for *DOFs* of the latter group. Therefore, a coupled *TN-MD* method was developed, which applies the direct sampling of the *TN* approach to *DOFs* of the first group only, while *DOFs* of the second group are sampled by *MD* simulations and subsequently incorporated into the *TN*. The remaining *TN-MD* method follows the steps introduced before for the standard *TN* approach.

In the following subchapters the *TN-MD method* is tested extensively. Therefore, the proton transfer through a small, water-filled model channel is analyzed using the *TN-MD method* as well as the standard *TN* approach. Along the presentation of the *TN* calculations, snapshots of the *MD* simulations are examined with regard to properties relevant for the transfer protons, e.g. hydrogen bonds.

4.1.1 Background, Methods and Settings I

The *TN-MD* method was used to investigate the proton transfer through a small model channel. In principle, the model channel was a cylinder filled with thirteen water molecules, an excess proton and two aspartate-like molecules at either end. Chemically speaking, each of the two aspartate-like molecules was composed of a carboxyl group, representing the side chain of the aspartate amino acid, and a bulky t-butyl structure, representing the backbone of the aspartate amino acid. Both parts were connected via the central carbon atom of the t-butyl structures. The t-butyl structures were spatially constrained, forming the top and bottom of the cylinder, while the cylinder walls were modeled by a harmonic potential setting on at a distance of 3 Å away from the cylindrical axis with a force constant of 500 kcal/(molÅ). Two distinct configurations with the excess proton located at either of the two aspartate-like molecules were modeled, representing the reactant and product state of the overall proton translocation through the model channel (cf. Fig. 4.1).

For the sampling of proton transfer pathways, *TNs*, i.e. weighted, simple graphs,

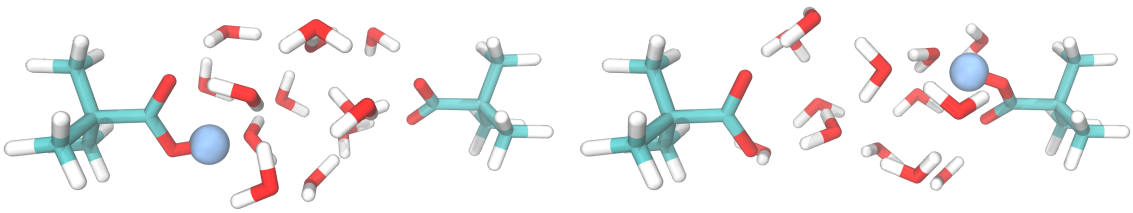


Figure 4.1: Model channel for the calculation of proton transfer pathways containing two aspartate-like molecules, thirteen water molecules, and an excess proton. Left: reactant state of the overall proton translocation, right: product state of the overall proton translocation. The blue sphere represents the excess proton, highlighting the change in protonation state.

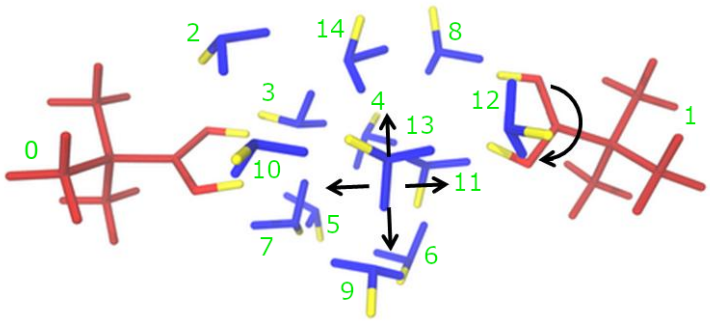


Figure 4.2: Model channel for the calculation of proton transfer pathways containing two aspartate-like molecules and thirteen water molecules, labeled in green from 0 to 14, as well an excess proton. Sampled protonation sites are indicated in yellow, while black arrows indicate the sampling of different dihedral angle conformations and water positions.

were constructed with nodes generated from various initial structures. These initial structures differed in their protonation state, their side chain dihedral angle conformations, and their water positions and orientations (cf. Fig. 4.2).

Here, three distinct approaches, to generate the initial structures, were compared:

1. Direct Sampling Approach

   Sampling of 17 different protonation states by placing the excess proton on either water molecule and on the carboxyl groups of the two aspartate-like molecules. For each protonation state the side chain dihedral angles were rotated in 45° intervals to sample individual side chain dihedral angle conformations. Overall, 1 088 initial states would need to be investigated for the direct sampling approach.

2. Direct Sampling Approach with Hydronium Ion Motions

   Sampling of 17 different protonation states by placing the excess proton on either water molecule and on the carboxyl groups of the two aspartate-like

molecules. For each protonation state the side chain dihedral angles were rotated in 45° intervals to sample individual side chain dihedral angle conformations. Furthermore, the motion of all hydronium ions was sampled by 2 Å translations in x-, y-, and z-direction (backward and forward). Overall, 6 080 initial states would need to be investigated for the direct sampling approach with hydronium ion motions.

3. Direct Sampling Approach with *MD* simulations

   Sampling of 17 different protonation states by placing the excess proton on either water molecule and on the carboxyl groups of the two aspartate-like molecules. For each protonation state the side chain dihedral angles were rotated in 45° intervals to sample individual side chain dihedral angle conformations. Furthermore, the motion of all water molecules was sampled by short *MD* simulations. Overall, 14 088 initial states, i.e. 1 088 initial states from the direct sampling approach and 1 000 initial states per *MD* simulation of either protonated water molecule, would need to be investigated for the direct sampling approach with *MD* simulations.

The former two approaches represent variations of the standard *TN* approach, while the latter approach is representing the *TN-MD method*. A comparison of all three approaches by the number of initial states to be investigated revealed that the *TN-MD method*, intended to reduce the number of initial states, displayed the highest amount of initial states to be investigated. The sampling of the former two approaches, however, was incomplete in terms of the sampling of the motion of the water molecules. Consequently, a comparison of the *TN-MD method* with the complete, standard *TN* approach (not performed, but generating about 10^{14} initial states) by the number of initial states to be investigated revealed an initial state reduction of almost 100 %.

The initial states of all three approaches were subjected to *QM* potential energy minimization according to the *ABNR* minimization algorithm using the *semi-empiric* quantum method *OM2* and a convergence criterion of 0.001 kcal/(molÅ). An evaluation of the *OM2* method for the computation of proton translocations along small *HBCs* is presented in Appendix F. Following the *QM* potential energy minimizations, all minimum energy conformations were assigned a network node or discrete state. This assignment was based on the conformation of the side chain dihedral angles of the two aspartate-like molecules, the protonation state of the system (defined by the location of the excess proton on either of the water molecules

or the aspartate-like molecules), and the position and orientation of all thirteen water molecules, commonly referred to as the water pattern of the system. All state characterizations were performed in comparison to the reactant state of the overall proton translocation.

Each assigned node was labeled as SC0.SC1.wi.P, where SC0 and SC1 refer to the number of dihedral angle steps of 45° of the side chains of molecule 0 and 1, i.e. the left and right aspartate-like molecule (cf. Fig. 4.2). It is important to note, that, due to the symmetry of the carboxyl group, a rotation of 180° (four steps) corresponds to the conformation of the reactant state of the overall proton translocation and is thus treated as a rotation of 0°. With the same argument, a backward rotation of three steps corresponds to a forward rotation of one step. Next, wi denotes the i-th water pattern of all thirteen water molecules of the model system. Different water patterns can be achieved by the translation of a single water molecule (or several water molecules) with a discretized displacement step of 2 Å, as well as the rotation of a single water molecule (or several water molecules) with a discretized rotation step of 22.5°. Consequently, the complete water pattern is defined by 6 digits per water molecule, three digits for the translation steps in x-, y, and z-direction, and three digits for the rotation steps around the x-, y-, and z-axis. Thus, the water patterns of the model system under investigation are composed of 78 digits. The index i, employed to label the individual water patterns, is just a counter for the compact notation of the water patterns. Its value is not quantifying the distance between the individual water patterns. The protonation state of the system is indicated by P and defined by the number of the molecule carrying the excess proton. In principle, multiple protonations (and deprotonations) are possible, but not sampled in this work.

The reactant state of the overall proton translocation is labeled 0.0.w0.0, corresponding to zero rotations of either of the side chain dihedral angles, zero translational or rotational displacements of all thirteen water molecules, and an excess proton located at the left aspartate-like molecule. A hypothetical node 0.2.w231.3 (SC0=0.SC1=2.wi=w231.P=3), for example, indicates the following displacements with respect to the reactant state of the overall proton translocation: zero rotation around the side chain dihedral angle of the left aspartate-like molecule, two rotation steps (90°) around the side chain dihedral angle of the right aspartate-like molecule, a water pattern with number 231, and the excess proton located on molecule 3 (a water molecule, cf. Fig. 4.2).

Transitions between pairs of nodes, v_i and v_k, were derived by *MEP* calculations

using the CPR algorithm as implemented in CHARMM [444]. These transitions represent the edges of the *TN*. As described before, the *MEP* connects two distinct nodes following a pathway along which all potential energy maxima correspond to first order saddle points or transition states. If an *MEP* contains more than one transition state, intermediate states are located along the *MEP*. According to the transition state theory, the transition state with the highest potential energy is rate-determining. Consequently, its relative potential energy is used as an edge weight, or cost, for going from node v_i to node v_k along the edge $e_{i,k}$ in the *TN*. Here, two nodes were considered as a neighboring pair, if they differed in each *DOF*, i.e. side chain dihedral angles, proton position, and water pattern, by at most one step.

All nodes and edges were then combined to form the *TN*. To determine the pathway which optimizes the transition time between the reactant and product state of the overall proton translocation the pathway which minimizes the sum over all transition times needs to be determined. The transition time between two nodes v_i and v_k is proportional to

$$w_{i,k} = exp\left[\frac{V_{i,k}}{k_B T}\right], \tag{4.1}$$

where $V_{i,k}$ is the potential energy of the transition state between nodes v_i and v_k, k_B is Boltzmann's constant, and T is the temperature (cf. eq 3.32). If $w_{i,k}$ is used as edge weight, the cost of a pathway $P(v_j, ..., v_n)$ along the *TN* is given by

$$C(P) = \sum_{j-1}^{n} w_{j,j+1}. \tag{4.2}$$

Due to the exponential weighting, this sum is dominated by the potential energy of the highest transition state. Hence, the optimal pathway connecting the reactant and product state of the overall proton translocation is the one for which the maximal transition state is minimal compared to all other pathways. In graph theory such a pathway is called the *MBP*. In the *TN* approach such pathways are determined by a variant of the Dijkstra algorithm. In addition, to the *MBP* the second, third (and so on) best pathways were determined.

All simulations presented in the following were performed by the CHARMM program [444]. *QM* potential energy minimizations and *CPR* calculations were performed by the CHARMM program interfaced to MNDO [445] (unless stated otherwise). The generation of the initial states, according to the direct sampling approaches introduced before, the assignment of nodes, the determination of neighbors, as well as the compilation and analysis of the *TNs* were carried out by existing

source code [325] and libraries from Noe et al [446].

4.1.2 Results and Discussion I

The following subchapter provides the three *TNs* introduced in the previous subchapter, starting with the *TN* following the direct sampling approach (approach one), continuing with the *TN* following the direct sampling approach with *MD* simulations (*TN-MD method*, approach three), and finally ending with the *TN* following the direct sampling approach with proton-carrying water motions (approach two). Along the presentation of the individual *TNs*, snapshots of the *MD* simulations are examined with regard to properties relevant for the transfer protons.

The *TN* calculation following the direct sampling approach generates 1 088 initial states and thus the maximal number of initial states possible for this setup, i.e. no states excluded due to steric clashes. Following the *QM* potential energy minimization, 208 distinct nodes remained, between which 4 193 neighbor pairs were identified. Consequently, 4 193 *CPR* calculations were performed to construct the *TN*. Reactant and product state of the overall proton translocation display similar potential energies, which is reasonable, due to their chemical equivalence. The *MBP* connecting the reactant and product state of the overall proton translocation contains a rate-determining, maximal transition barrier of 5 kcal/mol. Four other pathways with equal transition barrier heights can be identified, corresponding to the second, third, fourth, and fifth best pathway. In Fig. 4.3 all pathways up to a rate-determining, maximal transition barrier of 7 kcal/mol are depicted. Interestingly, all these pathways involve a node in which the side chain dihedral angle of the left aspartate-like molecule is rotated by 45° (one step), while the side chain dihedral angle of the right aspartate-like molecule corresponds to the configuration of the reactant state of the overall proton translocation. This state is abbreviated by SC0.SC1=1.0. Furthermore, all best pathways, except one, involve a node in which the side chain dihedral angle of the right aspartate-like molecule is rotated by 45°, while the side chain dihedral angle of the left aspartate-like molecule corresponds to the configuration of the reactant state of the overall proton translocation. This state is abbreviated by SC0.SC1=0.1. Finally, in all steps of the *TN*, depicted here, the water pattern differs, i.e. in every transition at least one of the thirteen water molecules undergoes a translational or rotational displacement as described before.

The *TN* calculation following the direct sampling approach with *MD* simulations, the *TN-MD method*, generates 1088 initial states by the direct sampling and thus the maximal number of initial states possible. These initial states are comple-

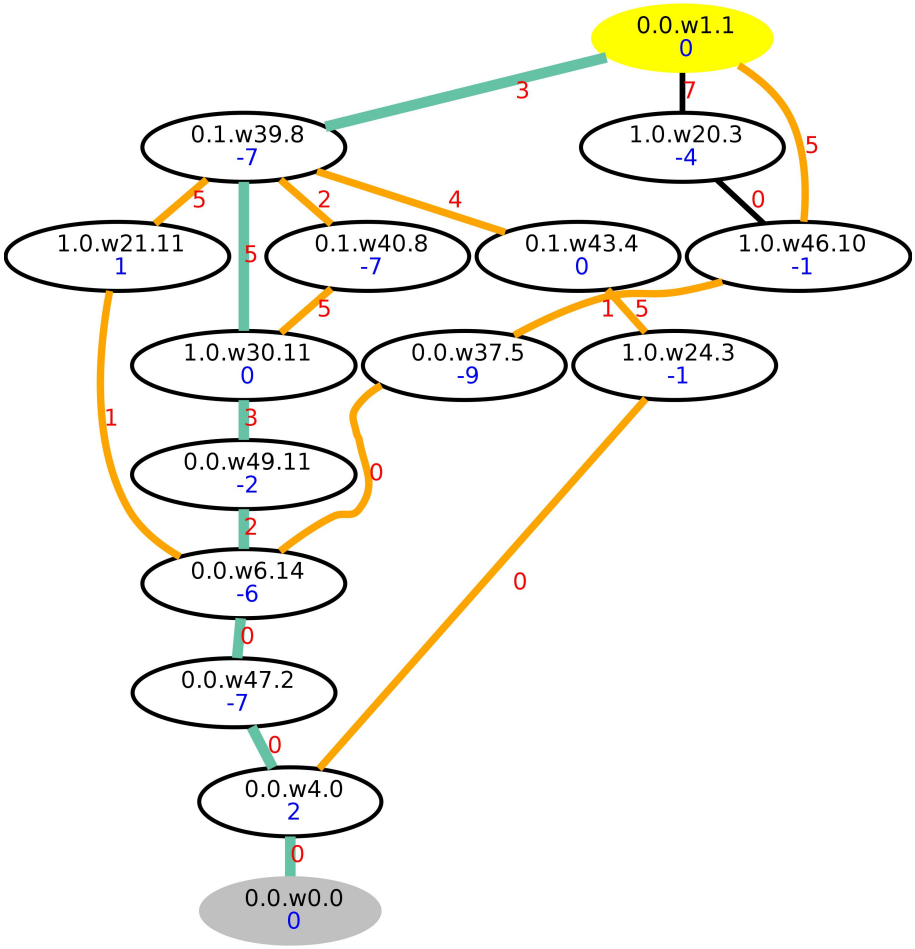


Figure 4.3: *TN* describing the translocation of a proton along the model channel, depicted in Fig 4.1, following the direct sampling approach, i.e. direct sampling of side chain dihedral angles and protonation states. The nodes are shown as ellipses labeled after the full set of *DOFs* used: side chains, water pattern, and proton position (SC0.SC1.wi.P). The ellipses filled in gray and yellow color correspond to the reactant and product state of the overall proton translocation, respectively. Edges are shown as lines connecting two nodes. Blue numbers are the potential energies of the respective nodes, red numbers are the potential energies of the highest transition states along the edges used as edge weight. All energies are in kcal/mol, relative to the reactant state of the overall proton translocation, and rounded to integer values for better readability. The *MBP* is bold and highlighted in turquoise, the 2nd, 3rd, 4th, and 5th best pathways are depicted in orange.

mented by states obtained from 13 *MD* simulations with varying protonation states, i.e. one *MD* simulation per hydronium ion. All *MD* simulations were performed with a time step of 1 fs and were stopped after 10^6 steps, providing 1 ns of *MD* simulation per hydronium ion. To propagate the model channel in time and maintain a constant temperature of 300 K the *Velocity-Verlet integrator* [337] and the *Nose-Hoover thermostat* [447, 448] were used, respectively. Necessary parameters for the *MD* simulations were obtained from the CHARMM force field [449] and

from *Sagnella et al* [450] for the hydronium ion. Consequently, water was treated as TIP3P water [451]. From either trajectory, snapshots were taken in 1 ps intervals, providing 1000 initial states per *MD* simulation, and thus 14088 initial states overall, i.e. no states excluded due to steric clashes.

Preceding the generation of the *TN*, the snapshots of all 13 *MD* simulations were examined with regard to properties relevant for the transfer of protons. This examination is presented in the following:

- Excursion

  The snapshots of all 13 *MD* simulations were subjected to two distinct potential energy minimizations, a *QM* potential energy minimization using the *OM2* method and a *MM* potential energy minimization using the same set of parameters as already used for the *MD* simulations, with a convergence criterion of 0.001 kcal/(molÅ). The potential energies of all minimized states are depicted in Fig 4.5 (bottom left and right) with respect to the distance of the hydronium ion to either aspartate-like molecule. The potential energies are relative, in comparison to the state with lowest potential energy. Interestingly, this state (and other states with marginally higher potential energies) does not correspond to the reactant or product state of the overall proton translocation. Instead, the lowest potential energies are found for states, in which the hydronium ion is hydrogen bonded to either of the aspartate-like molecules.

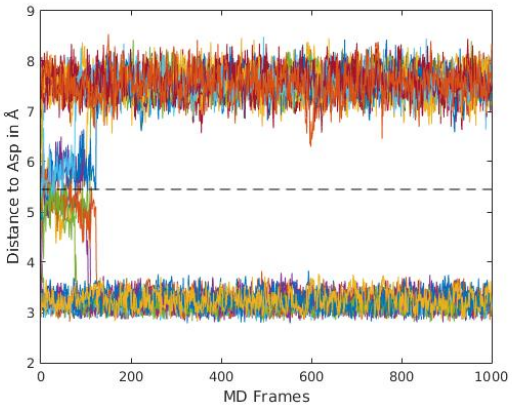


Figure 4.4: Position of the hydronium ions expressed as distance to the central carbon atom of the aspartate-like molecules (Asp). Time series of distances in the *MD* simulations. The dashed line indicates half the distance between the two aspartate-like molecules. The different colors correspond to the individual *MD* runs in which different water molecules are protonated.

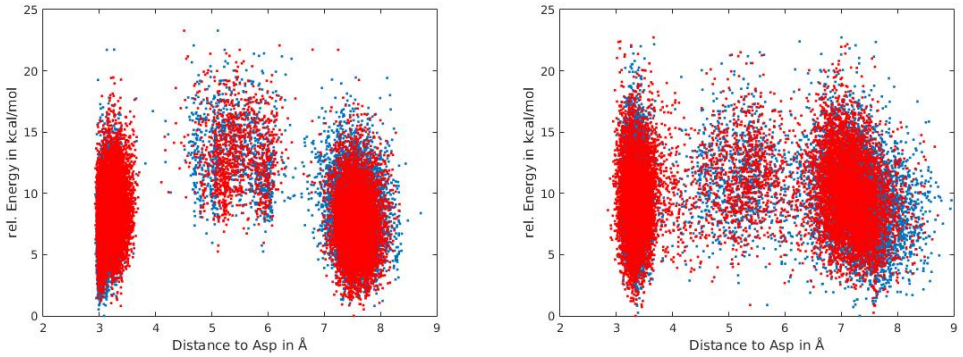


Figure 4.5: Position of the hydronium ions expressed as distance to the central carbon atom of the aspartate-like molecules (Asp). Left: Distances to Asp1 (blue) and Asp2 (red) and corresponding energies after *MM* potential energy minimizations. Right: Distances to Asp1 (blue) and Asp2 (red) and corresponding energies after *QM* potential energy minimizations.

In all thirteen *MD* simulations the hydronium ion is located, for most of the simulation time, close to either of the two aspartate-like molecules. *MD* simulations starting with a short distance between the hydronium ion and the aspartate-like molecules maintain the close proximity, while initially more centered hydronium ions move quickly, within ≈100 ps, towards one of the aspartate-like molecules (cf. Fig 4.4). This corresponds, to the hydronium ion positions following the two distinct potential energy minimizations (cf. Fig 4.5 left and right). Interestingly, however, following the *QM* potential energy minimizations an increased amount of centered hydronium ions, occupying distances and potential energy levels not observed following the *MM* potential energy minimizations, can be detected. The relative potential energy range for centered hydronium ions following the *MM* or *QM* potential energy minimization is ≈8-20 kcal/mol and ≈5-18 kcal/mol. Finally, although hydronium ion positions close to either of the aspartate-like molecules are favored, relative potential energies up to ≈23 kcal/mol are also observed for these configurations. Overall, the *QM* potential energy minimizations increase the number of states sampled by the *MD* simulations, providing states with hydronium ions in all regions of the model channel which is a necessity for an unbiased proton translocation along the channel.

Apart from the positions of the hydronium ion, the hydrogen bond is essential for efficient proton translocations. Therefore, the probability of either hydronium ion to form hydrogen bonds with surrounding water molecules and/or aspartate-like molecules was investigated along the *MD* simulations and following the *MM* or *QM* potential energy minimizations. The results are sum-

marized in Tab 4.1.

Table 4.1: Occurrence of hydrogen bonds formed with the hydronium ion in snapshots of classical *MD* simulations, and following *MM* or *QM* potential energy minimizations. The criterion for a hydrogen bond is a maximal donor acceptor (*DA*) distance of 3 Å and a maximal deviation of 20° from linear for the *DHA* angle.

Number of	Occurrence in proc		
hydrogen bonds	*MD*	Min(*MM*)	Min(*QM*)
0	2.57	1.98	1.98
1	20.97	8.57	8.50
2	46.72	24.89	24.77
3	29.72	64.53	64.73
4	0.02	0.03	0.02

The number of states in which no hydrogen bonds are formed, to and from the hydronium ion, is low (about 2 %). On the other hand, a single hydrogen bond, corresponding to a Zundel ion, is observed in more than 20 % of the states along the *MD* simulations, and in less than 10 % of the states following the potential energy minimizations. States with two and three hydrogen bonds correspond to Eigen-like ions and Eigen ions. Here, the most pronounced difference between the *MD* simulations and the states minimized in potential energy is observed. For the *MD* simulations, almost 50 % of all states correspond to Eigen-like ions, whereas the states minimized in potential energy correspond in almost 65 % to Eigen ions. The difference in distributions between *MM* and *QM* potential energy minimizations is negligible. Overall, the hydrogen bond distributions correspond to the proton transfer mechanism introduced before, where Eigen ions store the hydronium ions and Zundel ions serve as transition states between two Eigen-ion configurations. Consequently, the states generated from *MD* simulations provide sufficiently many hydrogen bonds to allow an unbiased proton translocation along the channel.

Following the Excursion, the snapshots of all 13 *MD* simulations, *QM* minimized in their potential energies, were used, in combination with the states from the direct sampling, to form a set of 7 462 distinct nodes. Between these nodes 134 527 neighbor pairs were identified. Consequently, 134 526 *CPR* calculations were performed to construct the *TN* following the direct sampling approach with *MD* simulations.

From the complete edge set, 47 468 edges (≈ 35 %) were exclusively based on nodes generated by the *MD* simulations. As before reactant and product state of the overall proton translocation display similar potential energies, due to their chemical equivalence. The *MBP* connecting the reactant and product state of the overall proton translocation consists of only three steps and displays a rate-determining, maximal transition barrier of 4 kcal/mol (cf. Fig. 4.6). Along the pathway, changes in all *DOFs*, except an rotation around the side chain dihedral angle of the left aspartate-like molecule, can be observed. The second best pathway, displaying a rate-determining, maximal transition barrier of 5 kcal/mol, corresponds to the *MBP* of the *TN* following the direct sampling approach. This aspect is based on a partial adjustment of the water molecules along the *QM* potential energy minimizations (cf. Fig. 4.5). Still, the nodes from the *MBP* of the *TN* following the direct sampling approach with *MD* simulations would be missed without a sampling of the positions and orientations of all water molecules and the hydronium ion. An evaluation of the TN using *DFT* functionals is presented in Appendix F.

The structures of the nodes along the *MBP* and the second best pathway from the *TN* following the direct sampling approach with *MD* simulations (cf. Fig. 4.6) are depicted in Fig. 4.7. Along both pathways, an Eigen-ion or Eigen-like ion is formed in all intermediate structures. The *MBP* starts with a proton transfer from the left aspartate-like molecule to a water molecule in close proximity. This transfer is associated with a re-orientation of the two water molecules forming hydrogen bonds to the proton accepting water molecule. Furthermore, a rotation around the side chain dihedral angle of the right aspartate-like molecule is observed along the proton translocation. In a second step, further re-orientations of two water molecules and again a rotation around the side chain dihedral angle of the right aspartate-like molecule, establishes an *HBC* through the entire model channel. The proton transfer to the product state of the overall proton translocation is concluded by a concerted proton transfer step along three hydrogen bonds. The subtle changes in the positions and orientations of certain water molecules, enabling the final transition over the rate-determining, maximal transition barrier, are those obtained by the *MD* sampling. According to *Hassanali et al* [452], the pre-orientation of the side chains of the aspartate-like molecules, as well as the pre-orientation of the water molecules, can be understood as the resting phase of the proton translocation, while the actual, concerted proton transfer along the *HBC* represents the active state.

The second best pathway in Fig. 4.7 starts with a re-orientation of all thir-

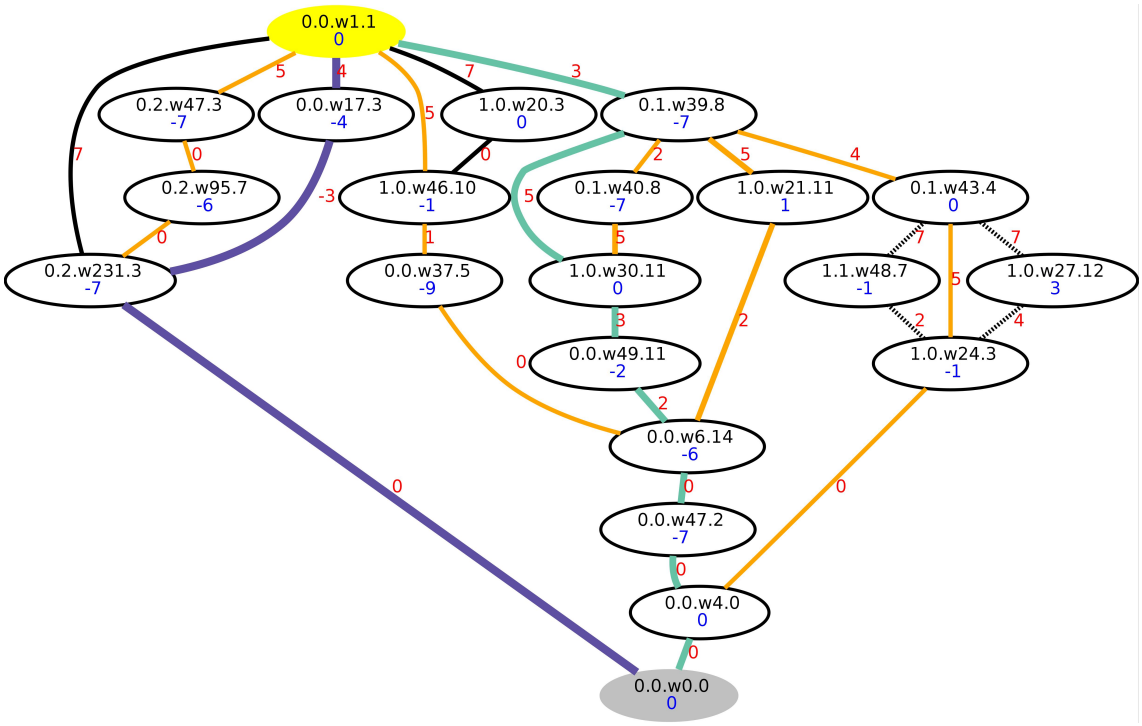


Figure 4.6: *TN* describing the translocation of a proton along the model channel, depicted in Fig 4.1, following the direct sampling approach with *MD* simulations, i.e. direct sampling of side chain dihedral angles and protonation states and *MD* simulations for the water motion, and the direct sampling approach with hydronium ion motions, i.e. direct sampling of side chain dihedral angles, protonation states, and the motion of the hydronium ions. The nodes are shown as ellipses, labeled after the full set of *DOFs* used: side chain dihedral angles, water pattern, and proton positions (SC0.SC1.wi.P). The ellipses filled in gray and yellow color correspond to the reactant and product state of the overall proton translocation, respectively. Edges are shown as lines connecting two nodes. Blue numbers are the potential energies of the respective nodes, red numbers are the potential energies of the highest transition states along that edges used as edge weight. All energies are in kcal/mol, relative to the reactant state, and rounded to integer values for better readability. The *MBP* is bold and highlighted in indigo, the 2nd best pathway is bold and highlighted in turquoise. In principle, the turquoise pathway displays the same rate-determining, maximal transition barrier as the best pathways shown in orange. It is, however, highlighted because it is the same pathway as the *MBP* from the *TN* following the direct sampling approach (cf Fig. 4.3). The dashed lines indicate additional edges and pathways, derived from a *TN* calculation following the direct sampling approach with hydronium ion motions.

teen water molecules. This re-orientation is followed by three individual proton transfer steps, associated with re-orientations of the proton-donating and proton-accepting water molecules, as well as two or three additional water molecules which are hydrogen bonded to them. In the last three steps, a hydronium ion is already hydrogen-bonded to the right aspartate-like molecule. Here, however, it is the rota-

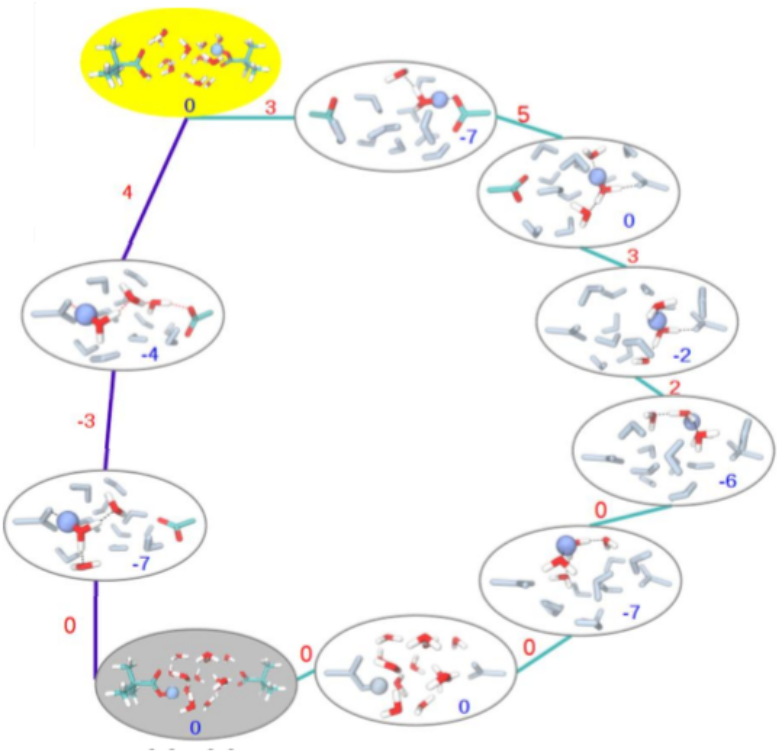


Figure 4.7: Structures of the nodes, intermediate states and end states, along the two best pathways from the *TN* following the direct sampling approach with *MD* simulations and the direct sampling approach with hydronium ion motions depicted in Fig. 4.6. Blue numbers are the potential energies of the respective nodes, red numbers are the potential energies of the highest transition states along the edges used as edge weight. All energies are in kcal/mol, relative to the reactant state of the overall proton translocation, and rounded to integer values for better readability. The *MBP* is highlighted in indigo, the 2nd best pathway in turquoise, respectively. For clarity, the bulky ends of the aspartate-like molecules are not depicted in the intermediate states. The part of the system that changes with respect to the previous state is highlighted in color. Big, blue spheres represent protons to be transferred or just being transferred. Colored water molecules indicate their change in position and/or orientation, colored side chains indicate a rotation around the side chain dihedral angles.

tion around the side chain dihedral angles of both aspartate-like molecules and the re-orientation of further water molecules, which precedes the proton transfer to the product state of the overall proton translocation. Interestingly, the entire second best pathway is contained in the *TN* following the direct sampling approach. This is especially notable for the first intermediate state, which differs from the reactant state of the overall proton translocation by the water orientation only. Since water re-orientations have not been sampled in this *TN*, the state in question, 0.0.w4.0, is a consequence of the *QM* potential energy minimization, leading away from the reactant state of the overall proton translocation. This fact underlines the importance

of an assignment procedure considering all *DOFs* and not just the *DOFs* sampled initially. In the present case, a limited assignment would be free of consequence, due to the fact that the transition barriers to and from the state, 0.0.w4.0, as well as the potential energy of the state itself, are all equal to the energy of the reactant state of the overall proton translocation. In other, especially more complex reactions, however, this may not be the case.

A third *TN* was calculated, following the direct sampling approach with hydronium ion motions, to check the reproducibility of the best pathways presented in Fig. 4.6. This setup generated 6 080 initial states and thus the maximal number of initial states possible for this setup, i.e. no states excluded due to steric clashes. Following the *QM* potential energy minimizations, 1 230 distinct nodes remained, between which 42 692 neighbor pairs were identified. Consequently, 42 692 CPR calculations were performed to construct the *TN*. As before reactant and product state of the overall proton translocation display similar potential energies, due to their chemical equivalence. The comparison of all best pathways up to a rate-determining, maximal transition barrier of 7 kcal/mol, shows an almost complete reproduction of the best pathways from the *TN* calculation following the direct sampling approach with *MD* simulations except for two pathways with rate-determining, maximal transition barriers of 7 kcal/mol (cf. Fig. 4.6).

Table 4.2: Number of *MBPs* within a potential energy range of 10 kcal/mol for the *TN* following the direct sampling approach (approach one), the *TN* following the direct sampling approach with hydronium ion motions (approach two), and the *TN* following the direct sampling approach with *MD* simulations (approach three).

	Number of Pathways		
Barrier (kcal/mol)	Approach 1	Approach 2	Approach 3
1	0	0	0
2	0	0	0
3	0	0	0
4	0	1	1
5	5	6	6
6	0	0	0
7	1	4	2
8	3	11	9
9	2	10	7
10	1	10	9
Up to 10	12	42	34

In Tab. 4.2 the number of best pathways up to a rate-determining, maximal transition barrier of 10 kcal/mol is summarized for all three *TNs* calculated. The number of best pathways grows considerably, when comparing the *TN* following the direct sampling approach of side chain dihedral angles and protonation states with the *TN* following the direct sampling approach of side chain dihedral angles and protonation states and *MD* simulations for the water motion. Most of the pathways identified are energetically similar but differ remarkably in their structural changes. Especially, the two best proton transfer pathways identified represent two distinct reaction mechanisms, i.e. a concerted proton transfer mechanism and a step wise proton transfer mechanism, which may not be identified by techniques with predefined reaction coordinates.

In conclusion, the *TN-MD method* provides a comprehensive and efficient way to capture all *DOFs* that are relevant in the proton transfer process. The best pathways determined display an interplay of all these *DOFs*. The optimal proton transfer pathway displays a concerted proton transfer, while the second best proton transfer pathway is step wise. In addition, several other pathways with similar energies were determined, which could not be determined by the direct sampling of side chain dihedral angles and protonation states only. The identification of ensembles of pathways will be even more important in complex systems with a huge variety of energetically feasible, yet mechanistically diverse, proton transfer pathways. The *TN-MD method* is an efficient way to capture this complexity and variability.

In some cases, however, a direct sampling of the positions of a few water molecules might still be necessary to achieve hydration states which are not sampled along the *MD* simulations and thus ignored in the *TN* calculation. These limitations need to be considered and dealt with. If done appropriately, the *TN-MD method*, is a powerful, bias-free method operating at the interface of the discrete and continuous world.

4.2 The TN Prediction Method

The *TN-MD method*, as introduced before, leads to a significant reduction of states and *MEP* calculations within the *TN* approach, while still providing an accurate description of the reaction of interest. In many processes of biological relevance, however, even this reduced number of initial states, nodes and *CPR* calculations is too high to be sampled within a finite amount of time. Therefore, it is, in most cases,

simply not possible to directly include all *DOFs* which might play a role in such processes. Consequently, these excluded or unsampled *DOFs*, e.g. a protonation state or the catalytic state of the active site in *CcO*, are more or less fixed along the *TN* calculations. To achieve a comprehensive description of the process of interest nonetheless, these unsampled *DOFs* need to be included indirectly.

Imagine for example a system containing (only) five amino acids with side chain dihedral angles χ_1 and χ_2. A pure investigation of these ten dihedral angles in 45° steps alone involves a sampling of $\approx 10^9$ initial states. A further inclusion of the protonation state increases this number 10-fold, when not considering the presence of any hydronium ions and much more if considering the presence of hydronium ions, while the inclusion of a further amino acid increases the number of initial states to investigate 8- to 390 625-fold, depending on the number of side chain dihedral angles involved. And remember, all this sampling does not include the motion of any water molecule.

A set of $\approx 10^9$ initial states results in a *TN* with $\approx 5 \cdot 10^{17}$ edges, when considering that all initial states end up as a node and the *TN* is complete. Imagine for once that $\approx 10^9$ *QM* potential energy minimizations and $\approx 5 \cdot 10^{17}$ *CPR* calculations, represent the maximal amount of calculations doable within a finite amount of time. (At this point, it should be emphasized that, apart from the little gedankenexperiment, these numbers of *QM* potential energy minimizations and *CPR* calculation pose insurmountable problems in reality when considering aspects like calculation costs and storage requirements and thus the *TNs* in reality are much smaller.) Then, the imagined *TN* is not posing any problem. Now, however, the inclusion of a further *DOF* is required, namely, the protonation state of an additional amino acid, which is either protonated or deprotonated. This additional sampling doubles the amount of *QM* potential energy minimizations and increases the amount of *CPR* calculations 4-fold. Thereby, both values are increased well above the maximal amount of doable calculations. Hence, an indirect inclusion of the protonation state is required, giving, at the end, two *TNs* (with protonated and deprotonated amino acid) with $\approx 10^9$ *QM* potential energy minimizations and $\approx 5 \cdot 10^{17}$ *CPR* calculations each. Thereby, both *TNs* calculations are doable within a finite amount of time. Coming back to reality this means, that for each unsampled *DOF* of interest a cost intensive re-calculation of the entire *TN* is required. This is especially problematic for the *CPR* calculations, which represent the rate-limiting factor of such *TN* calculations.

The problem of cost intensive, complete *TN* re-calculations is addressed by the *TN prediction method* with a shift in perspective, because, in principle, the intro-

duction of a proton in the imagined system (or any other change in the unsampled *DOFs*) is nothing else than a perturbation of the *TN* for the deprotonated case (or the *TN* in which the unsampled *DOFs* are not changed). Therefore, it can be expected that, obviously depending on the severity of the perturbation, the unperturbed and perturbed *TNs* share common features, which are conserved over the perturbation and consequently do not require re-calculations in the perturbed systems. Hence, the identification of such conserved features is a central aspect in reducing the calculation costs of perturbed *TNs*. A further aspect of interest is, that in many situations individual properties of the perturbed *TNs*, e.g. the maximal transition barrier of the *MBP* and the states visited along the *MBP*, are sufficient to decide if a more detailed analysis of the perturbed systems is required. The *TN prediction method* makes use of both of these aspects, using a coarse-grained representation of the unperturbed *TNs*, along with the unperturbed *MSTs* and corresponding sensitivity analyses, as conserved features over the perturbation and an additional coarse-graining of the perturbed *TNs* along their re-calculations to focus their re-calculations on the determination of important, perturbed network properties only.

In the following subchapters the *TN prediction method* is introduced in detail and tested extensively. Therefore, the proton transfer through a small, water-filled model channel, with an unsampled *DOF* in close proximity, is analyzed for an unperturbed system, i.e. a system containing the additional point charge in its initial configuration in terms of charge and position, and various perturbed systems, i.e. systems containing the additional point charge with decreased or increased charge values or an altered position, using the *TN prediction method* and the standard *TN* approach for comparison. Thereby, the computational costs of the predictions of the perturbed *TNs* (in contrast to the costs of a complete calculation of the perturbed *TNs*) and the accuracy of the predicted properties were of main interest. The entire source code of an implementation of the *TN prediction method* can be found in Appendix H.

In principle, other methods providing similar prediction results are conceivable, e.g. using robust analyses if the perturbation severity can be quantified in terms of edge weight intervals and if the network topologies do not change along the perturbation. Both requirements however, are not fulfilled in the *TN* perturbations presented in the following. Therefore, *TN* predictions based on robust analyses are not discussed further.

4.2.1 Background, Methods and Settings II

The *TN prediction method* was used to investigate the proton transfer through a small model channel, resembling the model channel introduced in subchapter 4.1.1 with an additional point charge in close proximity. Consequently, the reader is referred to subchapter 4.1.1 for a detailed description of the model channel. Here, the placement of the additional point charge is described along with its different configurations used to mimic *TN* perturbations (cf. Fig. 4.8). To obtain the position of the additional point charge, a triangular prism, spanned by the carbon atoms of the six methyl groups from the two t-butyl structures, was imagined. Thereby, a plane containing the central carbon atoms of both t-butyl structures, and thus the cylindrical axis, was defined parallel to one of the rectangular sides of the triangular prism. The initial position of the additional point charge was then defined by $\boldsymbol{d}$, a vector orthogonal to the plane with $|\boldsymbol{d}| = 6$ Å and the midpoint of the cylindrical axis as starting point (cf. Fig 4.8 top). Value and position of the additional point charge were constant within the individual *TN* calculations, which corresponds to one *TN* per point charge position or value. The initial charge of the point charge was set to $q = 0.050$. Other configurations of the additional point charge were achieved by circular translocations around the initial position with $|\boldsymbol{r}| = 0.5$ Å, 1.0 Å, or 2.0 Å and $\phi = 0°$, 45°, 90°, 135°, 180°, 225°, 270°, or 315°, parallel to the plane containing the cylindrical axis, charge decreases down to $q = 0.000$ (actually $q = 0.000001$ to keep the point charge included in the *QM/MM* setup) in steps of $\Delta q = 0.010$ or charge increases up to $q = 0.100$ in steps of $\Delta q = 0.001$ from $q = 0.050$ to $q = 0.060$, steps of $\Delta q = 0.002$ from $q = 0.060$ to $q = 0.070$, and steps of $\Delta q = 0.005$ from $q = 0.070$ to $q = 0.100$. Thereby, 51 configurations of the additional point charge were investigated overall.

In a first step, complete *TN* calculations were performed for the model system with all 51 configurations of the additional point charge. Therefore, two states, containing the excess proton at either of the two carboxyl groups, were generated and subsequently used as reactant and product state of the overall proton translocation. To construct the *TN*, different protonation states and side chain dihedral angle conformations of the carboxylated t-butyl structures were sampled by placing the excess proton on either of the water molecules or on one of the carboxyl groups and by rotating the side chain dihedral angles in 45° steps. This sampling approach corresponds to the direct sampling approach (approach one) introduced in subchapter 4.1.1. Overall, 1 088 initial states were generated for either *TN*, which corresponds to the maximal number of initial states possible under this setup. The

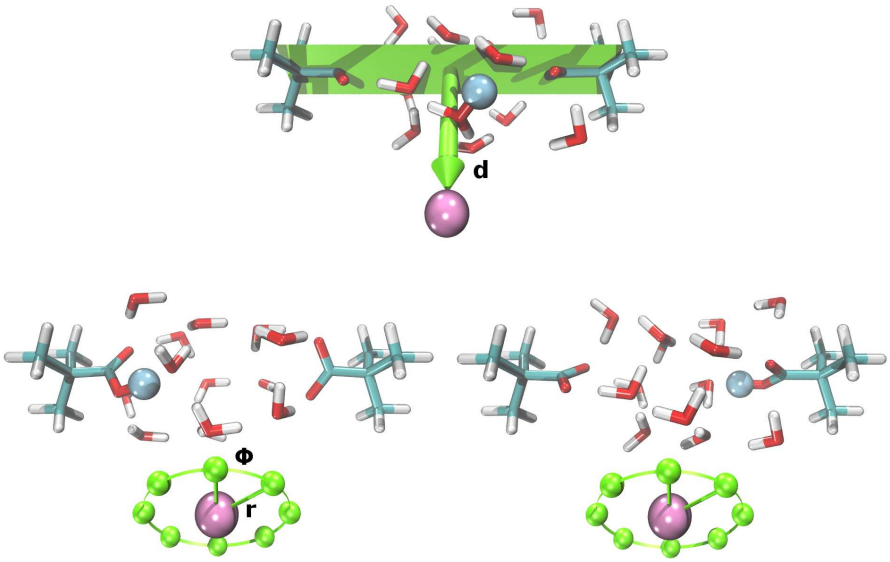


Figure 4.8: Model system for the calculation of proton transfer pathways with an additional point charge (pink sphere). Top: plane containing the cylindrical axis and the orthogonal vector $\boldsymbol{d}$ ($|\boldsymbol{d}|$ = 6 Å) to locate the initial position of the additional point charge, bottom left: reactant state, bottom right: product state. The blue sphere highlights an excess proton located on the left or right carboxyl group corresponding to the reactant or product state of the overall proton translocation. Circular translocations of the additional point charge, parallel to the plane with $|\boldsymbol{r}|$= 0.5 Å, 1.0 Å, or 2.0 Å and ϕ = 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°, are indicated by green spheres. Charge increases or decreases are not depicted.

sets of initial states were then subjected to *QM/MM*, potential energy minimizations according to the *ABNR* minimization algorithm and a convergence criterion of 0.001 kcal/(molÅ). Here, the water-filled cylinder and the additional point charge represented the *QM* and *MM* part, respectively. The *QM* part was treated with the *semi-empiric* quantum method *OM2* [244, 245] while the *MM* part was treated by the CHARMM force field [449]. Both parts were coupled electrostatically, which allowed the *MM* point charge to polarize the *QM* electron density. Furthermore, van der Waals interactions between both parts were modeled by the CHARMM force field. Following the minimizations all states were classified with respect to the reactant state of the overall proton translocation, regarding their side chain dihedral angles, protonation state, and water pattern, which gave the nodes of the *TN*. Side chain dihedral angles and protonation states were assigned according to the initial sampling. The motion of the water molecules was not sampled initially. Still, the pattern of all 13 water molecules, defined by their positions and orientations, was included in the state assignment. Two water patterns were considered to be unique, if at least one water molecule is translocated by 2.0 Å in the x-, y-, or z-direction or rotated by 22.5° around the x-, y-, or z-axis. To avoid chemical

equal states, which only differ in their atom labels an excessive atom re-labeling with respect to the reactant state of the overall proton translocation was performed using the Hungarian algorithm [335]. Following the determination of the node set, transitions between pairs of nodes, i.e. nodes which differed in each *DOFs* by at most one step, were calculated using the *CPR* algorithm. This provided the edges of the *TNs*, for which the potential energy of the rate-determining, maximal transition barrier along the respective *CPR* calculations with respect to the potential energy of the reactant state of the overall proton transfer reaction, was used as edge weight. Nodes and edges were then combined to form a weighted, simple graph. Finally, the *MBP* connecting the reactant and product state of the overall proton translocation was computed using Dijsktra's algorithm. The rate-determining, maximal transition barrier along this pathway is abbreviated with ω^* in the following.

In a second step, 50 *TN* predictions were performed, using the previously determined complete *TN* with the additional point charge in its initial configuration, termed here the initial or unperturbed *TN*, as starting point for the *TN* predictions of all other configurations of the additional point charge, termed here the perturbed *TNs*. To characterize the initial *TN,* its *MST* was determined using Kruskal's algorithm. The *MST* provides the *MBPs* between all pairs of nodes and is unique for distinct edge weights. In principle, every non-*MST* edge could be part of the *MST* if its weight is reduced appropriately. Two classes of non-*MST* edges may be distinguished, those for which a small weight reduction is sufficient to alter the *MST* and those for which a high weight reduction is required to alter the *MST*. Consequently, the classes of edges are named high-sensitive and low-sensitive edges, respectively (cf. Fig 4.9a). The sensitivity of all non-*MST* edges was calculated which allowed an ordering of the non-*MST* edges from high to low sensitivities. Later on, the *MST* of the initial *TN* was used as initial guess for the *MST* or *MBP* of the perturbed *TNs*, whereas the ordered initial non-*MST* edge lists were used for the refinement of the initial guess towards the perturbed *MST* or *MBP*.

Following the determination of the initial *MST* and the ordered initial non-*MST* edge list, the initial *TN* was coarse-grained (initial coarse-graining). In contrast to other coarse-graining techniques, the coarse-graining is based on transition barriers, using ω^* as upper bound, instead of a usual structure-related clustering [453]. For the initial coarse-graining, the eigenvectors of the *Laplacian* matrix $\mathbf{L}$, corresponding to the initial *TN*, were determined:

$$\mathbf{L} = \mathbf{D} - \mathbf{A}^*, \tag{4.3}$$

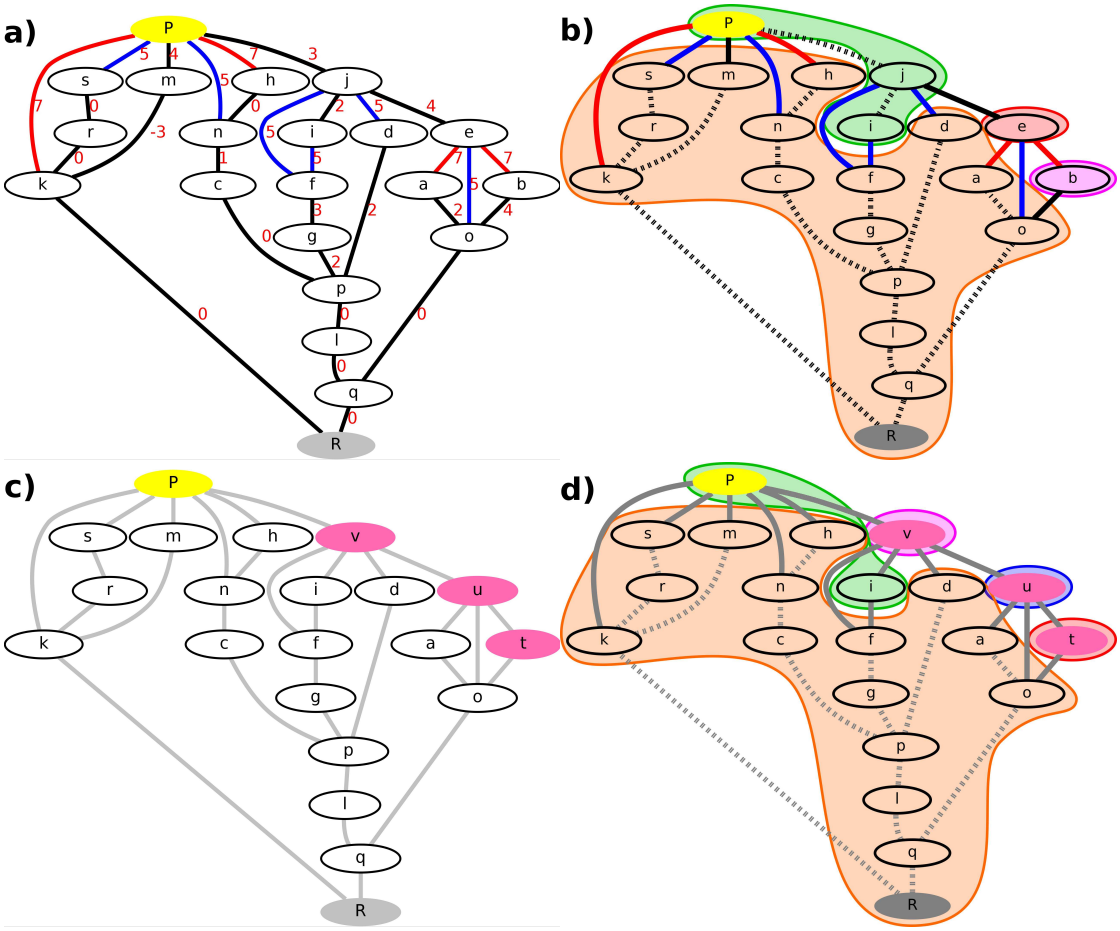


Figure 4.9: a) Initial *TN* containing 21 nodes and 29 edges, corresponding to the *TN* in Fig. 4.6. Nodes are shown as ellipses, edges are shown as lines. The reactant state is highlighted in gray, the product state in yellow, labeled R and P. Intermediate nodes are labeled from a to s. Red numbers represent the potential energies of the maximal transition states along the edges used as edge weight. All energies are in kcal/mol, relative to the potential energy of the reactant state of the overall proton translocation and rounded to integer values. Edges in black represent the *MST*, edges in blue represent high sensitive non-*MST* edges and edges in red represent low sensitive non-*MST* edges. b) Coarse-grained representation of the initial *TN* containing 4 coarse-grained nodes, represented by the orange, green, magenta, and red shaded areas, and 29 edges. Dashed lines represent negligible edges connecting nodes within the same coarse-grained node, solid lines represent edges connecting nodes within distinct coarse-grained nodes. Edge weights are not shown for a better visualization. c) Perturbed *TN* containing 21 nodes and 29 edges. The nodes b, e, and j are replaced by the nodes t, u, and v, highlighted in pink. Gray lines indicate edges to be calculated in an all-node *TN*. d) Coarse-grained representation of the perturbed *TN* based on the initial coarse-graining of the initial *TN* containing 5 coarse-grained nodes, represented by the orange, green, magenta, blue, and red shaded areas, and 29 edges. Dashed lines represent negligible edges, solid lines represent edges to be calculated.

where **D** is a diagonal matrix containing the degrees of all nodes of the initial *TN* and $\mathbf{A}^*$ is a special adjacency matrix of the initial *TN* with

$$a_{ij} = \begin{cases} 0 & \text{if } \omega_{ij} > \omega^* \\ 1 & \text{if } \omega_{ij} < \omega^* \end{cases} \tag{4.4}$$

to determine the connected components and thus the coarse-grained nodes. Due to the initial coarse-graining the initial *MBP* is, for example, reduced to a pathway with two coarse-grained nodes (each of them containing at least one of the original nodes) connected by an edge with ω^* as edge weight (cf. Fig. 4.9b). These coarse-grained nodes, however, should not be understood as single representative conformations, like an average conformation of the original nodes or the original node most similar to the average conformation, because such representative conformations are in most cases no local minima conformations, which is a requirement of the *CPR* algorithm, or they suffer from the initial guess-pathway problem. Instead, they should be interpreted as sets of original nodes between which barrier-free transitions, at least compared to ω^*, are possible. Thus, the initial coarse-graining step is not reducing the number of nodes to be investigated, but the number of edges. In other words, all edges connecting nodes within the same coarse-grained node can be excluded from further investigations. Thereby, the initial coarse-graining reduces the edge sets representing the initial *MST* and the ordered, initial non-*MST* edge lists to those edges connecting distinct coarse-grained nodes only. This concludes the analysis of the initial *TN*.

To start the predictions of the perturbed *TNs*, the sampling of initial states, the potential energy minimizations, and the node pairing is redone in the perturbed systems according to the same criteria of changes in the *DOFs* as for the initial *TN*. Thereby, a comprehensive picture of the topology of the perturbed *TNs* is achieved (cf. Fig 4.9c), meaning that all nodes of the perturbed node lists are calculated and all edges of the perturbed edge lists, which need to be calculated, are identified. Assuming that the coarse-grained nodes of the initial *TN* represent a conserved feature along the perturbations, all perturbed edges, connecting nodes within the same initial coarse-grained nodes, can be removed from the perturbed edges lists (cf. Fig 4.9d). Subsequently, the coarse-grained initial *MST* and the coarse-grained ordered initial non-*MST* edge list were adjusted to the perturbed *TNs*. This means, that all edges present in the initial edge list but absent in the perturbed edge lists were removed from the edge lists representing the coarse-grained initial *MST* and the coarse-grained ordered initial non-*MST* edge lists. On the other hand, edges

present in the perturbed edge lists but absent in the initial edge list were added to the edge list representing the coarse-grained ordered initial non-*MST* edge list. In order to achieve a proper integration of these edges into the sensitivity ranking of the coarse-grained ordered initial non-*MST* edge list the inverse of the maximal transition barrier of the *MBP* connecting the nodes in the initial *TN* was used as sensitivity value for these new edges. For edges connecting nodes which are not present in the initial node list, the sensitivity value was set to zero. If several edges were assigned the same sensitivity value they were ordered randomly. Thus, the coarse-grained initial non-*MST* edge list is ordered from high to low sensitivity values with a non-deterministic ordering of edges with equal sensitivity values.

Finally, the remaining coarse-grained non-*MST* edges were calculated. In few cases, theses calculations were already sufficient to obtain the coarse-grained *MST* or coarse-grained *MBP* of the perturbed *TNs*. If that was not the cases additional non-*MST* edges were calculated from the coarse-grained initial non-*MST* edge list in the order from high to low sensitivity values. The calculation of these additional edges allowed a further coarse-graining of the perturbed *TNs* ("on-the-fly" coarse-graining), in which coarse-grained nodes are merged if connected by an edge with edge weight below ω^*, thus reducing the coarse-grained initial non-*MST* edge list even further along the calculation of the perturbed *TNs*. The *TN* predictions were stopped once the coarse-grained *MSTs* or coarse-grained *MBPs* of the perturbed *TNs* resembled the results from the complete *TN* calculations or once no further edges were left to calculate. In Fig. 4.10 the *TN* prediction method is summarized in a flow chart.

In a third and final step, the *TN prediction method* was tested for an extended set of perturbations using each of the complete *TNs*, determined previously, as initial *TN* for the prediction of the *MSTs* or *MBPs* of all other *TNs*. Overall, this setup yields 702 and 600 *TN* predictions for the increase or decrease of the additional point charge or the translocation of the initial point charge around its initial position, respectively.

All potential energy minimizations and *CPR* calculations were performed with the CHARMM program [444] interfaced to MNDO [445]. The generation of the initial states, the node assignment, the neighbor search as well as the compilation and analysis of the *TNs* were performed by existing source code [325] and libraries from Noe et al [446]. The determination of the *MSTs* and the sensitivity analysis, the initial coarse-graining, and the "on-the-fly" coarse-graining during the *TN* prediction were performed with the Python progam presented in Appendix H.

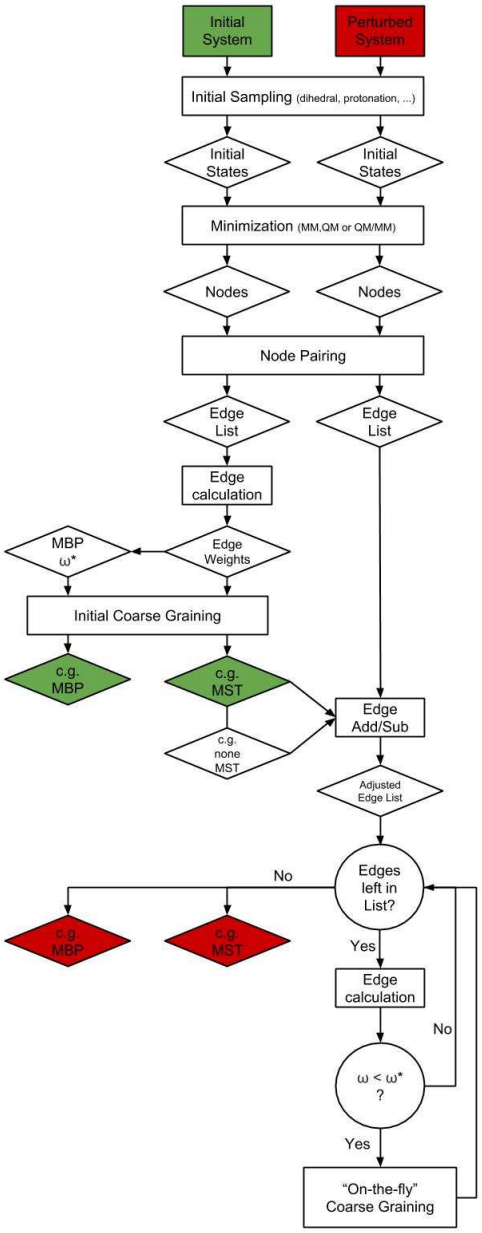


Figure 4.10: Flow chart representing the *TN* prediction method starting from an initial *TN* calculation. (Abbreviations are explained in the text.)

4.2.2 Results and Discussion II

In a first step 51 complete *TN* calculation were performed. These are:

1. The initial *TN* calculation.

2. 26 perturbed *TN* calculations with an increased or decreased value of the additional point charge.

3. 24 perturbed *TN* calculations with point charge translocations around the

initial position.

Here, only the results for the initial *TN* are presented in detail. The results for all other *TNs* are summarized in *Reidelbach et al* [443]. The calculation of the initial *TN* resulted in a network of 252 nodes and 20 316 edges. In comparison to the *TN* involving a direct sampling of the side chain dihedral angles and the protonation states without an additional point charge from subchapter 4.1.2, the number of nodes, and especially the number of edges, is increased. Possible explanations might be an ordering of the water molecules in similar water patterns, due to the additional point charge, sampling differences, due to different reactant and product states of the overall proton translocations, methodical differences, due to the calculation in *QM* only or *QM/MM* systems, or a combination of all these aspects. The *MBP* connecting the reactant and product state of the overall proton translocation in the initial *TN* contains three intermediate nodes and a rate-determining, maximal transition barrier of 5 kcal/mol. This pathway involves the re-arrangement of the side chain dihedral angles of both aspartate-like molecules, the protonation state, and the water pattern. It needs to be noted, that the change of the side chain dihedral angle of the right aspartate-like molecule, as well as the change in the protonation state, is not gradually. Instead back transition to previous states are observed (cf. Fig 4.11). The second and third best proton transfer pathways display rate-determining, maximal transition barriers of 8 kcal/mol and 10 kcal/mol, respectively. The *TN* following the direct sampling approach without an additional point charge in subchapter 4.1.2 displays the same rate-determining, maximal transition barrier of the *MBP*. The intermediate states involved, however, differ, due to the fact that the conformation of the reactant and product state of the overall proton transfer reaction differ. Still, some structural elements are conserved.

The properties of the *MBP* are affected by the increase, decrease, and translocation of the additional point charge. The rate-determining, maximal transition barrier of the *MBP* varies for increases and decreases of the additional point charge between 3 kcal/mol and 9 kcal/mol with an average rate-determining, maximal transition barrier of the *MBPs* of 5 ± 1.5 kcal/mol (5 ± 1.5 kcal/mol for the translocation of the additional point charge), while the number of intermediate states varies between 1 and 7 nodes with an average number of intermediate nodes of 5 ± 1 (5 ± 1 for the translocation of the additional point charge). Additionally, the general topology of the *TNs* is affected to a large extent. For translocations of the additional point charge 33 % to 69 % of the nodes and 52 % to 87 % of the edges present in the initial *TN* do not exist in the perturbed *TNs* (41 % to 60 % of the nodes and 58 %

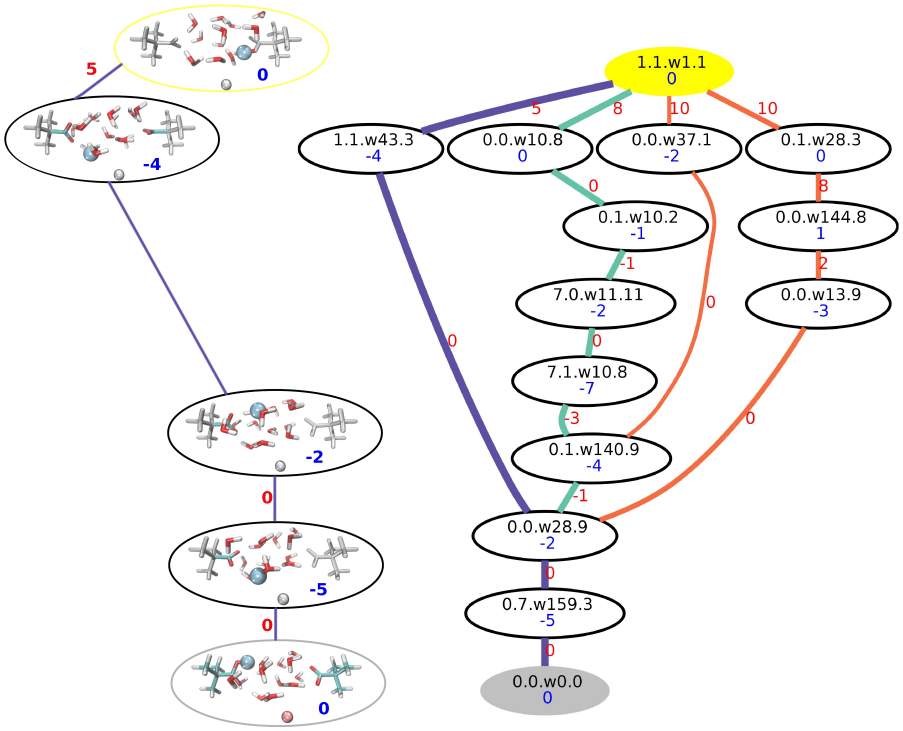


Figure 4.11: Right: Initial *TN* of a 13 water model system with an additional point charge from a complete *TN* calculation following the direct sampling approach, i.e. direct sampling of side chain dihedral angle rotations and protonation states. The nodes are shown as ellipses, labeled according to the side chain dihedral angles of the carboxylated t-butyl structures, the water pattern, and the protonation state (SC0.SC1.wi.P). The reactant state is shown in gray, the product in yellow. Edges are shown as lines. Blue numbers represent the potential energies of the nodes, red numbers the potential energies of the rate-determining, maximal transition barriers along the edges used as edge weight. All energies are in kcal/mol, relative to the reactant state of the overall proton translocation and rounded to integer values. The *MBP* is highlighted in indigo, the next best pathways are shown in turquoise and orange. Left: Detailed representation of the *MBP*. The reactant state is shown in gray, the product state in yellow. Edges are shown as lines. Blue numbers represent the potential energies of the nodes, red numbers the potential energies of the rate-determining, maximal transition barriers along the edges used as edge weight. All energies are in kcal/mol, relative to the reactant state of the overall proton translocation and rounded to integer values. Colored structure elements indicate structural changes along the transition according to the assignment of the *TN* calculation.

to 80 % of the edges for increases or decreases of the additional point charge). On the other hand, 30 % to 70 % of the nodes and 50 % to 87 % of the edges present in the perturbed *TNs* following the translocation of the additional point charge do not exist in the initial *TN* (38 % to 68 % of the nodes and 56 % to 82 % of the edges for increases or decreases of the additional point charge). Thus, in most cases the

perturbed *TNs* contain more "unknown" topological features than "known". Hence, for an efficient determination of the coarse-grained perturbed *MSTs* or *MBPs* the information provided by the initial *TN* is in most cases not sufficient. Consequently, further information about the perturbed *TNs* need to be acquired while they are predicted. This fact is acknowledged in the *TN* prediction method by the "on-the-fly" coarse-graining.

Following the calculation and analysis of the complete *TNs*, the initial *TN* was used to predict the coarse-grained *MSTs* or *MBPs* of all 50 perturbed *TNs*. In principle, the *TN* prediction method is composed of three distinct aspects. These are:

1. The usage of the *MST* of the initial *TN* as initial guess for the *MST* of the perturbed *TNs* followed by an ordered non-*MST* edge calculation according to pre-determined edge sensitivity values.

2. The initial coarse-graining using information from the initial *TN*.

3. The "on-the-fly" coarse-graining using information from the previous edge calculations for the next edge calculations.

In Fig. 4.12 the cost-reductive effect of either of the aspects of the *TN* prediction method is displayed by comparing the number of perturbed edges calculated along the *TN* predictions with the size of the complete edges lists of the perturbed *TNs*. Therefore, *TN* predictions using the *MST* and its sensitivity, the *MST*, its sensitivity, and the initial coarse-graining, as well as the complete method using the previous steps and the "on-the-fly" coarse-graining were performed. To study the effect of the non-deterministically ordered edges (due to similar sensitivity values) within the all-node or coarse-grained ordered non-*MST* edge lists all predictions were performed 1000 times per perturbation scenario. The *TN* predictions were stopped once the coarse-grained *MSTs* or coarse-grained *MBPs* of the predicted, perturbed *TNs* resembled the results from the complete *TN* calculations or, alternatively, once no further edges were left to calculate.

The prediction of the perturbed *MSTs* required average calculations of 99 %, 71 %, or 35 % of the non-*MST* edge lists for increases or decreases of the additional point charge and 97 %, 75 %, or 40 % for the translocation of the additional point charge around its initial position, using the *MST* and its sensitivity, the *MST*, its sensitivity, and the initial coarse-graining, or the complete method using the previous steps and the "on-the-fly" coarse-graining, respectively. For both types of perturbation the calculation costs do not depend on the severity of the perturbation,

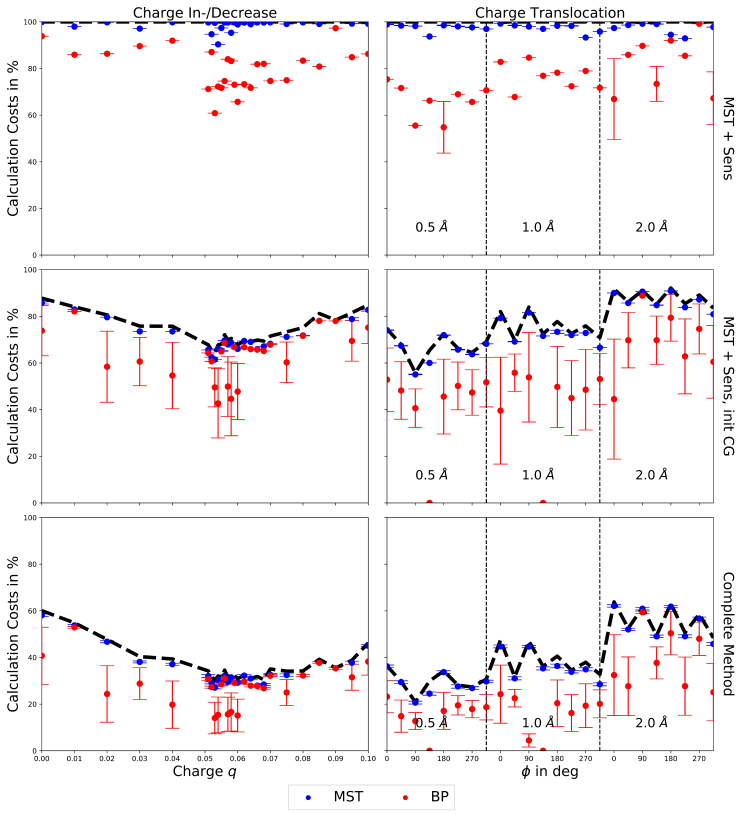


Figure 4.12: Average calculation costs (in % of the average length of the ordered non-*MST* edge lists without coarse-graining) for the (all-node or coarse-grained) prediction of the perturbed *MST* (blue) and *MBP* (red) from 1000 prediction runs per perturbation. Perturbations: charge increase or decrease (left column), charge translocation (right column). Top row: prediction by the *MST* and its sensitivity (all-node results), middle row: prediction by the *MST*, its sensitivity, and the initial coarse-graining (coarse-grained results), bottom row: complete method using the previous steps and the "on-the-fly" coarse-graining (coarse-grained results). The average length (as before in % of the average length of the ordered non-*MST* edge lists without coarse-graining) of the ordered non-*MST* edge lists (all-node or coarse-grained) used for the calculations is depicted in black, as dashed line for a better visualization.

when only using the *MST* and its sensitivity. However, when applying either of the coarse-graining procedures the calculation costs depend on the severity of the perturbation, e.g. increasing the calculation costs for translocations of 0.5 Å, 1.0 Å, and 2.0 Å. Interestingly, for charge increases or decreases the calculation costs are not symmetric, revealing a calculation cost difference of ≈12 % for $\Delta q = \pm\ 0.050$, when using the complete *TN prediction method*. The prediction of the perturbed *MSTs* for both types of perturbations is based on the *MST* of the initial *TN*. Hence, a possible bias of the predicted, perturbed *MSTs* towards the initial *MSTs* needs to be ruled out. Comparing the initial *MST* with the predicted, perturbed *MSTs*

reveals a maximal edge similarity of only 10 % and 20 % (on average 5 ± 3 % and 10 ± 6 %) for increases or decreases of the additional point charge and the translocation of the additional point charge around its initial position, respectively. Obviously, the node similarity of the *MSTs* equals the node similarity of *TNs*. Hence, the initial and predicted, perturbed *MSTs* are sufficiently different to indicate a bias-free *TN* prediction. Furthermore, the *TN prediction method* is able to reproduce all *MSTs* from the complete, perturbed *TN* calculations (without any initial guess), indicating once more a bias-free prediction.

The prediction of the perturbed *MBPs* required average calculations of 80 %, 64 %, or 28 % of the non-*MST* edge lists for increases or decreases of the value of the additional point charge and 75 %, 51 %, or 23 % for the translocation of the additional point charge around its initial position, using the *MST* and its sensitivity, the *MST*, its sensitivity, and the initial coarse-graining, or the complete method using the previous steps and the "on-the-fly" coarse-graining, respectively. In all cases the calculation costs of the *MBPs* depend on the severity of the perturbation. For charge translocation of $|\boldsymbol{r}|$=0.5 Å or 1.0 Å and ϕ=135° the calculation of the edges from the initial *MST* was already sufficient to determine the perturbed, coarse-grained *MBPs*. In all cases the prediction of the perturbed *MBPs* required less edge calculations than the predictions of the perturbed *MSTs*, which is trivial since the *MBP* is a subset of the *MST*.

The *TN prediction method* reduces the costs of the *MST* or *MBP* calculations by coarse-graining, and thus reducing, the ordered non-*MST* edge lists (cf. Fig. 4.12, dashed black lines). The most accurate determination of the perturbed *MSTs* or *MBPs* (at least within the error of the coarse-graining method itself, discussed later) is guaranteed by a complete calculation of the coarse-grained non-*MST* edge lists. Such calculations reduce the costs by 40 % up to 80 % compared to the costs of the respective complete *TN* calculations. In principle, further costs reductions are possible, at least for the prediction of the perturbed *MBPs* (cf. Fig 4.12, red dots). However, defining a lower edge calculation bound for the determination of the perturbed *MBPs* is problematic due to high fluctuations (up to ± 17 %) associated with the non-deterministic order of edges with equal sensitivity values within the coarse-grained ordered non-*MST* lists used for the *TN* predictions. Consequently, a complete calculation of the coarse-grained ordered non-*MST* edge lists should be performed for all determinations of the perturbed *MSTs* or *MBPs*. A further benefit of a complete calculation, compared to a pre-defined edge calculation bound, is that the amount of edges to be calculated is flexible, self-regulated by the *TN* prediction

method and constantly adjusted to the requirements of the perturbed *TNs*.

For complex reactions, a single reaction pathway, even if it is the optimal one, is often not enough to achieve a comprehensive description of the process of interest [103]. Hence, a prediction of the second, third, fourth, and fifth best pathways, next to the actual *MBPs*, would be desirable. In Fig. 4.13 the results for the prediction of the second, third, fourth, and fifth best pathway are depicted for increases or decreases of the additional point charge and the translocation of the additional point charge around its initial position using the *MST* and its sensitivity, the initial coarse-graining, and the "on-the-fly" coarse-graining.

The average calculation costs for the prediction of the perturbed second, third, fourth, and fifth best pathways fluctuate, more or less, around the calculation costs for the perturbed *MBPs*. Hence, a complete calculation of the coarse-grained non-*MST* edge lists (dashed black lines), guarantees not only the most accurate prediction of the perturbed *MBPs* and *MSTs*, but also the most accurate prediction of the second, third, fourth, and fifth best pathways. Consequently, the *TN prediction method* is able to provide a proper description of complex reactions in a cost efficient manner.

For all perturbation scenarios, the initial and "on-the-fly" coarse-graining steps achieved significant cost reductions. At the same time, however, these coarse-graining steps represent potential sources of error regarding important properties of the perturbed *TNs*, e.g. the rate-determining, maximal transition barriers of the perturbed *MBPs* or other properties related to the *MSTs*. Here, the accuracy of the *TN prediction method* is analyzed with respect to the accuracy of the predicted rate-determining, maximal transition barriers of the perturbed *MBPs*.

The initial coarse-graining is based on the assumption that the coarse-graining of the initial *TNs* is a conserved feature over the *TN* perturbations, thereby allowing the neglect of edges connecting nodes within the same coarse-grained node from the perturbed edge lists. In principle, node additions or subtractions to or from a coarse-grained node are possible and occur frequently. Still, it is required that all nodes within the same coarse-grained can be reached by crossing barriers below the initial rate-determining, maximal transition barrier only. In Fig 4.14 "Perturbation 1" is not fulfilling this requirement, i.e. due to the subtraction of node *a* from the coarse-grained node, barriers of 6 kcal/mol need to be crossed in order to reach every node within the coarse-grained node, while the initial coarse-graining is assuming that all nodes can be reached by crossing barriers below 2 kcal/mol. Hence, the rate-determining, maximal transition barrier of the perturbed *MBP* would be 6 kcal/mol

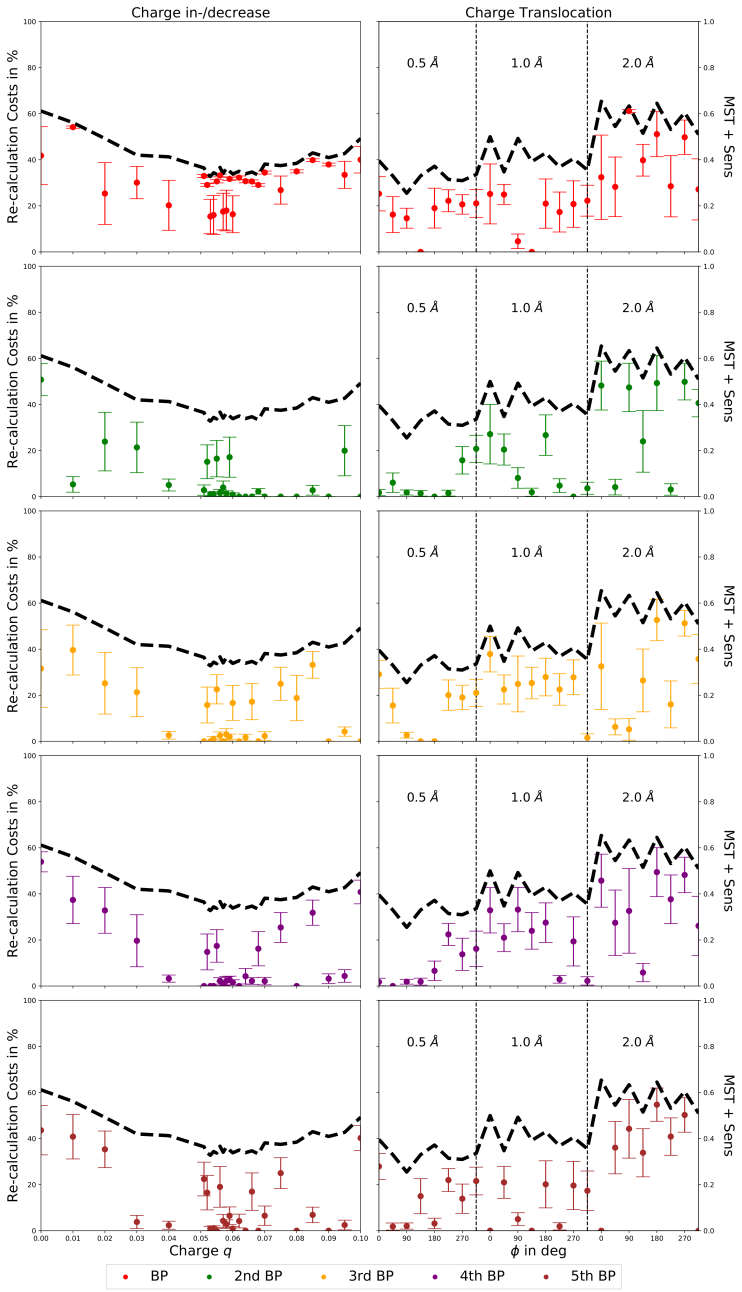


Figure 4.13: Average calculation costs (in % of the average length of the ordered non-*MST* edge lists without coarse-graining) for the coarse-grained prediction of the perturbed *MBP* (red), the second best pathway (green), the third best pathway (orange) the fourth best pathway (purple), and the fifth best pathway (brown) from 1000 prediction runs per perturbation using the *MST* and its sensitivity, the initial coarse-graining, and the "on-the-fly" coarse-graining. Perturbations: increase and decrease of the additional point charge (left column), translocation of the additional point charge around its initial position (right column). The average length (as before in % of the average length of the ordered non-*MST* edge lists) used for the predictions is depicted in black, as dashed line for a better visualization.

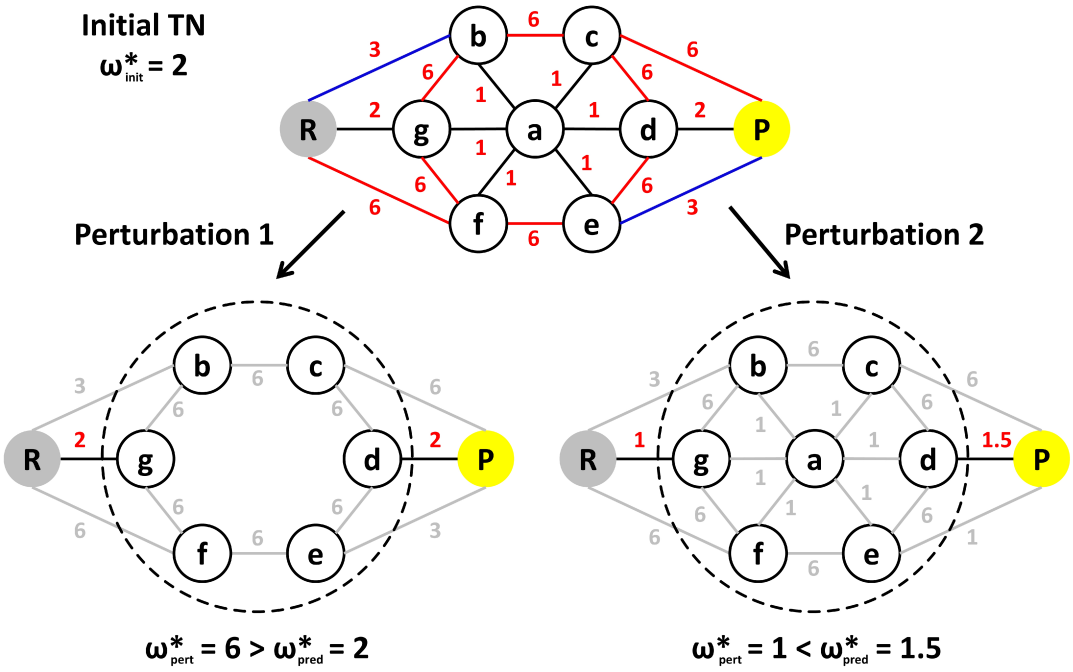


Figure 4.14: Initial *TN*: Nodes are depicted as circles, edges as lines. Reactant and product state are depicted in gray and yellow, labeled R and P, respectively. Other nodes are labeled from a to g. The edge weights are in kcal/mol and shown in red. Perturbed *TNs*: Node labeling as before, calculated edges are depicted in black, edges to be calculated are depicted in gray. The dashed circles indicate the main coarse-grained node according to the initial coarse-graining. Other coarse-grained nodes are the reactant and product state.

if all nodes are considered and 2 kcal/mol if the initial coarse-graining is applied. These findings indicate that the initial coarse-graining is prone to rate-determining, maximal transition barrier underestimations, while overestimations are ruled out.

The "on-the-fly" coarse-graining combines two coarse-grained nodes if they are connected by an edge with edge weight below the initial rate-determining, maximal transition barrier, thereby neglecting all further edges connecting the two coarse-grained nodes. This setup ensures an exact refinement of the perturbed rate-determining, maximal transition barrier if the perturbed rate-determining, maximal transition barrier is greater than or equal to the initial rate-determining, maximal transition barrier. For the opposite case, however, the "on-the-fly" coarse-graining could stop the refinement of the perturbed rate-determining, maximal transition barrier too early. A typical situation is depicted in Fig. 4.14 by "Perturbation 2". Here, the initial coarse-graining provided three coarse-grained nodes. Following the calculation of the edges belonging to the coarse-grained initial *MST* in the perturbed system all three coarse-grained nodes will be combined by the "on-the-

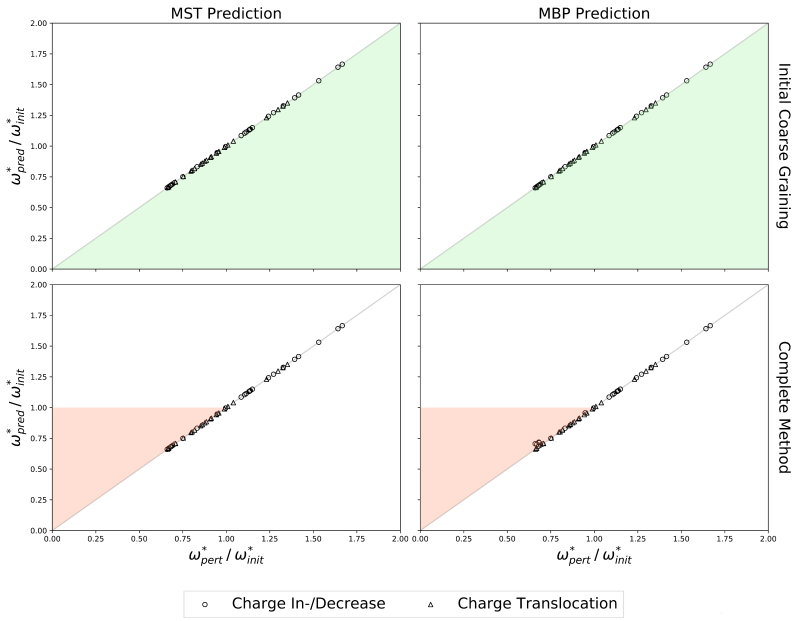


Figure 4.15: Comparison of ω^* from complete *TN* calculations (ω^*_{pert}) and *MST* (left column) or *MBP* (right column) predictions (ω^*_{pred}, averaged over 1000 prediction runs per perturbation scenario) following the initial coarse-graining (top row) or the complete prediction method (bottom row). The results for increases and decreases of the additional point charge are represented by circles, while the results for charge translocations around the initial position of the additional point charge are represented by triangles. Shaded areas indicate potential error regions, underestimations in green, overestimations in red.

fly" coarse-graining, thus preventing any further edge calculations. Thereby, the rate-determining, maximal transition barrier of the perturbed *MBP* would be 1 kcal/mol if all nodes are considered and 1.5 kcal/mol if the "on-the-fly" coarse-graining is applied. Hence, the "on-the-fly" coarse-graining step is prone to rate-determining, maximal transition barrier overestimations for situations in which the perturbed rate-determining, maximal transition barrier is lower than the initial rate-determining, maximal transition barrier, while underestimations are ruled out.

To check the accuracy of the *TN prediction method* the rate-determining, maximal transition barriers of the perturbed *MBPs* from complete *TN* calculations were compared with those accessible after the initial coarse-graining and those finally predicted.

The rate-determining, maximal transition barriers accessible after the initial coarse-graining are in perfect agreement to the barriers determined by the complete *TN* calculations (cf. Fig. 4.15), regardless of the type or severity of the perturbation or the ratio of the predicted and initial rate-determining, maximal transition barriers. Hence, the PES associated with the system under investigation is most

likely stabilizing the coarse-grained nodes, rendering perturbations as depicted in Fig. 4.14 by "Perturbation 1" unlikely. The rate-determining, maximal transition barriers predicted by the complete *TN prediction method* also agree with the barriers determined by the complete *TN* calculations. However, for situations in which the perturbed rate-determining, maximal transition barriers are smaller than the initial rate-determining, maximal transition barrier, slight inaccuracies can be observed for individual perturbation scenarios. The inaccuracies reported here, however, are well below the *RMSDs* reported in proton transfer benchmarks for the *semi-empiric* quantum method *OM2* [454]. Hence, the inaccuracies of the *TN prediction method* are negligible compared to the intrinsic error of the *semi-empiric* quantum method used. Nevertheless, the potential risk of rate-determining, maximal transition barrier over- or underestimations should not be forgotten.

Finally, the *TN prediction method* was tested for a larger set of perturbations, using each of the *TNs* derived from the complete *TN* calculations as initial *TN* for the prediction of the *MSTs* or *MBPs* of all other *TNs*. Thereby, the number of *TN* predictions was increased from 26 and 24 to 702 and 600 for the increase or decrease of the additional point charge or the translocation of the additional point charge around its initial position, respectively. Furthermore, the perturbation severities are increased from $\Delta q_{max} = \pm\, 0.050$ to $\Delta q_{max} = \pm\, 0.100$ and $|\boldsymbol{r}|_{max} = 2.0$ Å to $|\boldsymbol{r}|_{max} = 4.0$ Å, which allowed a testing of the *TN prediction method* for more pronounced perturbations. Additionally, this setup allowed the investigation of the influence of the initial *TN* and the similarity of initial and perturbed *TNs* on the prediction of the *MSTs* or *MBPs* of the perturbed *TNs*.

For the correct predictions of the perturbed *MSTs* or *MBPs* different amounts of edges needed to be calculated on average. For the *MSTs* not a single correct prediction was observed, when less than 20 % of the perturbed edge sets were calculated. On the other hand, correctly predicted *MBPs* were achieved in 16 % or 12 % of all *TN* predictions using only 5 % of the perturbed edge lists, for the increase or decrease of the additional point charge and the translocation of the additional point charge around its initial position, respectively. However, once calculations of more than 20 % of the perturbed edge lists are performed the increase in correctly predicted perturbed *MSTs* is higher than the increase in correctly predicted perturbed *MBPs*. In numbers, correct predictions of the perturbed MSTs or MBPs are achieved in 82 % and 90 % (for the increase or decrease of the value of the additional point charge) or 50 % and 71 % (for the translocation of the additional point charge around its initial position) of all *TN* predictions with a calculation of 50 %

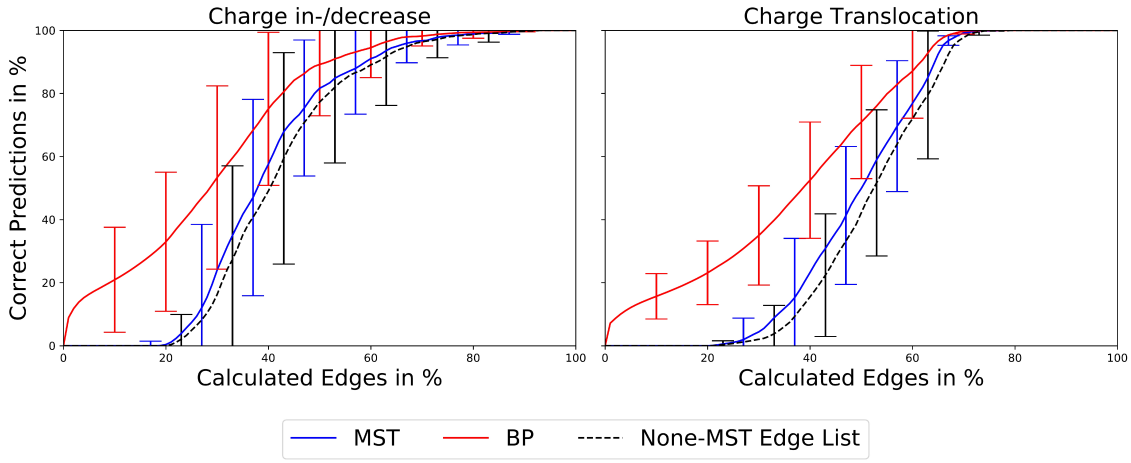


Figure 4.16: Correct predictions of the perturbed *MSTs* (blue) or *MBPs* (red) for individual numbers of calculated edges averaged over the different initial *TNs*. The dashed black lines indicate the number of correct *MST* or *MBP* predictions by calculating the complete coarse-grained ordered non-*MST* edge lists averaged over the different initial *TNs*. Left: increases or decreases of the additional point charge, right: translocation of the additional point charge around its initial position.

of the perturbed edge lists. With calculations of 60 % of the perturbed edge lists the number of correct predictions is above 90 % for all cases, except the perturbed *MST* predictions for charge translocations around the initial position (cf. Fig 4.16).

The dashed black lines in Fig. 4.16 indicate the number of correct *MST* or *MBP* predictions by calculating the complete coarse-grained non-*MST* edge lists and thus the length of the coarse-grained non-*MST* edge lists. The costs for these calculations display the same behavior as those for calculations stopped once the correct perturbed *MSTs* are reached, shifted by at most 5 % to higher edge calculation costs, while the difference to calculations stopped once the correct perturbed *MBPs* are reached is much larger. In 77 % and 42 % (or 72 %) of all *TN* predictions the number of edges in the non-*MST* edge lists was reduced by 50 % (or 40 %), due to the coarse-graining steps in the *TN prediction method* for the increase or decrease of the value of the additional point charge or the translation of the additional point charge around its initial position, respectively. Hence, in 77 % or 72 % of all *TN* predictions the predicted *MSTs* or *MBPs* are unequivocally correct (at least within the coarse-graining errors described before) when only calculating of 50 % or 60 % of the perturbed edge lists, because there are simply no further edges left to calculate. As already described before, further cost reductions are possible, but inevitably bear the risk of incorrect predictions of the perturbed *MSTs* or *MBPs*, since there are substantial amounts of edges left to be calculated. Taken as a whole, the number of correct predictions (stopped after the prediction of the *MSTs*, *MBPs* or the com-

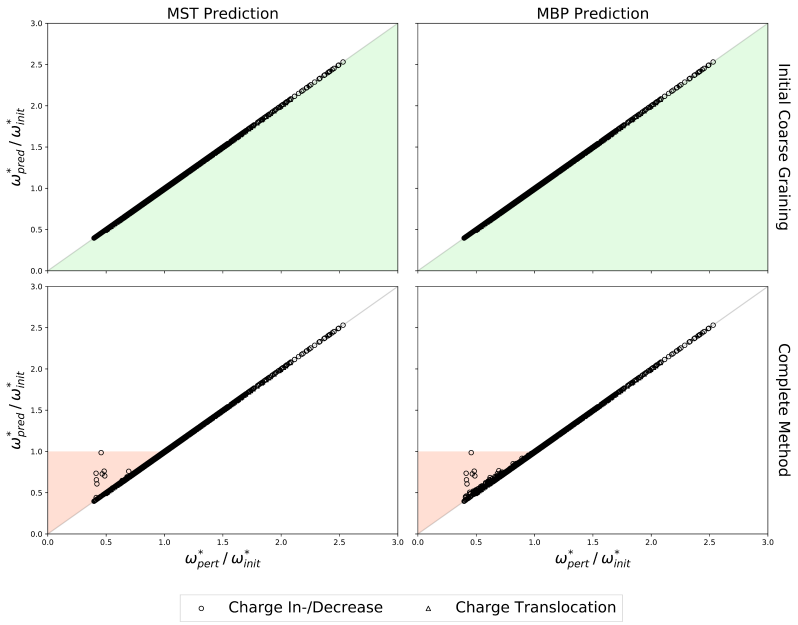


Figure 4.17: Comparison of ω^* from complete *TN* calculations (ω^*_{pert}) and *MST* (left column) or *MBP* (right column) predictions (ω^*_{pred}, averaged over 1000 prediction runs per perturbation scenario) following the initial coarse-graining (top row) or the complete method (bottom row). The results for increases and decreases of the additional point charge are represented by circles, while the results for charge translocation around the initial position of the additional point charge are represented by triangles. Shaded areas indicate potential error regions, underestimations in green, overestimations in red.

plete calculation of the coarse-grained ordered non-*MST* edge lists) depends to a large extent on the initial *TN* or the similarity of the initial and perturbed *TN* (cf. Fig 4.16).

Once again, the accuracy of the rate-determining, maximal transition barriers along the *MBPs* was checked by comparing the barriers of the complete *TN* calculations with those accessible after the initial coarse-graining and the complete *TN* prediction procedure (cf. Fig. 4.17).

As before, not a single rate-determining, maximal transition barrier underestimation was observed following the initial coarse-graining of the perturbed *TNs*, thus supporting the hypothesis that the PES of the studied model system is preventing such deviations. The same holds for the rate-determining, maximal transition barriers derived by the complete *TN prediction method* if the perturbed rate-determining, maximal transition barriers are higher than the initial rate-determining, maximal transition barriers. For the opposite case, however, deviations from the perturbed rate-determining, maximal transition barriers were observed in 12 of all 1 302 combinations of initial and perturbed *TNs*. 5 out of 12 of these overestimations occurred

in *TN* predictions in which the complete *TN* with q=0.000 was used as initial *TN*, which is intuitively obvious since the rate-determining, maximal transition barrier of the complete *TN* with q=0.000 is the highest with 9 kcal/mol and thus provides the maximal range for barrier overestimations. On the other hand, 11 out of 12 overestimations occurred in *TN* predictions in which the perturbed *TNs* displayed the lowest rate-determining maximal transition barrier of 3 kcal/mol (4 kcal/mol for the remaining one), once again providing the maximal range for barrier overestimations. Hence, further investigations might be necessary for situations in which the perturbed rate-determining, maximal transition barriers are lower than the initial ones.

Overall, complete calculations of the coarse-grained non-*MST* edge lists provide significant cost reductions paired with accurate predictions of the perturbed *TNs*. In conclusion, it needs to be said that the *TN prediction method* is flexible in costs, self-regulated along the prediction, and constantly adjusted to the requirements of the perturbed *TNs* to be predicted. On the downside, further investigations might be necessary for an accurate determination of important network properties like the rate-determining, maximal transition barrier of the perturbed *MBPs*. In principle, the *TN prediction method* follows Boruvka's algorithm combined with an excessive pre-processing.

Hence, the *TN prediction method* allows an efficient investigation of *DOFs*, which need to be sampled indirectly. Thereby, providing another step towards an accurate and complete description of complex biological reactions using the *TN* approach.

Chapter 5

Results II: D-Channel Proton Transfer

In the following chapter, the transfer of an excess proton along the D-channel of *CcO* is investigated during the $P_R \rightarrow F$ transition using the *TN-MD method*. However, due to the high number of *DOFs*, potentially relevant for the translocation, a complete sampling of the side chain dihedral angle rotations and protonation states, following the direct sampling approach, next to the *MD*-based water sampling, is not possible, at least within a finite amount of time. Consequently, several preliminary investigations preceded the calculation of the *TN* to identify the most relevant *DOFs* for the proton transfer. These investigations included an analysis of the side chain dihedral angle conformation of N139, proposed to serve as a steric gate of the D-channel [156], as well as an investigation of the dehydrated region surrounding N139 in the so called "open" state, proposed to serve as a hydration gate of the D-channel [156]. These preliminary investigations allowed the refinement of an initial proton transfer pathway along the D-channel. Finally, the *TN*, describing the proton translocation along the D-channel, was sampled by five individual *TN* calculations along the initial proton transfer pathway which were ultimately combined to a single *TN*.

Throughout this chapter a model system of the D-channel is employed containing the most important amino acids of the D-channel along with several water molecules, embedded in a fixed, uncharged hull. A more sophisticated description of the D-channel model system is provided in subchapter 5.1. Next to the description of the model D-channel, this subchapter summarizes the preliminary investigations and provides all settings for the *TN* calculations along the D-channel. The results of the final *TN* calculations are presented and discussed in subchapter 5.2.

5.1 Background, Methods, and Settings III

To elucidate the proton transfer along the D-channel of *CcO* a model system of the D-channel was constructed from a snapshot of an existing *MD* simulation, subsequently published together with other *MD* simulations in *Bagherpoor et al* [455], including SU I and II of *CcO* from *R. sphaeroides* at a resolution of 2.0 Å (PDB [456] entry: 2GSM [186]) embedded in a lipid bilayer of *phosphatidylcholines* and solvated in TIP3P water [451]. Here, the protein and the lipid bilayer were parameterized according to the CHARMM22 force field [457, 449] and the CHARMM36 extension for lipids [458], while the setup of the system and the parametrization of the co-factors, e.g. *heme a* or *heme* a_3, followed the scheme of [242].

The model system of the D-channel contains the highly conserved amino acid residues D132 and E286 (*R. sphaeroides* numbering), representing the entrance and exit of the D-channel, respectively, as well as all hydrophilic amino acid residues and water molecules in between. The hydrophilic amino acid residues are: H26, Y33, N121, N139, S142, S197, S200, S201, and N206. Furthermore, all amino acid residues, within 10 Å to these eleven amino acid residues and all water molecules within 15 Å to D132 are included. Overall the model system of the D-channel is comprised of 4 003 atoms, representing 188 amino acid residues and 331 water molecules. From these atoms, only 283 atoms are flexible and charged, according to the same parameter set employed for the *MD* simulation or the *semi-empiric* quantum method *OM2* [244, 245], while all remaining atoms are fixed and uncharged and thus serving as steric hull enclosing the channel and its entrance. The set of flexible atoms includes the side chains of H26, Y33, N121, D132, N139, S142, S197, S200, S201, N206, and E286, i.e. no backbone atoms of type N, O, C, C_α, HN, HO, H, and H_α, and 33 water molecules (cf. Fig. 5.1). In case of a *QM* treatment of the flexible atoms, link atoms are introduced between the C_α and C_β atoms for all amino acid residues under consideration to compensate the cuts of covalent bonds at the boundary of *QM* and *MM* treatment.

The *MD* simulation underlying the D-channel model system displayed the following charges for the side chains of the flexible amino acid residues:

1. Zero charge for H26 (HSE), Y33, N121, D132, N139, S142, S197, S200, and S201.

2. Charge of -1 for E286.

Consequently, the overall charge of the D-channel model system is -1, representing

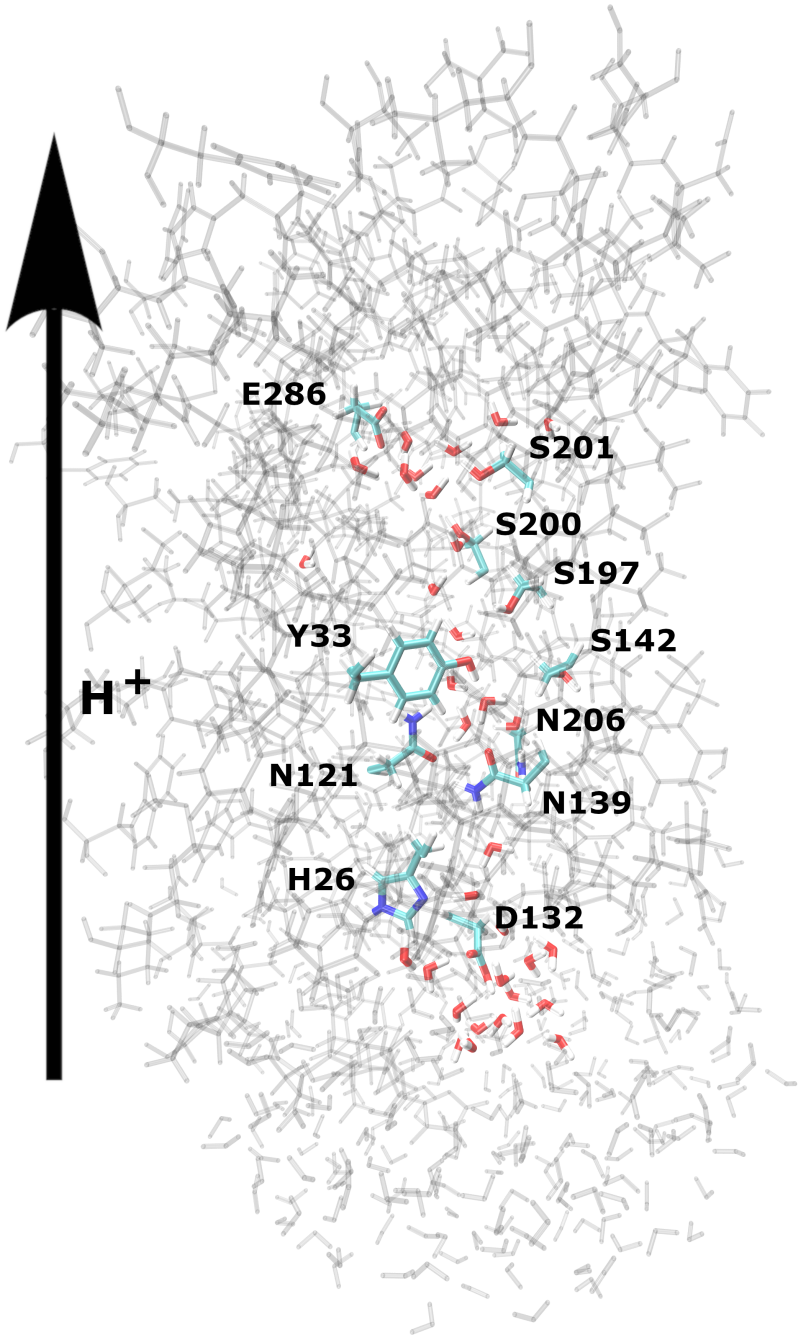


Figure 5.1: Model system of the D-channel, containing a flexible and charged set of atoms, including the side chains of H26, Y33, N121, D132, N139, S142, S197, S200, S201, N206, and E286 and 33 water molecules (colored), and a fixed and uncharged set of atoms (gray), with protonated D132, deprotonated H26, and deprotonated E286, representing the reactant state of the proton translocation along the D-channel.

a state in which the excess proton is about to be transferred from D132 to E286. This state is said to be the reactant state of the proton translocation along the

D-channel. To investigate the proton translocation along the D-channel a second state of the model system was constructed. Here, the excess proton is located at E286 instead of D132. This state is said to be the product state of the proton translocation along the D-channel. Following their *QM* potential energy minimization using the *semi-empiric* quantum method *OM2* and the *ABNR* minimization algorithm with a 0.001 kcal/(molÅ) convergence criterion both states of the model channel displayed a potential energy difference of ≈9 kcal/mol in favor of the product state. From a qualitative perspective the potential energy difference between the reactant and product state is reasonable, i.e. allowing a proton transfer from D132 to E286. Shortly thereafter, however, multi-state reactive *MD* simulations (*MS-RMDs*), based on metadynamics, performed by *Liang et al* [206] to investigate the proton transfer during the $P_R \rightarrow F$ transition, revealed a free energy difference of ≈6 kcal/mol in favor of the product state. Although direct comparisons between potential and free energy values need to be treated with caution, it may be deduced from *Liang et al's MS-RMD* calculations [206], that the model system of the D-channel employed within the framework of this thesis, is overestimating the energy difference of reactant and product state. The magnitude of this overestimation, however, is rather small when considering the extensive set of simplifications introduced along the derivation of the model channel.

The preliminary investigations of the proton transfer along the D-channel started with twelve short, 5 ns long, *MD* simulations in the D-channel model system following the same setup as described before and different protonation states, i.e. 6 *MD* simulations with protonated D132 and six *MD* simulations with protonated E286. These *MD* simulations allowed the re-production, as well as possible extensions, of individual results derived by *Henry et al* [156]. These aspects are summarized in the following:

- Conformation of N139

  Henry et al [156] proposed a gating function of the side chain of N139, allowing or preventing the formation of a continuous *HBC* through the D-channel. Both states were defined by the χ_1 and χ_2 dihedral angles, i.e. "closed" state, (χ_1,χ_2)=(-165°,41°), and "open" state, (χ_1,χ_2)=(-75°,-70°), with the latter state being ≈4 kcal/mol less favorable in terms of free energy (cf. Fig 5.9 and 5.10). Along the short *MD* simulations performed in the D-channel model system both states of N139 were observed. Interestingly, however, with rather different probabilities. In all three *MD* simulations with protonated E286, starting in the "closed" state of N139, not a single transition to the "open"

state had been observed, while the remaining *MD* simulations with protonated E286, starting in the "open" conformation of N139 transitioned quickly, within the heating or equilibration phase of the *MD* simulations, in the "closed" conformation. Nearly, the opposite can be observed for the *MD* simulation with protonated D132. Here, all but one *MD* simulations, starting in the "closed" conformation of N139 ended up in the "open" conformation, while the *MD* simulations in the "open" conformation stayed in the "open" conformation.
The *MD* simulations with protonated D132 revealed a further interesting aspect. Namely, the χ_2 dihedral angle of N139 is flipped by 180° compared to the state definitions from *Henry et al* [156]. That means, on a conformational level, that the carboxamide group of N139 is flipped. Thereby, the amide group of N139, representing the more positively charged part of the side chain, is always pointing towards the deprotonated end residue of the D-channel. According to their results, *Henry et al* [156] suggested a misassignment of the carboxamide group of N139 in the crystal structure of *CcO*, resulting in a metastable rotamer related to the global minimum by a 180° flip of χ_2. The short *MD* simulations performed within the framework of this thesis, however, simply suggest a different protonation state of D132 and E286 at either end of the D-channel.

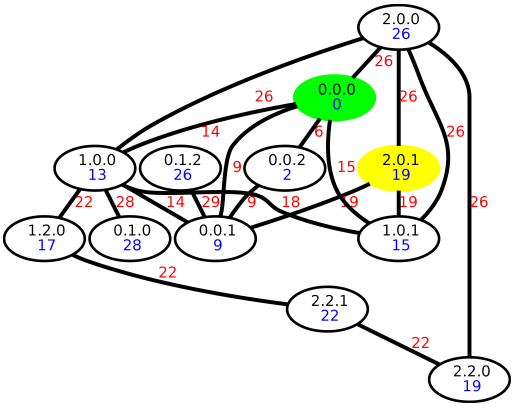


Figure 5.2: TN for individual conformations of the asparagine gate with protonated D132 at the D-channel entrance. The nodes are shown as ellipses, labeled according to the side chain dihedral angle χ_2 of N139 and χ_1 and χ_2 of the opposing N121 ($\chi_{2.N139}$·$\chi_{1.N121}$·$\chi_{2.N121}$). The "open" conformation of N139, according to Henry et al [156], is shown in yellow. The "open" conformation of N139 with flipped χ_2 dihedral angle is shown in green. Edges are shown as lines. Blue numbers represent the potential energies of the nodes, red numbers the potential energies of the rate-determining, maximal transition barriers along the edges used as edge weight. All energies are in kcal/mol, relative to the "open" conformation of N139 with flipped χ_2 dihedral angle, and rounded to integer values.

The second aspect was further investigated by small *MM TN* calculations focusing on the conformation of χ_2 of N139 and the conformation of χ_1 and χ_2 of the opposing N121 for different protonation states of D132 and E286. χ_1 of N139 was artificially restrained to the "open" conformation.

To generate the *TN* depicted in Fig. 5.2, the side chain dihedral angles χ_2 of N139 and χ_1 and χ_2 of N121 were sampled following the direct sampling approach (cf. subchapter 4.1) with a sampling interval of 120°, giving rise to 27 distinct initial states. Following the *MM* potential energy minimization using the *ABNR* minimization algorithm and the subsequent state assignment, 15 distinct states remained. Between these states, 47 neighbor pairs were identified using an *RMSD* criterion of 2.5 Å, instead of the criteria applied in subchapter 4.1. Consequently, 47 *CPR* calculations were performed to connect all neighbor pairs. In Fig. 5.2 only a subset of the *TN* is depicted, leaving out all edges, and related nodes, with relative edge weights above 30 kcal/mol. The "open" conformation of N139 with flipped χ_2 dihedral angle (χ_1,χ_2)=(-62°,118°), i.e. the amide group of the side chain of N139 pointing towards the deprotonated E286, was used as starting structure. According to the *TN* this structure is also the global minimum of all sampled conformations in the model system of the D-channel with protonated D132 and deprotonated E286. The relative potential energies of all other states are, in most cases, much higher, i.e. well above 10 kcal/mol. Notable exceptions are two states, displaying a rotation of the side chain dihedral angle χ_2 of N121 only. The "open" conformation of N139 according to *Henry et al* [156], i.e. the amide group of the side chain of N139 pointing towards the protonated D132, can be assigned to nodes with $\chi_{2.N139}$ equal to 2. The minimal, relative potential energy of all states with $\chi_{2.N139}$ equal to 2 is 19 kcal/mol.

In Fig. 5.3 the *TN* for the model system of the D-channel with protonated E286 is depicted. Once again, the side chain dihedral angles χ_2 of N139 and χ_1 and χ_2 of N121 were sampled following the direct sampling approach (cf. subchapter 4.1) with a sampling interval of 120°, giving rise to 27 distinct initial states. Following the *MM* potential energy minimization using the *ABNR* minimization algorithm and the subsequent state assignment, 12 distinct states remained. Between these states 31 neighbor pairs were identified using an *RMSD* criterion of 2.5 Å, instead of the criteria applied in subchapter 4.1. Consequently, 31 *CPR* calculations were performed to connect all neighbor pairs. In Fig. 5.3 only a subset of the *TN* is depicted, leaving out all

edges with relative edge weights above 30 kcal/mol. As before, the "open" conformation of N139 with flipped χ_2 dihedral angle (χ_1,χ_2)=(-64°,121°), i.e. the amide group of the side chain of N139 pointing towards the protonated E286, was used as starting structure. In contrast to the *TN* displayed in Fig. 5.2, however, this structure is not the global minimum of all conformations sampled. Instead, the "open" conformation of N139 according to *Henry et al* [156], i.e. the amide group of the side chain of N139 pointing towards the deprotonated D132, is favored by 1 kcal/mol. Overall, however, χ_2 conformations of N139 with the amide group of the side chain pointing towards the deprotonated D132 or the protonated E286 are more or less similar in their relative potential energies. The latter finding is somewhat contradicting the observed χ_2 dihedral angle conformations of N139 along the *MD* simulations. A possible explanation might be the artificially restrained χ_1 dihedral angle of N139.

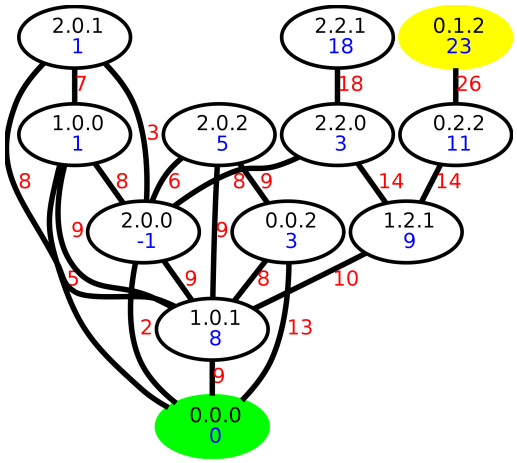


Figure 5.3: TN for individual conformations of the asparagine gate with protonated E286 at the D-channel end. The nodes are shown as ellipses, labeled according to the side chain dihedral angle χ_2 of N139 and χ_1 and χ_2 of the opposing N121 ($\chi_{2.N139}$·$\chi_{1.N121}$·$\chi_{2.N121}$). The "open" conformation of N139, according to *Henry et al* [156], is shown in yellow. The "open" conformation of N139 with flipped χ_2 dihedral angle is shown in green. Edges are shown as lines. Blue numbers represent the potential energies of the nodes, red numbers the potential energies of the rate-determining, maximal transition barriers along the edges used as edge weight. All energies are in kcal/mol, relative to the "open" conformation of N139 with flipped χ_2 dihedral angle, and rounded to integer values.

To further investigate the flipping behavior of the χ_2 dihedral angle of N139 a third *TN* was calculated. Here, both end residues of the D-channel were deprotonated. Consequently, the excess proton was located on a water molecule, forming an hydronium ion, which was spatially constraint shortly above the

asparagine gate. The resulting *TN* is depicted in Fig. 5.4.

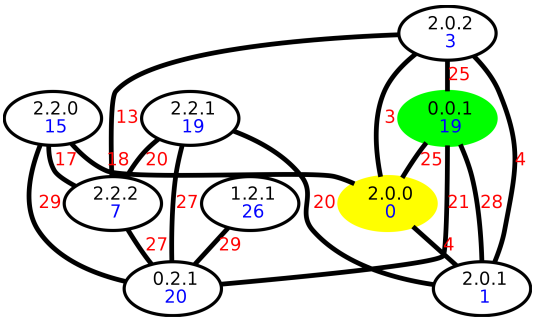


Figure 5.4: TN for individual conformations of the asparagine gate with a spatially constrained hydronium ion shortly above the asparagine gate. The nodes are shown as ellipses, labeled according to the side chain dihedral angle χ_2 of N139 and χ_1 and χ_2 of the opposing N121 ($\chi_{2.N139}$·$\chi_{1.N121}$·$\chi_{2.N121}$). The "open" conformation of N139, according to Henry et al [156], is shown in yellow. The "open" conformation of N139 with flipped χ_2 dihedral angle is shown in green. Edges are shown as lines. Blue numbers represent the potential energies of the nodes, red numbers the potential energies of the rate-determining, maximal transition barriers along the edges used as edge weight. All energies are in kcal/mol, relative to the "open" conformation of N139 with flipped χ_2 dihedral angle, and rounded to integer values.

As for the *TNs* depicted in Figs 5.2 and 5.3, the side chain dihedral angles χ_2 of N139 and χ_1 and χ_2 of N121 were sampled following the direct sampling approach (cf. subchapter 4.1) with a sampling interval of 120°, giving rise to 27 distinct initial states. Following the *MM* potential energy minimization using the *ABNR* minimization algorithm and the subsequent state assignment, 10 distinct states remained. Between these states 20 neighbor pairs were identified using an *RMSD* criterion of 2.5 Å, instead of the criteria applied in subchapter 4.1. Consequently, 20 *CPR* calculations were performed to connect all neighbor pairs. In contrast to the previous *TN* calculations the "open" conformation of N139 according to *Henry et al* [156], i.e. the amide group of the side chain of N139 pointing towards the deprotonated D132, was used as starting structure. According to the *TN*, this structure is also the global minimum of all sampled conformations in the model system with an hydronium ion above the asparagine gate. The "open" state of N139 with flipped χ_2 dihedral angle, i.e. the amide group of the side chain of N139 pointing towards the deprotonated E286, can be assigned to nodes with $\chi_{2.N139}$ equal to 0. The minimal, relative potential energy of all states with $\chi_{2.N139}$ equal to 0 is 19 kcal/mol.

Overall, the evaluation of different *MD* simulations and *TN* calculations proposed a conformational dependency of the χ_1 and χ_2 dihedral angles of N139

on the protonation state of the D-channel. These aspects were further analyzed, and subsequently confirmed, in *Bagherpoor et al* [455] and *Ghane et al* [459].

- Hydration of the asparagine region

  Apart from the gating function of the side chain of N139, *Henry et al* [156] proposed a hydration gate surrounding N139 in the "open" conformation. To analyze the hydration gate, the water distribution along the D-channel was investigated by short, 5 ns long, *MD* simulations in the D-channel model system with protonated D132 at the D-channel entrance. Obviously, 5 ns are not sufficient to describe long-scale water diffusions. However, important aspects of the water mobility along the D-channel can be derived.

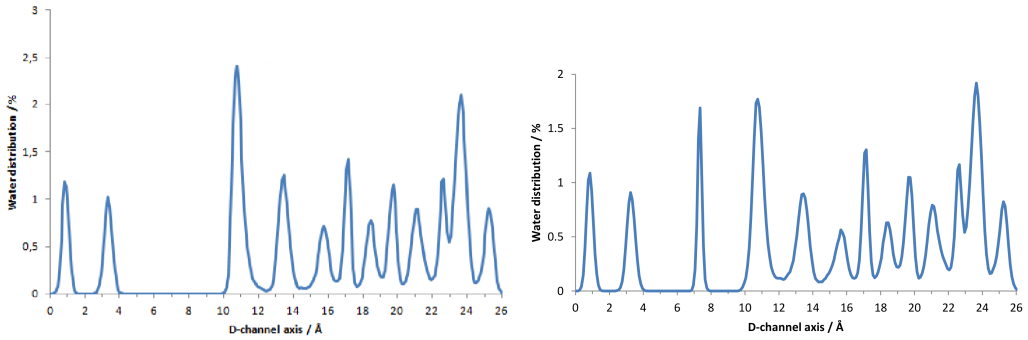


Figure 5.5: Water distribution along the D-channel axis, defined as the straight line connecting the C_α atoms of D132 and E286, averaged over 5 ns *MD* simulations with protonated D132 at the D-channel entrance. Left: Water distribution without an artificially placed water molecule in between N121 and N139. Right: Water distribution with an artificially placed water molecule in between N121 and N139.

The water distributions displayed in Fig. 5.5 correspond to 5 ns *MD* simulations of the D-channel model system without (left) and with (right) an artificially placed water molecule in between N121 and N139. For *MD* simulations without an artificially placed water molecule in between N121 and N139, the first two peaks of the water distribution, located 0.85 Å and 3.35 Å away from the D-channel entrance, are rather narrow and do not interconvert, i.e. the water molecules are localized. Therefore, *Henry et al* [156] referred to this region as being "solid-like". Henry's "solid-like" region, however, contains three water molecules. Reason for this difference is the conformation of D132. Here, D132 points inside the D-channel, while it was flipped out and replaced by a water molecule in Henry's work. The subsequent region from 4.35 Å to

9.75 Å, containing N139 and N121, is void of any water molecule, although N139 is in the "open" conformation for the entire *MD* simulation due to the protonation of D132. Following the hydration gap, ten interconverting water peaks can be observed in the region up to E286. Hence, the flexibility of the water molecules in the upper region is higher than the flexibility of the water molecules at the channel entrance. Interestingly, the global maximum of the water distribution is located directly after the water gap, presumably indicating the side from which the water gap is hydrated. For *MD* simulations with an artificially placed water molecule in between N121 and N139, the artificially placed water molecule bridges the water gap at ≈7 Å. Interestingly, the water molecule stays at this location for the entire *MD* simulation, resulting in a sharp peak. Consequently, the distribution of all other water molecules is not affected by the additional water molecule, except a significant reduction of the peak at 10.75 Å and an increased inter-conversion. Once again indicating the side from which the water gap could be hydrated.

To further investigate the hydration of the water gap in between N121 and N139 *MM CPR* calculations were performed to determine the transition barrier for the diffusion of a water molecule in between the asparagines. According to the previous paragraphs, four distinct cases needed to be considered. These are:

1. The hydration of the asparagine region from the flexible or "solid-like" region.
2. The hydration of the asparagine region with the amide group of N139 pointing towards D132 or E286.

These calculations indicated a hydration of the asparagine region from the flexible region (as already indicated by Fig. 5.5) with the amide group of N139 pointing towards D132.

The preliminary investigations (summarized above) allowed the determination of an initial proton transfer pathway through the D-channel connecting D132 and E286. Three crucial steps of the proton transfer are:

1. Opening of N139.
2. Hydration of the asparagine region from the flexible side.
3. Proton transfer from D132 to E286.

Along the derivation of the initial proton transfer pathway additional steps were added to the pathway, leading to the final, initial proton transfer pathway sequence:

1. Proton transfer from D132 to H26.
2. Concerted opening of N139 and hydration of the asparagine region.
3. Formation of a continuous *HBC*.
4. Proton transfer from H26 to E286.
5. Closing of N139.

Hence, the proton transfer along the D-channel is represented by five sub-transitions, which were calculated using the *CPR* algorithm and the *semi-empiric* quantum method *OM2*. The four intermediate states along the pathway were modeled and minimized according to the minimization of the reactant and product state of the overall proton translocation. The key aspects of the initial proton transfer pathway, in terms of relative potential energies of the intermediate states and relative transition barriers, are summarized in Tab. 5.1 and depicted in Fig. 5.6.

Table 5.1: The five steps of the initial proton transfer pathway with potential energies (E) for the initial, transition, and final state of either sub-transition. All potetial energies are in kcal/mol and relative to the potential energy of the reactant state of the overall proton translocation, i.e. the initial state of transition $1 \rightarrow 2$.

Transition	Initial E in kcal/mol	TS E in kcal/mol	Final E in kcal/mol
$1 \rightarrow 2$	0.00	3.12	-1.52
$2 \rightarrow 3$	-1.52	9.87	0.46
$3 \rightarrow 4$	0.46	0.89	-5.03
$4 \rightarrow 5$	-5.03	6.87	-3.15
$5 \rightarrow 6$	-3.15	2.79	-9.06

The initial proton transfer pathway starts with a transfer of the excess proton from D132 to H26 via three water molecules involving only a minor transition barrier of ≈ 3 kcal/mol. Following the proton transfer, N139 starts to open up, allowing the hydration of the asparagine region in a concerted fashion. The TS with a transition barrier of ≈ 10 kcal/mol is reached in the middle of the opening. Subsequently, a continuous *HBC* is formed through the asparagine region with a transition barrier of ≈ 1 kcal/mol to finally transfer the excess proton from H26 to E286 with a transition barrier of ≈ 7 kcal/mol. This second proton transfer step involves several water

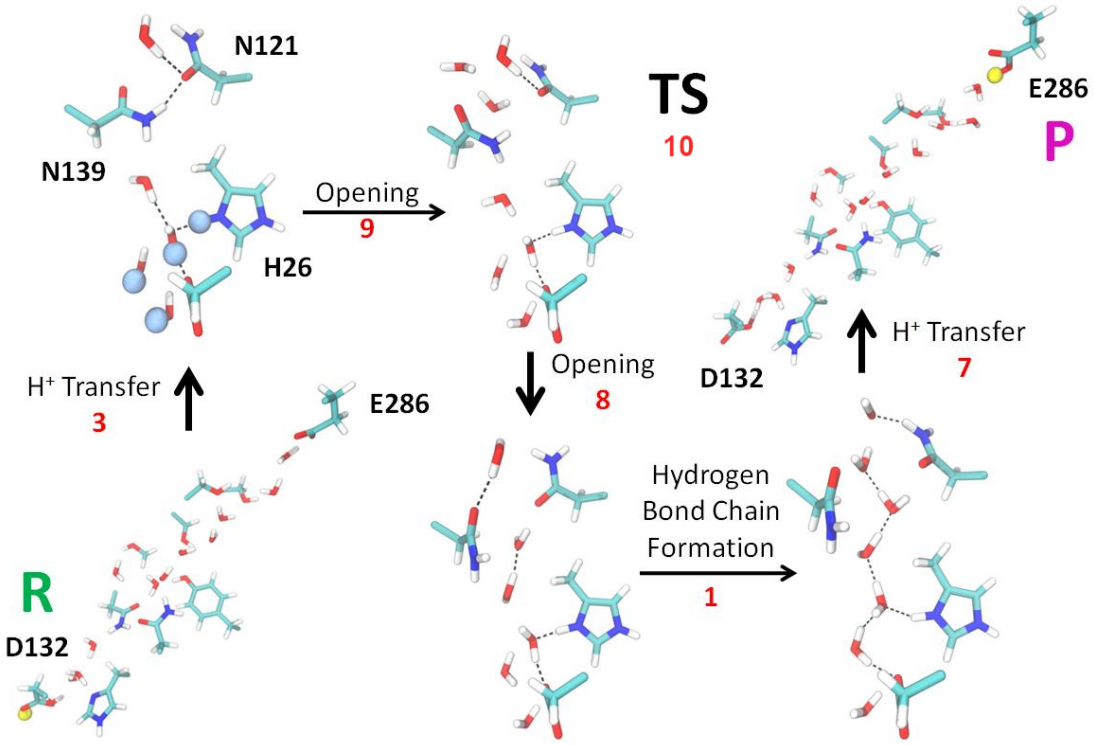


Figure 5.6: Initial proton transfer pathway from D132 to E286 along the D-channel of CcO. Reactant (bottom left) and product (top right) state of the overall proton translocation are labeled R and P, while the initial and final position of the excess proton is indicated by yellow spheres. Other depicted states are: the lower part of the D-channel with protonated H26 and transferred protons indicated by blue spheres (top left), the transition state (TS) of the overall proton translocation along the D-channel with a partially "open" N139 (top middle), the lower part of the D-channel with "open" N139 and a hydrated asparagine region (bottom middle), and the lower part of the D-channel with a continuous *HBC* through the asparagine region (bottom right). Transitions are indicated by black arrows. Transition barriers are shown in red and rounded to integer kcal/mol values. Hydrogen bonds are depicted by dashed black lines.

molecules, the side chain of Y33, as well as the side chain of S201. Interestingly, however, not the excess proton located at H26 is translocated to E286, but the proton hole located at E286 is translocated to H26. The proton translocation along the D-channel is concluded by the dehydration and closing of the asparagine region. With a transition barrier of ≈ 3 kcal/mol the closure of the D-channel is energetically much more favorable than the preceding opening, which is in accordance to *Bagherpoor et al* [455].

MD simulations from *Ghane et al* [459] with protonated H26 oppose the initial proton transfer pathway depicted in Fig. 5.6. Here, the protonated H26 forms, within a few ns of the *MD* simulation, a long lasting, rarely broken hydrogen bond

with the deprotonated D132, rendering a further translocation of the excess proton towards E286 unlikely. Such an effect has not been observed in the model system of the D-channel. On the other hand, *Liang et al* [206] report, that N139 opens up spontaneously, while the excess proton is approaching, which is in accordance to *MD* simulations with a spatially constrained hydronium ion beneath the asparagine gate from *Ghane et al* [459]. Comparing the transition barriers of the concerted opening and hydration of the asparagine region for the excess proton located at D132 or H26, respectively, reveals a transition barrier difference of ≈ 4 kcal/mol in favor of a protonated H26. These findings coincide with the results from *Liang et al* [206] and *Ghane et al* [459], relating the opening and hydration of the asparagine region with the distance of the asparagine region to the excess proton. Hence, it may be deduced that a protonated H26, or a proton in close vicinity to H26, is only occurring rather transiently, while the excess proton moves towards the asparagine region. Due to the setup of the model system and the absence of any kinetic energy, however, the approaching proton is only expressed as protonated H26 in the model system. Nevertheless, the main effect of the approaching proton, i.e. the reduction of the transition barrier for the concerted opening and hydration of the asparagine region, is preserved.

Apart from the protonation of H26, a further discussion of the hydration of the asparagine region, the formation of the *HBC*, and the subsequent proton translocation from H26 to E286 is required. The steps of the initial proton transfer pathway, depicted in Fig. 5.6, follow the *classic view* on the translocation of protons, among others proposed for *CcO* by *Henry et al* [156]. Here, the excess proton resides at a fixed location, e.g. H26, until a continuous *HBC* is formed, which facilitates the translocation of the proton. On the other hand, *Liang et al* [206] proposed a dynamic proton translocation. Here, the excess proton forms its own *HBC* while it is transferred. Such an effect has also been observed in carbon nanotubes [155]. Consequently, the asparagine region is hydrated from the "solid-like" region in *Liang et al*'s pathway [206], while it is hydrated from the flexible region in the initial proton transfer pathway, which is in accordance to *Henry et al*'s suggestions [156].

The rate-determining, maximal transition barriers of the overall proton translocation along the D-channel from *Liang et al*'s pathway [206] and the initial proton transfer pathway (cf. Fig. 5.6) are more or less equal, showing a difference of less than 1 kcal/mol (caution, again comparing free and potential energies). The transition states, on the other hand, differ considerably. For *Liang et al*'s pathway [206] the transition state is assigned to a conformation in which the charged species, the

hydronium ion, is located directly in between N121 and N139, while it is the midpoint of the opening of the asparagine region in the initial proton transfer pathway. This difference is remarkable, rendering the rate-determining, maximal transition barrier equality a coincidence and not a necessity.

To further investigate the discrepancies between the initial proton transfer pathway and the pathway suggested by *Liang et al* [206], five *TNs* were calculated along the five sub-transitions introduced in Tab. 5.1 and Fig. 5.6, using the *TN-MD method*. The main focus of these *TN* calculations was the opening of the asparagine region to further investigate the state corresponding to the rate-determining, maximal transition barrier. Consequently, the *TN* describing the opening of the asparagine gate received the most attention. The setup for all five *TN* calculations is summarized in Tab. 5.2.

Table 5.2: The setup of five *TN-MD* calcaultions. *DOFs* sampled directly: side chain dihedral angle rotations and protonation states. *MD* sampling: motion of water molcules. Number of initial states, nodes, and edges for every *TN* calculation. Individual settings of the direct sampling of side chain dihedral angle rotations and protonation state changes as well as the sampling of water configurations by *MD* simulations with the number of snapshots investigated (cf. Fig. 5.7): 1a) χ_1 and χ_2 of residue 4 in 45° intervals, 1b) Protonation of residues 0, 4, 27, 32, 38, 39, 43, and 44, 1c) Two 1 ns *MD* simulations with residue 4 or 32 protonated and snapshots taken every 10 ps, 2a) χ_1 and χ_2 of residues 3 and 5 in 45° intervals, 2b) Two 1 ns *MD* simulations with protonated residue 0, residue 5 in the "open" conformation, and residue 20 in between or above residues 3 and 5 and snapshots taken every 10 ps, 3a) One 1 ns *MD* simulation with a continuous *HBC* through residues 3 and 5 and snapshots taken every 2 ps, 4a) Protonation of residues 0, 11, 15, 16, 18, 20, 23, 25, 27, 32, and 41, 4b) Twelve 100 ps *MD* simulations with residue 0, 11, 15, 16, 18, 20, 23, 25, 27, 32, or 41 protonated and snapshots taken every 10 ps, and 5a) χ_1 and χ_2 of residues 3 and 5 in 90° intervals.

Transition	Init	Nodes	Edges	Direct Sampling	MD
$1 \rightarrow 2$	776	60	951	SC^{1a} and P^{1b}	200^{1c}
$2 \rightarrow 3$	4296	685	11283	SC^{2a}	200^{2b}
$3 \rightarrow 4$	500	283	2557	-	500^{3a}
$4 \rightarrow 5$	132	77	1388	P^{4a}	120^{4b}
$5 \rightarrow 6$	256	221	57	SC^{5a}	-

The five *TNs* describing the individual sub-transitions of the proton transfer through the D-channel were subsequently combined to a single *TN* describing the overall proton translocation from D132 to E286 along the D-channel. The combined *TN* is presented and discussed in the following subchapter.

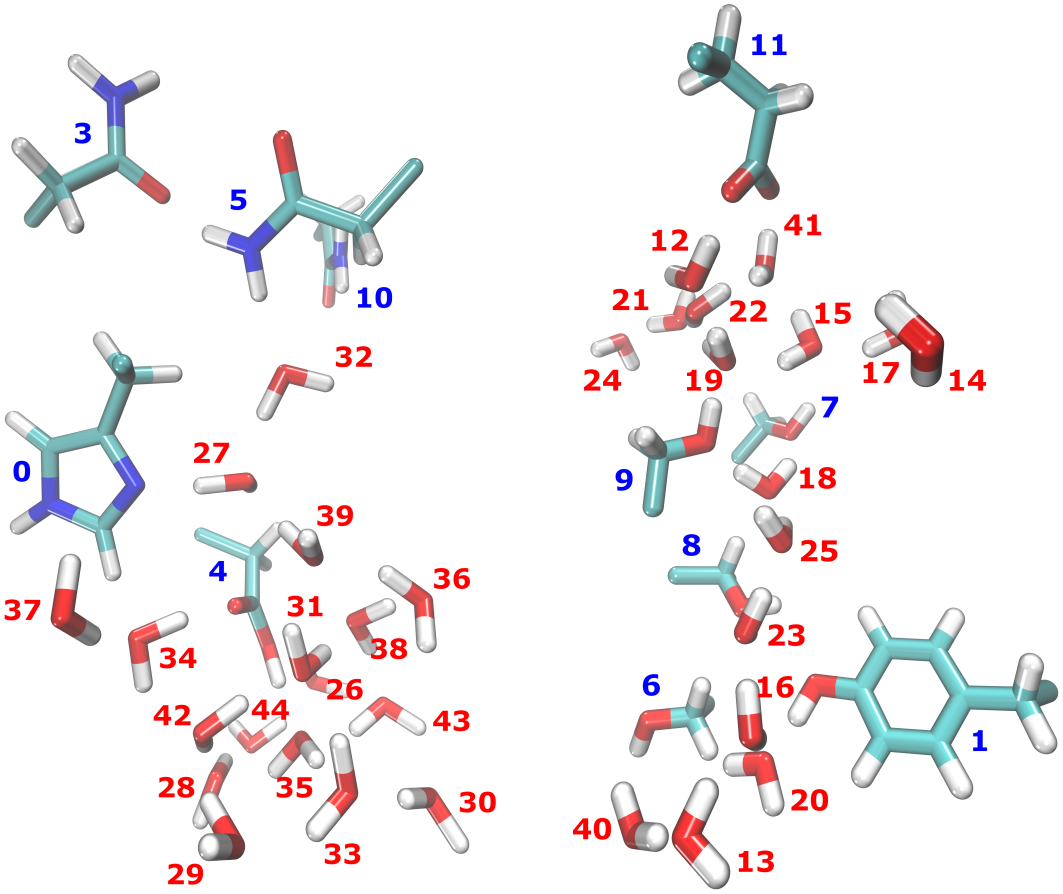


Figure 5.7: Flexible and charged region of the D-channel of *CcO* with residue numbers in blue and red for the different amino acids side chains and water molecules, respectively. Left: Lower part of the D-channel including D132, H26, N121, N139, N206, and 17 water molecules. Right: Upper part of the D-channel including Y33, S142, S197, S200, S201, E286, and 16 water molecules.

5.2 Results and Discussion III

The proton translocation along the D-channel from D132 to E286 was investigated by five *TN* calculations (cf. Tab. 5.2) using the *TN-MD method*. The combined *TN* contains 1 088 nodes connected by 13 394 edges. In this TN, each node is defined by 27 side chain dihedral angle conformations, a water pattern including the translocation and rotation of 33 water molecules, and a protonation state. All these *DOFs* are assigned in comparison to the reactant state of the overall proton translocation. Side chain dihedral angle rotations, water molecule translations, and water molecule rotations are discretized in 45° rotation steps around the side chain dihedral angle, 2.0 Å translation steps in x-, y-, and z-direction, and 22.5° rotation steps around the x-, y-, and z-axis, respectively. The 27 side chain dihedral angles

considered in the state assignment are listed in Tab. 5.3.

Table 5.3: The 27 side chain dihedral angles considered in the state assignment of the combined *TN*. Side chain dihedral angles displaying differences to the reatant state of the overall proton translocation in at least one node are indicated in bold.

Residue	Side chain dihedral angles		
H26	$\boldsymbol{\chi_1}$	$\boldsymbol{\chi_2}$	
Y33	χ_1	$\boldsymbol{\chi_2}$	$\boldsymbol{\chi_3}$
N121	χ_1	$\boldsymbol{\chi_2}$	χ_3
D132	$\boldsymbol{\chi_1}$	$\boldsymbol{\chi_2}$	
N139	$\boldsymbol{\chi_1}$	$\boldsymbol{\chi_2}$	$\boldsymbol{\chi_3}$
S142	χ_1	$\boldsymbol{\chi_2}$	
S197	χ_1	$\boldsymbol{\chi_2}$	
S200	$\boldsymbol{\chi_1}$	$\boldsymbol{\chi_2}$	
S201	χ_1	$\boldsymbol{\chi_2}$	
N206	$\boldsymbol{\chi_1}$	$\boldsymbol{\chi_2}$	$\boldsymbol{\chi_3}$
E286	χ_1	χ_2	$\boldsymbol{\chi_3}$

From the 27 side chain dihedral angles considered in the state assignment of the complete TN, 19 side chain dihedral angles display differences to the reactant state of the overall proton translocation in at least one node (cf. Tab. 5.3 bold), giving rise to 338 unique side chain dihedral angle combinations. The side chain dihedral angles displaying a maximal amount of differences in all nodes are, in decreasing order, $\chi_{2.N139}$, $\chi_{3.Y33}$, $\chi_{3.N139}$, and $\chi_{1.N139}$. Interestingly, the side chain dihedral angle $\chi_{3.Y33}$ displays the second most variation, despite its omission in the direct sampling of side chain dihedral angles. At this point, the observant reader may notice, that the side chain dihedral angle $\chi_{3.Y33}$, as well as the side chain dihedral angles $\chi_{3.N121}$, $\chi_{3.N139}$, $\chi_{2.S142}$, $\chi_{2.S197}$, $\chi_{2.S200}$, $\chi_{2.S201}$, and $\chi_{3.N206}$, used for the state assignment are not considered as usual side chain dihedral angles (cf. Tab. 3.1). Consequently, these side chain dihedral angles are defined in the following:

Table 5.4: Atoms involved in the sidechain dihedral angles of tyrosine, asparagine, and serine (AA, one letter code). Additional side chain dihedral angles are indicated in bold.

AA	χ_1	χ_2	χ_3	χ_4	χ_5
N	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma N_{\delta,1}$	$\mathbf{C_\beta C_\gamma N_{\delta,1} H_{\delta,1}}$	-	-
S	$NC_\alpha C_\beta O_\gamma$	$\mathbf{C_\alpha C_\beta O_\gamma H_\gamma}$	-	-	-
Y	$NC_\alpha C_\beta C_\gamma$	$C_\alpha C_\beta C_\gamma C_{\delta,1}$	$\mathbf{C_\epsilon C_\zeta O_\eta H_\eta}$	-	-

The side chain dihedral angle χ_3 of asparagine describes the rotation of its amide group. Due to the symmetry of the amide group such rotations are usually of less importance. In proton transfer reactions, however, slight rotations of the amide group could affect the existence of individual hydrogen bonds and thus the stability of larger hydrogen bond networks required for the transfer of protons. Thereby, the inclusion of the side chain dihedral angle χ_3 of asparagine in the state assignment is justified. The side chain dihedral angles χ_2 and χ_3 of serine and tyrosine, on the other hand, describe the rotation of their hydroxyl groups. Once again, such rotations are usually of less importance. In proton transfer reactions, however, slight rotations of the hydroxyl groups could affect the proton transfer indirectly, i.e. forming or breaking hydrogen bonds stabilizing or destabilizing larger hydrogen bond networks required for the transfer of protons, or directly, i.e. participating in an *HBC*. Thereby, the inclusion of the side chain dihedral angles χ_2 and χ_3 of serine and tyrosine in the state assignment is justified.

Apart from the side chain dihedral angles, the complete *TN* contains 527 unique water patterns and variations in the protonation state of residues 0, 1, 4, 11, 13, 16, 20, 26, 27, 29, 31, 33, 36, 38, 39, 41, 42, and 43. In Fig. 5.8 the six best proton transfer pathways along the D-channel are presented.

The optimal proton transfer pathway from D132 to E286 starts with two proton transfer steps from D132 (cf Fig. 5.9 left) to a water molecule in close proximity (residue 26) and further on to a second water molecule (residue 43). Both these steps occur without the rotation or re-orientation of side chain dihedral angles and water molecules, respectively. Then, following a re-orientation of some water molecules the excess proton is translocated towards H26 (cf. Fig. 5.9 right). All three transitions involve only minor relative transition barriers of 0 kcal/mol or -1 kcal/mol. Interestingly, however, with -3 kcal/mol or -4 kcal/mol the relative potential energy of states with the excess proton located at a water molecule or H26, respectively, is lower than

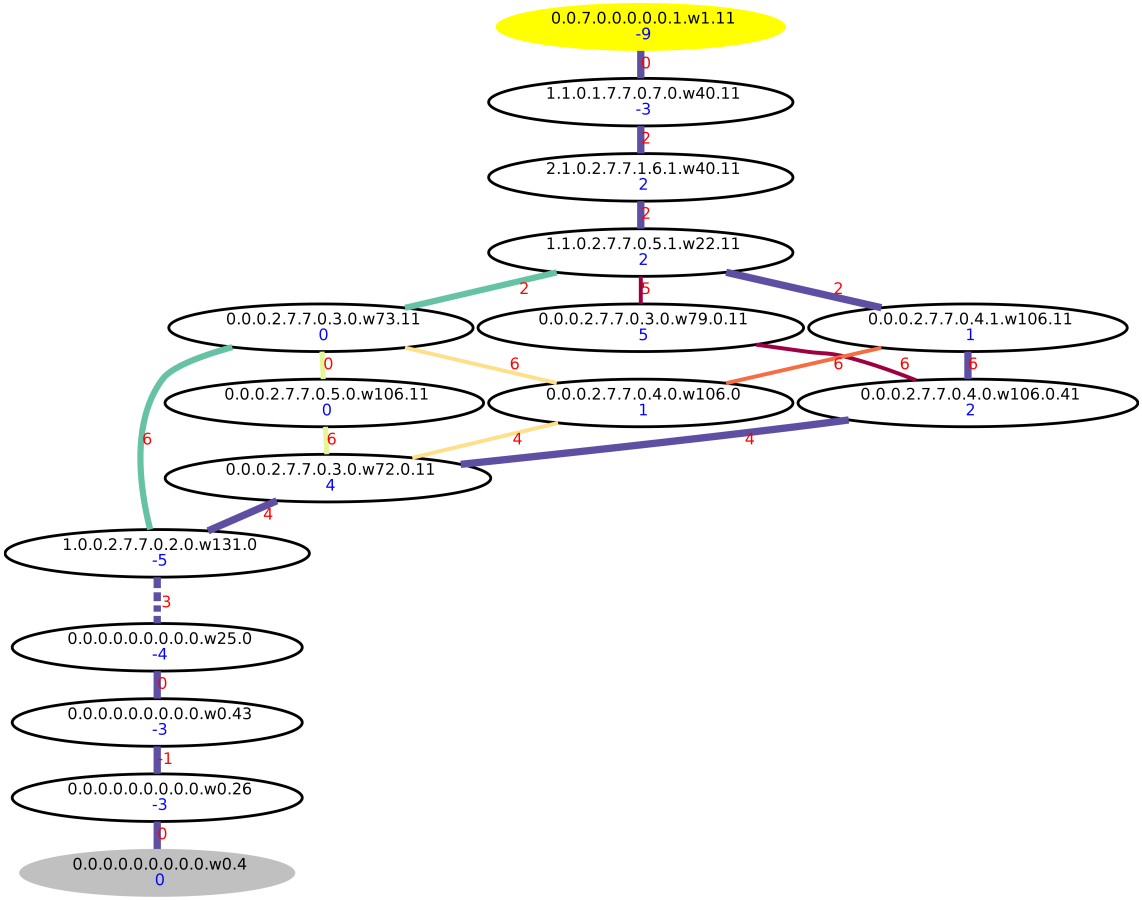


Figure 5.8: TN describing the translocation of a proton along a model system of the D-channel of *CcO*, from D132 at the D-channel entrance to E286 at the D-channel exit, following five individual *TN* calculations using the *TN-MD method*, i.e. direct sampling of side chain dihedral angles and protonation states and *MD* simulations for the motion of the water molecules (cf. Tab. 5.2 for a detailed description of either *TN* calculation). The nodes are shown as ellipses, labeled after a subset of *DOFs* ($\chi_{1.Y33}$·$\chi_{2.N121}$·$\chi_{2.D132}$·$\chi_{1.N139}$·$\chi_{2.N139}$·$\chi_{3.N139}$·$\chi_{2.S200}$·$\chi_{2.S201}$·$\chi_{3.E286}$·wi.P). The proton positions follow the residue numbering from Fig. 5.7. The ellipses filled in gray and yellow color correspond to the reactant and product state of the overall proton translocation, respectively. Edges are shown as lines. The dashed edge represents 18 sub-transitions (Appendix G). Blue numbers are the potential energies of the respective nodes, red numbers are the potential energies of the highest transition states along the edges used as edge weight. All energies are in kcal/mol, relative to the reactant state, and rounded to integer values for better readability. The *MBP* is bold and highlighted in indigo, the 2nd, 3rd, 4th, 5th, and 6th best pathways are bold and highlighted in turquoise, green, yellow, orange, and red, respectively.

the relative potential energy of the reactant state of the overall proton translocation with the excess proton located at D132. These relative potential energies indicate, once again, an unfavorable reactant state of the overall proton transfer reaction. As soon as the excess proton is located at H26, the side chain of N139 starts to open up,

thereby allowing the hydration of the asparagine region (cf. Fig. 5.10 left). These processes occur in concerted fashion along 16 states of the pathway, involving 16 different water patterns, clockwise rotations of χ_3 of Y33, χ_2 of N121, χ_1 of N139, χ_2 of N139, χ_3 of N139, and χ_2 of S200, and counter-clockwise rotations of χ_2 of N121, χ_1 of N139, χ_2 of N139, χ_3 of N139, and χ_2 of S200, while the protonation state is unchanged. The relative transition barriers of these transitions are ranging from -6 kcal/mol to 3 kcal/mol, with relative potential energies ranging from -7 kcal/mol to 0 kcal/mol. Hence, the maximal transition barrier of the opening of N139 and the hydration of the asparagine region is 3 kcal/mol. Following the formation of a continuous *HBC* through the asparagine region, in two steps of the pathway, the proton hole transfer from E286 to H26 is initiated. Therefore, a proton is transferred, in two steps of the pathway, from a cluster of water molecules beneath E286 towards E286 and stored at, or in close proximity to, E286 (cf. Fig. 5.10 right). Along these four transitions, χ_2 of S201 is rotated clockwise. The relative transition barriers of all four transitions are ranging form -1 kcal/mol to 4 kcal/mol with relative potential energies ranging from -5 kcal/mol to 4 kcal/mol. In a subsequent transition, the proton hole is transferred towards H26, and χ_3 of E286 is rotated clockwise (cf. Fig. 5.11 left). This transition involves a relative transition barrier of 6 kcal/mol and is thus the rate-determining, maximal transition barrier of the proton translocation along the D-channel of CcO. Finally, the asparagine region is dehydrated and N139 rotates into the closed conformation to reach the product state of the overall proton translocation in four steps of the pathway (cf. Fig. 5.11 right). The closing of the D-channel involves relative transition barriers ranging from 0 kcal/mol to 2 kcal/mol with relative potential energies ranging from -9 kcal/mol to 2 kcal/mol.

Comparing the best proton transfer pathway from Fig. 5.8 with the initial proton transfer pathway it can be noticed that the rate-determining, maximal transition barrier of the overall proton translocation along the D-channel is reduced from $\approx$ 9 kcal/mol to $\approx$ 6 kcal/mol. Hence, the determined transition barrier is significantly lower than the transition barrier derived by *Liang et al* [206]. Furthermore, the low transition barrier of 3 kcal/mol determined for the opening and hydration of the asparagine region, is in conflict with results from *Bagherpoor et al* [455], who performed eight microseconds of *MD* simulations of the *CcO* complex without observing an open and hydrated asparagine region. Interestingly, however, the difference in rate-determining, maximal transition barriers, as well as the difference in the energy difference of reactant and product state of the overall proton transloca-

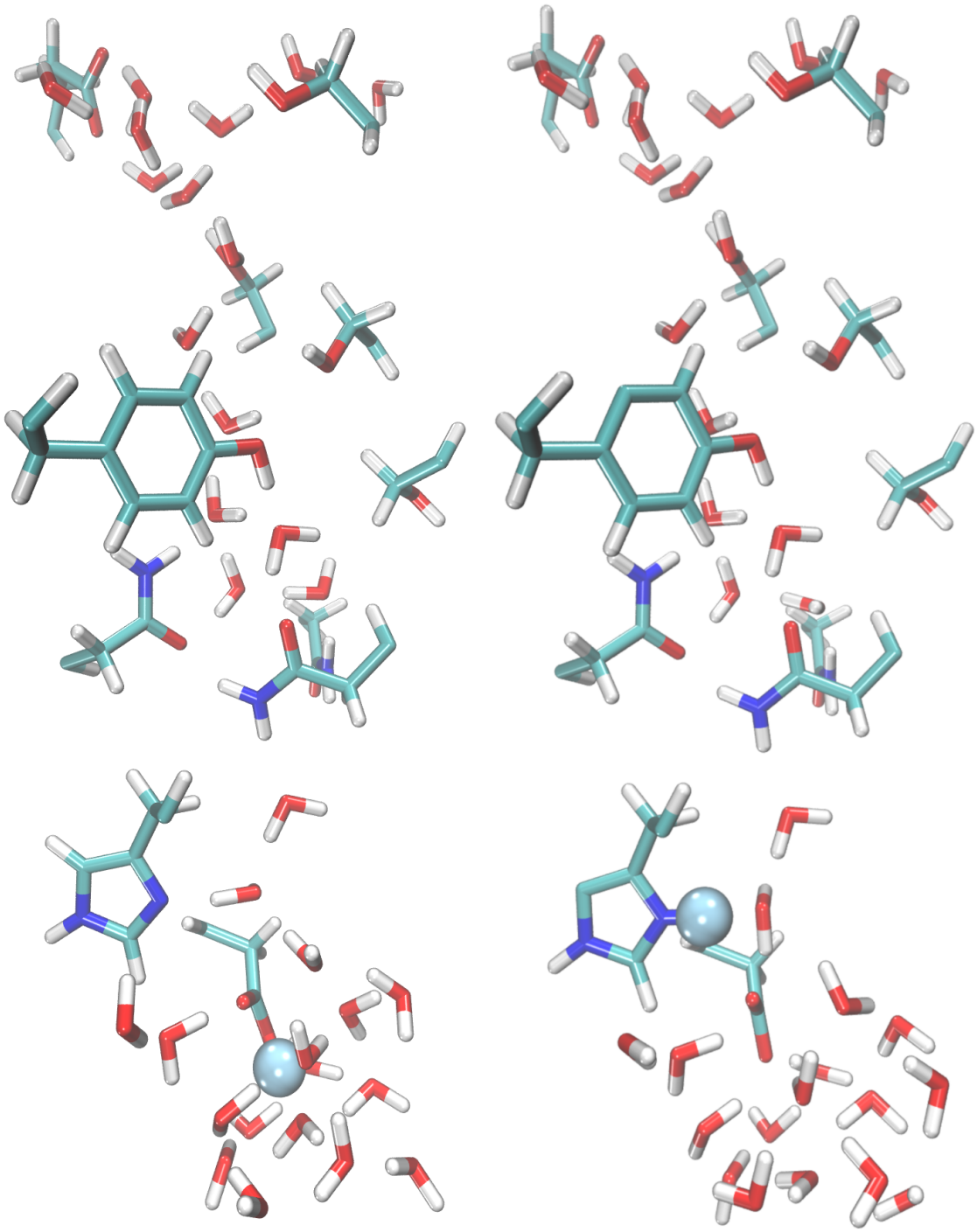


Figure 5.9: Left: Reactant state of the overall proton translocation along the D-channel of *CcO*. State 1, 0.0.0.0.0.0.0.0.0.w0.4, of the *MBP* from Fig 5.8. Right: Intermediate state of the proton translocation along the D-channel of *CcO*. State 4, 0.0.0.0.0.0.0.0.0.w25.0, of the *MBP* from Fig 5.8.

tion, both compared to *Liang et al* [206], correspond to $\approx$ 3 kcal/mol. Assuming that the *TN* calculations of *CcO*'s D-channel, performed within the framework of this thesis, are based on an unfavorable reactant state of the overall proton translo-

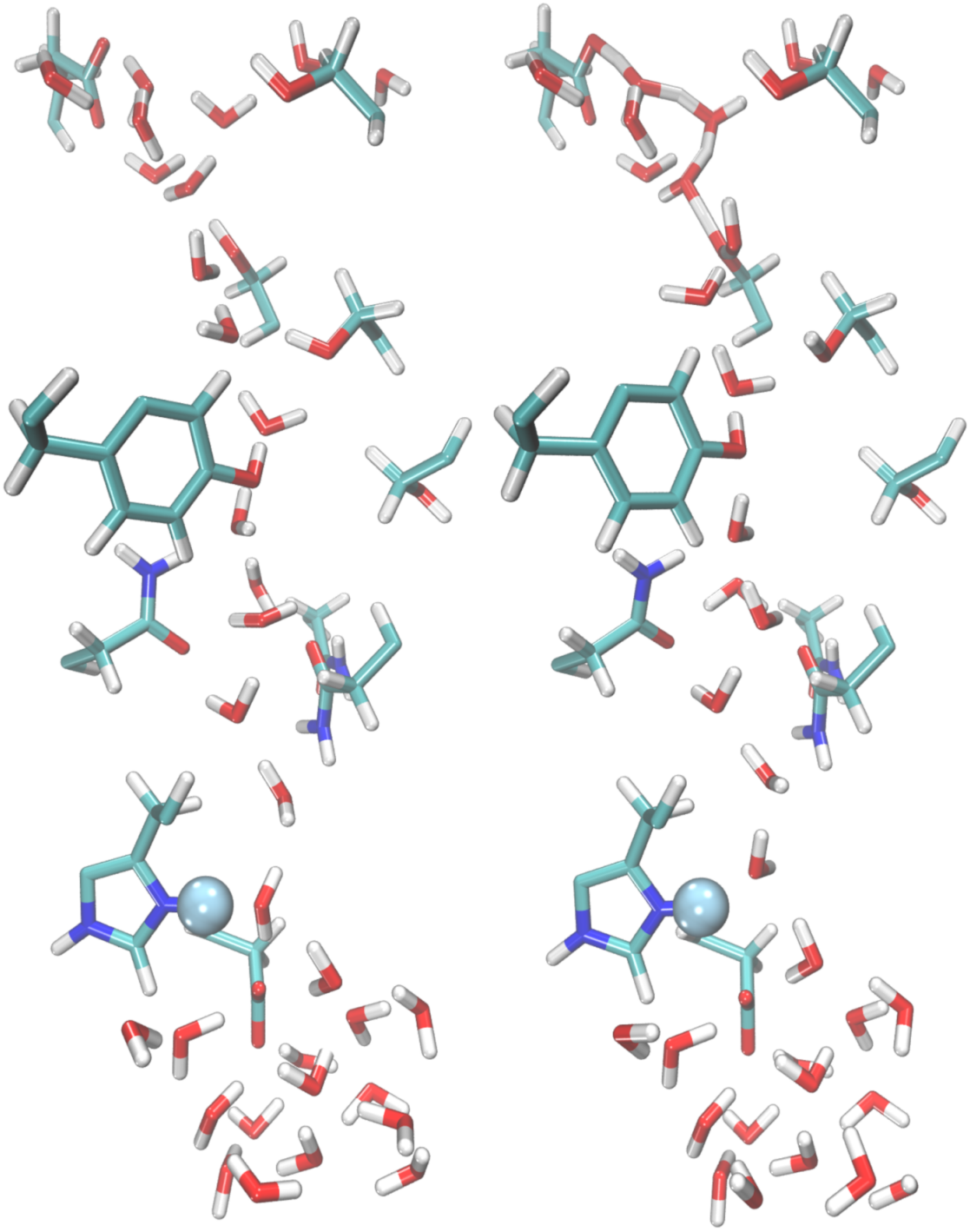


Figure 5.10: Left: Intermediate state of the proton translocation along the D-channel of *CcO*. State 20, 3.0.0.2.7.7.1.0.0.w22.0, of the *MBP* from Fig. 5.8. Right: Intermediate state of the proton translocation along the D-channel of *CcO*. State 24, 0.0.0.2.7.7.0.4.0.w106.0.41, of the *MBP* from Fig. 5.8.

cation, an increase in both barriers can be imagined for an optimal reactant state of the overall proton translocation. If the energy difference between the reactant and product state of the overall proton translocation, proposed by *Liang et al* [206],

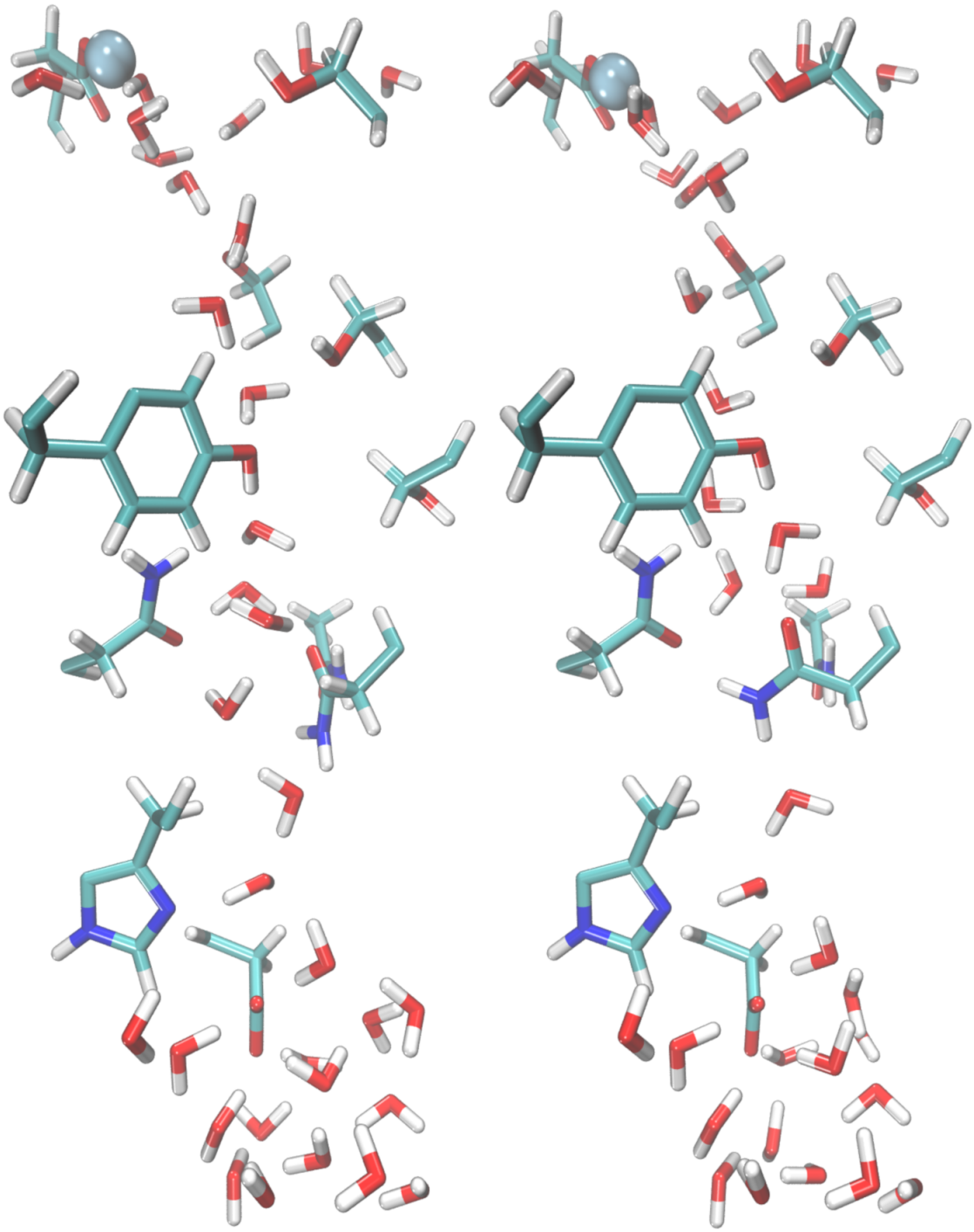


Figure 5.11: Left: Intermediate state of the proton translocation along the D-channel of *CcO*. State 25, 0.0.0.2.7.7.0.4.1.w106.11, of the *MBP* from Fig. 5.8. Right: Product state of the proton translocation along the D-channel of *CcO*. State 29, 0.0.7.0.0.0.0.0.1.w1.11, of the *MBP* from Fig. 5.8.

is assumed to be correct, than the rate-determining, maximal transition barrier of the overall proton translocation, determined by the *TN-MD method*, is in complete accordance to the results from *Liang et al* [206]. A similar increase in the transition

barrier of the opening and hydration of the asparagine region would also resolve the conflict with the results from *Bagherpoor et al* [455]. Due to the reduced size and parametrization of the model system of the D-channel, used for the *TN-MD* calculations, an overestimation of the potential energy of all states investigated is likely. This is especially true for the reactant state of the overall proton translocation along the D-channel, where the greater surrounding is believed to assist the proton uptake via D132 by increasing its pK_A value [460].

A further conflict of the initial proton transfer pathway with the results from *Liang et al* [206] was the conformation of the transition state corresponding to the rate-determining, maximal transition barrier, i.e. the midpoint of the opening and hydration of the asparagine region for the initial proton transfer pathway and a state with an hydronium ion in between the asparagines for the pathway proposed by *Liang et al* [206]. The *TN-MD method* shifts the transition state from the midpoint of the opening and hydration of the asparagine region to the final proton transfer step. In Fig. 5.12 a detailed representation of the transition state corresponding to the rate-determining, maximal transition barrier of the overall proton translocation is provided.

According to Fig. 5.12 the transition state corresponding to the rate-determining, maximal transition barrier of the overall proton translocation contains, according to the *TN-MD method*, an hydroxide ion in between the asparagines instead of an hydronium ion as proposed by *Liang et al* [206]. More generally, however, the transition state is defined in both pathways by a charged species in between the asparagines. The identity of the charged species is most likely determined by the mechanism underlying the proton translocation, i.e. proton transport from D132 to E286 along the optimal proton transfer pathway from *Liang et al* [206] or proton hole transport from E286 to H26 along the optimal proton transfer pathway determined by the *TN-MD method*. Consequently, the optimal proton transfer pathway determined by the *TN-MD method* in a minimal D-channel model system is reproducing the most important aspects of the pathway proposed by *Liang et al* [206], which is remarkable considering the fundamental difference in the methods used for the analysis. The confirmation of experimental results using a wide range of techniques is at the heart of science and especially important for the acceptance of computational studies.

Apart from the determination of the optimal proton transfer pathway, the *TN-MD method* also provides a range of next best proton transfer pathways (cf. Fig. 5.8). In Tab. 5.6 the number of optimal proton transfer pathways along the D-

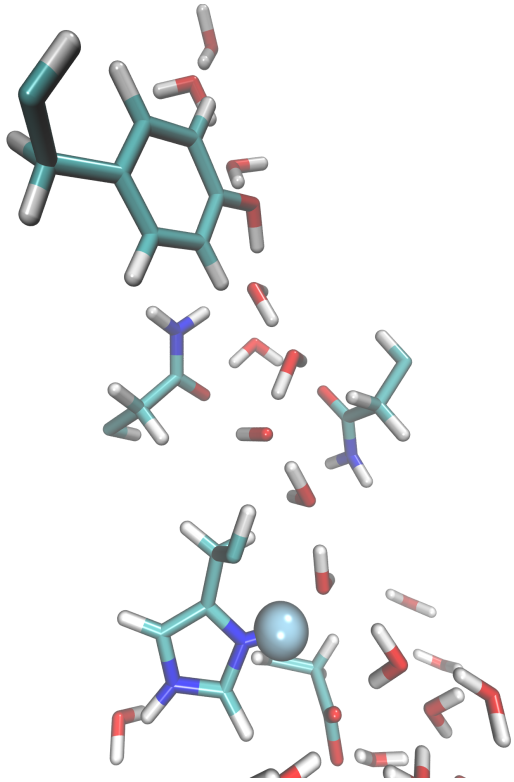


Figure 5.12: Conformation of the transition state corresponding to the rate-determining, maximal transition barrier along the optimal proton transfer pathway, according to the *TN-MD* calculations, connecting D132 at the D-channel entrance and E286 at the D-channel end.

channel is summarized up to a relative transition barrier of 20 kcal/mol.

Table 5.6: Number of proton transfer pathways along the D-channel of *CcO* for different rate-determining, maximal transition barriers.

Rel. transition barrier in kcal/mol	No. of pathways
6	21
7	9
8	3
9	0
10	2
11	2
12	1
13	6
Up to 20	53

The *TN-MD method* allows the determination of 21 distinct proton transfer path-

ways with rate-determining, maximal transition barriers of 6 kcal/mol and 32 additional proton transfer pathways up to rate-determining, maximal transition barriers of 20 kcal/mol. These pathways differ in their side chain dihedral angle rotations, their water patterns, and their protonation states (e.g. cf. Fig. 5.8). Hence, the *TN-MD method* is reproducing the multiplicity of transition pathways expected for complex transitions.

Chapter 6

Conclusion and Outlook

The previous chapters of this thesis provided a thorough introduction of various proton transfer reactions, along with their outstanding biological and chemical relevance as well as consequences arising from malfunctions. Furthermore, the *TN* approach was introduced, and subsequently refined by the *TN-MD method* and the *TN prediction method*, to allow an unbiased investigation of proton transfer reactions along the channels of proton pumps, in particular the D-channel of *CcO*.

At the beginning of this thesis an objective was set, namely, providing another piece of structure-function analysis to cure *oxidative phosphorylation* related diseases like Alzheimer. How many people will be cured from such diseases, on the spot, following the publication of this thesis? No one. How many people will discuss the additional insight, concerning the structure and function of the D-channel, following the publication of this thesis? Few, at most. So, what does this thesis actually provide? This thesis provides an opportunity. The opportunity to see and investigate the world of proton transfer from a different perspective, a discretized perspective. The sheer amount of *DOFs* involved in the transfer of protons along the channels of proton pumps is challenging, rendering biased investigations in most cases an unwanted necessity. Therefore, in many cases, different parts of a proton transfer system are treated with different precision levels, the proton transfer system is limited to different pre-defined transitions, or the proton transfer system is driven along a pre-defined direction. Consequently, the outcome of such investigations depends, more or less, on the knowledge gathered before. While this may not affect the determination of the optimal transition pathway, it will most likely affect the determination of alternative transition pathways, especially if they are mechanistically diverse. Thereby, a comprehensive picture of proton transfer reactions, including all relevant proton transfer pathways, is often out of reach. The *TN* ap-

proach, invented more than ten years ago, addressed this problem by including and investigating all relevant *DOFs* of a proton transfer reaction in a discretized manner, e.g. protonating all protonatable sites, rotating all side chain dihedral angles with a certain step size, and moving all water molecules with a certain step size. Thereby, the dependence on previously gathered knowledge is reduced, which allows a bias-free investigation of the proton transfer reaction. This holds at least if the set of *DOFs* is chosen properly. A minimal set of *DOFs* to sample can be derived by comparing the reactant and product state of the overall proton translocation in terms of internal coordinates, e.g. dihedral angles. Internal coordinates displaying differences need to be included in the sampling. Additional DOFs to sample can be derived by the sampling of initial proton transfer pathways using common transition pathway finding algorithms, as well as correlation, communication and hydrogen bond analysis on existing *MD* simulation data. Following their sampling, the resulting discrete structures are minimized, assigned a state, connected by *MEPs* and ultimately combined to a weighted, simple graph (or *TN*), which can be analyzed using standard graph theoretical algorithms. Early on, however, it had been noticed that the undeniable beauty of the *TN* approach, is, at the same time, its biggest drawback, rendering the investigation of infeasible amounts of discrete states an unwanted necessity. Consequently, for a feasible *TN* calculation, either the set of sampled *DOFs* needs to be chosen inappropriately small, or the step size along the sampling needs to be chosen inappropriately high. Obviously, both requirements prohibit a comprehensive analysis of proton transfer reactions.

The *TN-MD method*, developed within the framework of this thesis, attenuates this drawback, combining the benefits of the *TN* approach, i.e. the complete sampling of individual *DOFs* irrespective of barrier heights, with the benefits of the *MD* simulation, i.e. the generation of a sequence of reasonable states of individual *DOFs*. By focusing the *TN* approach on *DOFs*, giving rise to a minor number of states separated by substantial transition barriers, e.g. side chain dihedral angle rotations, and applying the *MD* simulations to *DOFs*, giving rise to a larger number of states separated by lower transition barriers, e.g. water motions, the *TN* calculation costs can be reduced significantly. *TN-MD* calculations in a small model channel for the transfer of protons confirmed the cost reductive effect. At the same time, these *TN-MD* calculations maintained the expected variability of proton transfer pathways, and even displayed completely distinct proton transfer mechanisms, i.e. concerted or step wise proton transfer reactions along the model channel. A different approach to reduce the number of *DOFs* to investigate within

a single *TN* calculation is the indirect inclusion of certain *DOFs*, i.e. calculating several *TNs* with different configurations of unsampled *DOFs*. An indirectly sampled protonation state, for example, requires two distinct *TN* calculations with or without the additional proton. The different configurations of unsampled *DOFs* can be seen as a perturbation of an initial configuration of the unsampled *DOFs*, which gives rise to the assumption that certain features of the *TNs* are conserved over the perturbations and do not need to be re-calculated. Conserved features, which proved to be efficient are a coarse-grained representation of the initial *TN*, according to the transition barriers used as edge weight, the initial *MST*, and a corresponding sensitivity analysis. Furthermore, in most cases, the determination of specific properties of the perturbed *TNs*, e.g. the rate-determining, maximal transition barrier of the perturbed *MBP*, is enough to decide if a thorough investigation of the perturbed system is required. The combination of both these aspects within the *TN* prediction method provided calculation cost reductions of the perturbed *TNs* by up to 80 % for proton transfer reactions in a small model channel with an unsampled point charge in close vicinity. In addition to the costs reductions, the *TN* prediction method maintained important properties of the perturbed *TNs* in almost all *TN* predictions.

Apart from the method development, the proton transfer along the D-channel of *CcO* has been investigated using the *TN-MD method* and a reduced D-channel model system. Thereby different aspects of previous analyses concerning the D-channel itself, as well as the proton transfer along the D-channel, have been confirmed, e.g. rate-determining transition barrier of the optimal proton transfer pathway, or extended, e.g. the flipping behavior of the asparagine gate for different protonation states. The confirmation (and extension) of experimental results, using a wide range of techniques, is at the heart of science and building the foundation of succeeding investigations, which will ultimately lead to curative treatments of *oxidative phosphorylation* related diseases.

At the same time, however, the *TN-MD* calculations along the D-channel revealed the necessity of further costs reductions, due to infeasible high numbers of discrete states required for the sampling of different side chain dihedral angle conformations and protonation states. Within the framework of this thesis this was accomplished by *TN-MD* calculations along an initial proton transfer pathway, which inevitably introduces a bias in the resulting *TN*. Therefore, additional methods need to be developed which reduce the costs of the *TN* approach for large numbers of *DOFs* which require a discrete sampling. For the conformation of side chain dihedral

angles, for example, it is imaginable to sort out states by *MM* potential energy minimizations preceding the actual *QM* potential energy minimizations. Furthermore, an initial investigation of different protonation states might be accomplished by flexible *QM* regions surrounding the protonation state of interest. Both proposals, obviously, require an elaborated testing.

In conclusion, the methods developed within the framework of this thesis, provide a big leap towards an efficient and bias-free investigation of complex proton transfer reactions. However, as mentioned before further method developments are required.

Bibliography

[1] GM Cooper and RE Hausman. An overview of cells and cell research. In *The Cell: A Molecular Approach, Sixth Edition*, chapter 1, pages 3–42. Sinauer Associates, Sunderland, Ma, 2013.

[2] C Woese, O Kandler, and M Wheelis. Towards a natural system of organisms: Proposal for the domains archea, bacteria, and eucarya. *PNAS*, 87:4576–4579, 1990.

[3] JA Lake, E Henderson, M Oakes, and MW Clark. Eocytes: A new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *PNAS*, 81:3786–3790, 1984.

[4] E Mayr. Two empires or three? *PNAS*, 95:9720–9723, 1998.

[5] T Cavalier-Smith. Only six kingdoms of life. *Proc. R. Soc. B*, 271:1251–1262, 2004.

[6] JM Archibald. The eocyte hypothesis and the origin of eukaryotic cells. *PNAS*, 105:20049–20050, 2008.

[7] TA Williams, PG Foster, CJ Cox, and TM Embley. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*, 504:231–236, 2013.

[8] A Nasir, KM Kim, G Caetano-Anolles. Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms archaea, bacteria, and eukarya. *BMC Evol. Biol.*, 12:156, 2012.

[9] S Luketa. New views on the megaclassification of life. *Protistology*, 7:218–237, 2012.

[10] AJ Roger and AGB Simpson. Evolution: Revisiting the root of the eukaryote tree. *Current Biology*, 19:165–167, 2009.

[11] MA Ruggiero, DP Gordon, TM Orrell, N Bailly, T Bourgoin, RC Brusca, T Cavalier-Smith, MD Guiry, PM Kirk, EV Thuesen. A higher level classification of all living organisms. *PLOS ONE*, 10, 2015.

[12] B Alberts, A Johnson, J Lewis, M Raff, K Roberts, and P Walter. Cells and genomes. In *Molecular biology of the Cell, Fourth Edition*, chapter 1, pages 3–46. Garland Science, New York, 2002.

[13] K Lohmann. Über die pyrophosphatfraktion im muskel. *Naturwissenschaften*, 17:624–625, 1929.

[14] CH Fiske, Y Subbarow. Phosphocreatine. *J. Biol. Chem.*, 81:629–679, 1929.

[15] GD Novelli. Amino acid activation for protein synthesis. *Annual Review of Biochemistry*, 36:449–484, 1967.

[16] N Mishra, R Tuteja, and N Tuteja. Signaling through map kinase networks in plants. *Archives of Biochemistry and Biophysics*, 452:55–68, 2006.

[17] G Dahl. Atp release through pannexon channels. *Philosophical Transactions of the Royal Society B: Biological Science*, 370, 2015.

[18] DW Rogers. The gibbs free energy. In *Concise Physical Chemistry*, chapter 6, pages 84–92. John Wiley and Sons, Inc., Hoboken, NJ, 2010.

[19] JJ Zimmermann, A von Saint AndrÚ-von Arnim, and J McLaughlin. Cellular respiration. In *Pediatric Critical Care, Fourth Edition*, chapter 74, pages 1058–1072. Mosby Inc., Philadelphia, PA, 2011.

[20] B Alberts, A Johnson, J Lewis, M Raff, K Roberts, and P Walter. Cell chemistry and biosynthesis. In *Molecular biology of the Cell, Fourth Edition*, chapter 2, pages 47–128. Garland Science, New York, 2002.

[21] SM Houten and RJA Wanders. A general introduction to the biochemistry of mitochondrial fatty acid beta-oxidation. *J Inherit Metab Dis*, 33:469–477, 2010.

[22] M Pester, K-H Knorr, MW Friedrich, M Wagner, and A Loy. Sulfate-reducing microorganisms in wetlands - fameless actors in carbon cycling and climate change. *Frontiers in Microbiology*, 3:72, 2012.

[23] H Lodish, A Berk, SL Zipursky, P Matsudaira, D Baltimore, and J Darnell. Electron transport and oxidative phosphorylation. In *Molecular Cell Biology, Fourth Edition*, chapter 16. W. H. Freeman, New York, 2000.

[24] GM Cooper and RE Hausman. Bioenergetics and metabolism. In *The Cell: A Molecular Approach, Sixth Edition*, chapter 11, pages 421–458. Sinauer Associates, Sunderland, Ma, 2013.

[25] W Martin, M Hoffmeister, C Rotte, and K Henze. An overview of endosymbiotic models for the origins of eukaryotes, their atp-producing organelles (mitochondria and hydrogenosomes) and their heterotrophic lifestyle. *Biol. Chem.*, 382:1521–1539, 2001.

[26] M Wikström, V Sharma, VRI Kaila, JP Hosler, and G Hummer. New perspectives on proton pumping in cellular respiration. *Chemical Reviews*, 115:2196–2221, 2015.

[27] B Alberts, A Johnson, J Lewis, M Raff, K Roberts, and P Walter. Energy conversion: Mitochondria and chloroplasts. In *Molecular biology of the Cell, Fourth Edition*, chapter 16, pages 767–830. Garland Science, New York, 2002.

[28] P Mitchell. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature*, 191:144–148, 1961.

[29] S DiMauro and P Rustin. A critical approach to the therapy of mitochondrial respiratory chain and oxidative phosphorylation diseases. *Biochimica et Biophysica Acta*, 1792:1159–1167, 2009.

[30] Rolf Luft, Denis Ikkos, Genaro Palmieri, Lars Ernster, and Björn Afzelius. A case of severe hypermetabolism of nonthyroid origin with a defect in the maintenance of mitochondrial respiratory control: a correlated clinical, biochemical, and morphological study. *The Journal of Clinical Investigation*, 41:1776–1804, 1962.

[31] DC Wallace. Mitochondrial diseases in man and mouse. *Science*, 283:1482–1488, 1999.

[32] M Walker, RW Taylor, and DM Turnbull. Mitochondrial diabetes. *Diabetic Medicine*, 22:18–20, 2005.

[33] M Hüttemann, I Lee, A Pecinova, P Pecina, K Przyklenk, and JW Doan. Regulation of oxidative phosphorylation, the mitochondrial membrane potential, and their role in human disease. *Journal of Bioenergetics and Biomembranes*, 40:445–456, 2008.

[34] JA Smeitink, M Zeviani, DM Turnbull, and HT Jacobs. Mitochondrial medicine: A metabolic perspective on the pathology of oxidative phosphorylation disorders. *Cell Metabolism*, 3:9–13, 2006.

[35] AHV Schapira. Human complex i defects in neurodegenerative diseases. *Biochimica et Biophysica Acta*, 1364:261–270, 1998.

[36] WD Parker, CM Filley, and JK Parks. Cytochrome oxidase deficiency in alzheimer's disease. *Neurology*, 40:1302, 1990.

[37] W Balemans, L Vranckx, N Lounis, O Pop, J Guillemont, K Vergauwen, S Mol, R Gillisen, M Motte, D Lançois, M De Bolle, K Bonroy, H Lill, K Andries, D Bald, and A Koul. Novel antibiotics targeting respiratory atp synthesis in gram-positive pathogenic bacteria. *Antimicrobial Agents and Chemotherapy*, 56:4131–4139, 2012.

[38] A Pérez, A Noy, F Lankas, FJ Luque, and M Orozco. The relative flexibility of b-dna and a-rna duplexes: database analysis. *Nucleic Acids Research*, 32:6144–6151, 2004.

[39] A Leo-Macias, P Lopez-Romero, D Lupyan, D Zerbino, and AR Ortiz. An analysis of core deformations in protein superfamilies. *Biohysical Journal*, 88:1291–1299, 2005.

[40] C Micheletti. Comparing proteins by their internal dynamics: Exploring structure-function relationships beyond static structural alignments. *Physics of Life Reviews*, 10:1–26, 2013.

[41] N Ahalawat and RK Murarka. Conformational changes and allosteric communications in human serum albumin due to ligand binding. *Journal of Biomolecular Structure and Dynamics*, 33:2192–2204, 2015.

[42] A Hospital, JR Gõni, M Orozco, and JL Gelpi. Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry*, 8:37–47, 2015.

[43] RO Dror, RM Dirks, JP Grossman, H Xu, and DE Shaw. Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics*, 41:429–452, 2012.

[44] JA McCammon, BR Gelin, and M Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.

[45] M Karplus and JA McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9:646–652, 2002.

[46] LV Bock, GF Schröde, I Davydo, N Fischer, H Stark, MV Rodnina, AC Vaiana, and H Grubmüller. Energy barriers and driving forces in trna translocation through the ribosome. *Nature Structural & Molecular Biology*, 20:1390–1396, 2013.

[47] PC Whitford, SC Blanchard, JHD Cate, and KY Sanbonmatsu. Connecting the kinetics and energy landscape of trna translocation on the ribosome. *PLoS Computational Biology*, 9, 2013.

[48] M Chavent, AL Duncan, and MSP Sansom. Molecular dynamics simulations of membrane proteins and their interactions: from nanoscale to mesoscale. *Current Opinion in Structural Biology*, 40:8–16, 2016.

[49] J Warnau, V Sharma, AP Ganiz-Hernandez, A Di Luca, O Haapanen, I Vattulainen, M Wikström, G Hummer, and VRI Kaila. Redox-coupled quinone dynamics in the respiratory complex i. *PNAS*, 2018.

[50] L Euro, O Haapanen, T Ròg, I Vattulainen, A Suomalainen, and V Sharma. Atomistic molecular dynamics simulations of mitochondrial dna polymerase gamma: Novel mechanisms of function and pathogenesis. *Biochemistry*, 56:1227–1238, 2017.

[51] B Pandey, A Grover, and P Sharma. Dynamics of dof domain-dna interaction in wheat: Insights from atomistic simulations andfree energy landscape. *Journal of Cellular Biochemistry*, pages 1–12, 2018.

[52] DE Shaw, RO Dror, JK Salmon, JP Grossman, KM Mackenzie, JA Bank, C Young, MM Deneroff, B Batson, KJ Bowers, E Chow, MP Eastwood, DJ Ierardi, JL Klepeis, JS Kuskin, RH Larson, K Lindorff-Larsen, P Maragakis, MA

Moraes, S Piana, Y Shan, and B Towles. Millisecond-scale molecular dynamics simulations on anton. *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, 2009.

[53] J Zhang, F Xu, Y Hong, Q Xiong, and J Pan. A comprehensive review on the molecular dynamics simulation of the novel thermal properties of graphene. *RSC Advances*, 5:89415–89426, 2015.

[54] JL Sussman, D Lin, J Jiang, NO Manning, J Prilusky, O Ritter, and EE Abola. Protein data bank (pdb): Database of three-dimensional structural information of biological macromolecules. *Acta Crystallographica Section D*, 54:1078–1084, 1998.

[55] JG Almeida, AJ Preto, PI Koukos, AMJJ Bonvin, and IS Moreira. Membrane proteins structures: A review on computational modeling tools. *Biochimica et Biophysica Acta - Biomembranes*, 1859:2021–2039, 2017.

[56] DA Case et al. Amber18. 2018.

[57] C Oostenbrink, A Villa, AE Mark, WF van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The gromos force-field parameter sets 53a5 and 53a6. *Journal of Computational Chemistry*, 25:1656–1676, 2004.

[58] N Foloppe, and AD MacKerrell Jr. All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of Computational Chemistry*, 21:86–104, 2000.

[59] MP Allen. Introduction to molecular dynamics simulation. In *Computational Soft Matter: From Synthetic Polymers to Proteins*, chapter 1, pages 1–28. John von Neumann Institute for Computing, Jülich, 2004.

[60] J Wang, RM Wolf, JW Caldwell, PA Kollman, and DA Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25:1157–1164, 2004.

[61] JD Durrant and JA McCammon. Molecular dynamics simulations and drug discovery. *BMC Biology*, 9, 2011.

[62] JD Chodera, DL Mobley, MR Shirts, RW Dixon, K Branson, and VS Pande. Alchemical free energy methods for drug discovery: Progress and challenges. *Current Opinion in Structural Biology*, 21:150–160, 2011.

[63] PS Nerenberg and T Head-Gordon. New developments in force fields for biomolecular simulations. *Current Opinion in Structural Biology*, 49:129–138, 2018.

[64] TP Senftle, S Hong, MM Islam, SB Kylasa, Y Zheng, YK Shin, C Junkermeier, R Engel-Herbert, MJ Janik, HM Aktulga, T Verstraelen, A Grama, and ACT van Duin. Exploring the role of the third active site metal ion in dna polymerase with qm/mm free energy simulations. *npj Computational Materials*, 2, 2016.

[65] P Cieplak, J Caldwell, and P Kollman. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and n-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *Journal of Computational Chemistry*, 22:1048–1057, 2001.

[66] CM Baker, PEM Lopes, X Zhu, B Roux, and AD MacKerell Jr. Accurate calculation of hydration free energies using pair-specific lennard-jones parameters in the charmm drude polarizable force field. *Journal of Chemical Theory and Computation*, 6:1181–1198, 2010.

[67] JW Ponder, C Wu, P Ren, VS Pande, JD Chodera, MJ Schnieders, I Haque, DL Mobley, DS Lambrecht, RA DiStasio Jr., M Head-Gordon, GNI Clark, ME Johnson, and T Head-Gordon. Current status of the amoeba polarizable force field. *Journal of Physical Chemistry B*, 114:2549–2564, 2010.

[68] TR Lucas, BA Bauer, and S Patel. Charge equilibration force fields for molecular dynamics simulations of lipids, bilayers, and integral membrane protein systems. *Biochimica et Biophysica Acta - Biomembranes*, 1818:318–329, 2012.

[69] J-P Martinez. The hartree-fock method: from self-consistency to correctsymmetry. *annalen der physik*, 529, 2016.

[70] RO Jones. Density functional theory: Its origins, rise to prominence, and future. *Reviews of Modern Physics*, 87, 2015.

[71] MJS Dewar, EG Zoebisch, EF Healy, and JJP Stewart. Ami: A new general purpose quantum mechanical molecular model1. *Journal of the American Chemical Society*, 107:3902–3909, 1985.

[72] W Weber and W Thiel. Orthogonalization corrections for semiempirical methods. *Theoretical Chemistry Accounts*, 103:495–506, 2000.

[73] DR Stevens and S Hammes-Schiffer. Exploring the role of the third active site metal ion in dna polymerase with qm/mm free energy simulations. *Journal of the American Chemcial Society*, 140:8965–8965, 2018.

[74] T Maximova, R Moffatt, B Ma, R Nussinov, and A Shehu. Principles and overview of sampling methods for modeling macromolecular structure and dynamics. *PLOS Computational Biology*, 2016.

[75] Some accelerated/enhanced sampling techniques.

[76] SJ Marrik and DP Tieleman. Perspective on the martini model. *Chemical Society Reviews*, 42:6801–6822, 2013.

[77] J Barnoud and L Monticelli. Coarse-grained force fields for molecular simulations. In *Molecular Modeling of Proteins. Methods in Molecular Biology (Methods and Protocols)*, chapter 7, pages 125–149. Humana Press, New York, 2015.

[78] J Kästner. Umbrella sampling. *WIREs Computational Molecular Science*, 1:932–942, 2011.

[79] H Grubmüller. Predicting slow structural transitions in macromolecular systems: conformational flooding. *Physical Review E*, 52:2893–2906, 1995.

[80] S Park and K Schulten. Calculating potentials of mean force from steered molecular dynamics simulations. *Chemical Physics*, 120:5946–5961, 2004.

[81] E Darve and A Pohorille. Calculating free energies using average force. *Journal of Chemical Physics*, 115:9169–9183, 2001.

[82] A Barducci, M Bonomi, and M Parrinello. Metadynamics. *WIREs Computational Molecular Science*, 1:826–843, 2011.

[83] Y Sugita, A Kitao, and Y Okamoto. Multidimensional replica-exchange method for free-energy calculations. *The Journal of Chemical Physics*, 113:6042–6051, 2011.

[84] SG Itoh, A Damjanović, and BR Brooks. ph replica-exchange method based on discrete protonation states. *Proteins - Structure, Function, Bioninformatics*, 79:3420–3439, 2011.

[85] M Meli and G Colombo. A hamiltonian replica exchange molecular dynamics (md) method for the study of folding, based on the analysis of the stabilization determinants of proteins. *International Journal of Molecular Sciences*, 14:12157–12169, 2013.

[86] W Quapp. Analysis of the concept of minimum energy path on the potential energy surface of chemically reacting systems. *Theoretical Chemistry Accounts*, 66:245–260, 1984.

[87] S Fischer and M Karplus. Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states with many degrees of freedom. *Chemical Physics Letters*, 194:252–261, 1992.

[88] H Jónsson, G Mills, and KW Jacobsen. Nudged elastic band method for finding minimum energy paths of transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*, chapter 16, pages 385–404. World Scientific, Lerici, 1998.

[89] H Eyring. The activated complex in chemical reactions. *The Journal of Chemical Physics*, 3:107–115, 1935.

[90] G Mills, GK Schenter, DE Makarov, and H Jónsson. Raw quantum transition state theory. In *Classical and Quantum Dynamics in Condensed Phase Simulations*, chapter 17, pages 405–421. World Scientific, Lerici, 1998.

[91] H Jónsson, G Mills, and KW Jacobsen. Nudged elatic band method for finding minimum energy paths of transitions. In B Berne, G Cicotti, and DF Coker, editor, *Classical and Quantum Dynamics in Condensed Phase Simulations*, chapter 16, pages 385–404. WORLD SCIENTIFIC PUB CO INC, Lerici, 1998.

[92] S Fischer and M Karplus. Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chemical Physics Letters*, 194:252–261, 1992.

[93] IV Ionova, and EA Carter. Ridge method for finding saddle points on potential energy surfaces. *The Journal of Chemical Physics*, 98:6377–6388, 1993.

[94] MJS Dewar, EF Healy, and JJP Stewart. Location of transition states in reaction mechanisms. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics*, 80:227–233, 1984.

[95] G Henkelman, and H Jonsson. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *The Journal of Chemical Physics*, 111:7010–7022, 1999.

[96] R Yu and C De Jonghe. Proton-transfer mechanism in $lapo_4$. *The Journal of Physical Chemistry C*, 111:11003–11007, 2007.

[97] X Ke and I Tanaka. Atomistic mechanism of proton conduction in solid $cshso_4$ by a first-principles study. *Physical Review B*, 69, 2004.

[98] A-N Bondar, M Elstner, S Suhai, JC Smith, and S Fischer. Mechanism of primary proton transfer in bacteriorhodopsin. *Structure*, 12:1281–1288, 2004.

[99] R Friedman, S Fischer, E Nachliel, S Scheiner, and M Gutman. Minimum energy pathways for proton transfer between adjacent sites exposed to water. *The Journal of Physical Chemistry B*, 111:6059–6070, 2007.

[100] G Henkelmann, G Johannesson, and H Jonsson. Methods for finding saddle points and minimum energy paths. In Steven D Schwartz, editor, *Theoretical Methods in Condensed Phase Chemistry*, chapter 10, pages 269–300. Springer Netherlands, Dordrecht, 2002.

[101] F Noé, F Ille, JC Smith, and S Fischer. Automated computation of low-energy pathways for complex rearrangements in proteins: Application to the conformational switch of ras p21. *Proteins*, 59:534–544, 2005.

[102] F Noé. *Transition Networks. Computational Methods for the Comprehensive Analysis of of Complex Rearrangements in Proteins.* PhD thesis, Ruprechts-Karl Universität Heidelberg, 2006.

[103] F Noé and S Fischer. Transition networks for modeling the kinetics of conformational change in macromolecules. *Current Opinion in Structural Biology*, 18:154–162, 2008.

[104] RS Berry and R Breitengraser-Kunz. Topography and dynamics of multi dimensional interatomic potential surfaces. *Physical Review Letters*, 74:3951–3954, 1995.

[105] PN Mortenson and DJ Wales. Energy landscapes, global optimizationand dynamics of the polyalanine $ac(ala)_8nhme$. *Journal of Chemicl Physics*, 114:6443–6453, 2001.

[106] DA Evand and DJ Wales. The free-energy landscape and dynamics of met-enkephalin. *Journal of Chemicl Physics*, 119:9947–9955, 2003.

[107] DJ Wales. Discrete path sampling. *Molecular Physics*, 100:3285–3305, 2002.

[108] DJ Wales. Energy landscapes: Calculating pathways and rates. *International Review in Physical Chemistry*, 25:237–282, 2006.

[109] O Ivchenko, CS Whittleston, JM Carr, P Imhof, S Goerke, P Bachert, and DJ Wales. Proton transfer pathways, energy landscape, and kinetics in creatine-water systems. *Journal of Physical Chemistry B*, 118:1969–1975, 2014.

[110] S Somani and DJ Wales. Energy landscapes and global thermodynamics for alanine peptides. *Journal of Chemical Physics*, 139, 2013.

[111] D Rapport, CJ Galvin, DY Zubarev, A Aspuru-Guzik. Complex chemical reaction networks from heuristic-aided quantum chemistry. *Journal of Chemical Theory and Computation*, 10:897–907, 2014.

[112] FA Stillinger. Exponential multiplicity of inherent structures. *Physical Review E*, 59:48–51, 1999.

[113] MRA Blombrg and PEM Siegbahn. Different types of biological proton transfer reactions studied by quantum chemical methods. *Biochimica et Biophysica Acta - Bioenergetics*, 1757:969–980, 2006.

[114] GR Desiraju. Chemistry beyond the molecule. *Nature*, 412:397–400, 2001.

[115] P Ball. In *Life's Matrix: A Biography of Water*. 1999.

[116] F Martin and H Zipse. Charge distribution in the water molecule - a comparison of methods. *Journal of Computational Chemistry*, 26:97–105, 2005.

[117] D Marx. Proton transfer 200 years after von grotthus: Insights from ab inition simulations. *ChemPhysChem*, 7:1848–1870, 2006.

[118] N Bjerrum. Structure and properties of ice. *Science*, 115:385–390, 1952.

[119] GG Malenkov. Liquid water and ices: understanding the structure and physical propertie. *Journal of Physics Condensed Matter*, 21, 2009.

[120] L Pauling. In *The Nature of the Chemical Bond and the Structure of Molecules and Crystals. An Introduction to Modern Structural Chemistry*. Oxford University Press, 1940.

[121] GG Malenkov, D Tytik, and EA Zheligovskaya. Hydrogen bonds in computer-simulated water. *Journal of Molecuar Liquids*, 82:27–38, 1999.

[122] AG Kalinichev and JD Bass. Hydrogen bonding in supercritical water: a monte carlo simulation. *Chemical Physics Letters*, 231:301–307, 1994.

[123] R Kumar, JR Schmidt, and JL Skinner. Hydrogen bonding definitions and dynamics in liquid water. *The Journal Chemical Physics*, 126, 2007.

[124] GG Malenkov. Geometry of structures consisting of water molecules in hydrated crystals. *Journal of Structural Chemistry*, 3:206–226, 1962.

[125] M Chaplin. Anomalous properties of water.

[126] Y Marechal. Transfer of protons as a third fundamental property of h-bonds. In *Proton Transfer in Hydrogen-Bonded Systems*, chapter 1, pages 1–17. 1991.

[127] A Polian and M Grimsditch. New high-pressure phase of h_2o: Ice x. *Physical Review Letters*, 52, 1984.

[128] EF Caldin. Tunneling in proton-transfer reactions in solution. *Chemical Reviews*, 69:135–156, 1969.

[129] JN Collie and T Tickle. Lxvi. - the salts of dimethylpyrone, and the quadrivalence of oxygen. *Journal of the Chemical Society*, 75:710–717, 1899.

[130] G Zundel and H Metzger. Energiebänder der tunnelnden Überschuss-protonen in flüssigen säuren. eine ir-spektroskopische untersuchung der natur der gruppierungen $h_5o_2^+$. *Zeitschrift für physikalische Chemie*, 58:225–245, 1968.

[131] E Wicke, M Eigen, and T Ackermann. Über den zustand des protons (hydroniumions) in wässriger lösung. *Zeitschrift für physikalische Chemie*, 1:340–364, 1954.

[132] M Eigen. Proton transfer, acid-base catalysis, and enzymatic hydrolysis. part i: Elementary processes. *Angewandte Chemie International Edition*, 3:1–19, 1964.

[133] ML Higgins. The role of hydrogen bonds in conduction by hydrogen and hydroxyl ions. *Journal of the American Chemical Society*, 53:3190–3191, 1931.

[134] ML Higgins. Hydrogen bridges in ice and liquid water. *Journal of Physical Chemistry*, 40:723–731, 1936.

[135] ME Tuckerman, K Laasonen, M Sprik, and M Parrinello. Ab initio molecular dynamics simulation of the solvation and transport of h3o$^+$ and oh$^-$ ions in water. *Journl of Physical Chemistry*, 99:5749–5752, 1995.

[136] ME Tuckerman, K Laasonen, and M Sprik. Ab initio molecular dynamics simulation of the solvation and transport of hydronium and hydroxyl ions in water. *Journl of Physical Chemistry*, 103, 1995.

[137] CJT de Grotthus. Sur la décomposition de leau et des corps quelle tient en dissolution À laide de lélectricité galvanique. *Annali di Chimica*, 58:54–73, 1806.

[138] H Engelhardt. Protonic conduction in ice. In *Physics and Chemsitry of Ice*, pages 226–235. University of Toronto Press, 1973.

[139] N Agmon. Mechanism of hydroxide mobility. *Chemical Physics Letters*, 319:247–252, 2000.

[140] R Ludwig. New insight into the transport mechanism of hydrated hydroxide ions in water. *Angwandte Chemie International Edition*, 42:258–260, 2003.

[141] ME Tuckerman, D Marx, and M Parrinello. The nature and transport mechanism of hydrated hydroxide ions in aqueous solution. *Nature*, 417:925–929, 2002.

[142] A Botti, F Bruni, S Imberti, MA Ricci, and AK Soper. Solvation of hydroxyl ions in water. *Journal of Chemical Physics*, 119, 2003.

[143] JD Watson and FHC Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[144] W Kabsch and C Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22:2577–2637, 2003.

[145] AJ Dingley, F Cordier, and S Grzesiek. An introduction to hydrogen bond scalar couplings. *Concepts in Magnetic Resonance*, 13:103–127, 2001.

[146] LAH van Bergen, M Alonso, A Palló, L Nilsson, F De Proft, and J Messens. Revisiting sulfur h-bonds in proteins: The example of peroxiredoxin ahpe. *Scientfic Reports*, 6:30369, 2016.

[147] SK Panigrahi and GR Desiraju. Strong and weak hydrogen bonds in drug-dna complexes: a statistical analysis. *Journal of Biosciences*, 32:677–691, 2007.

[148] CA Wraight. Chance and design - proton transfer in water, channels and bioenergetic proteins. *Biochimica et Biophysica Acta - Bioenergetics*, 1757:886–912, 2006.

[149] L Onsager. Ion passages in lipid bilayers. *Science*, 156, 1967.

[150] JF Nagle and HJ Morowitz. Molecular mechanisms for proton transport in membranes. *PNAS*, 75:298–302, 1978.

[151] JF Nagle and S Tristram-Nagle. Hydrogen bonded chain mechanism for proton conduction and proton pumping. *Journal of Membrane Biolgy*, 74:1–14, 1983.

[152] TE Decoursey. Voltage-gated proton channels and other proton transfer pathways. *Physical Review*, 83:475–579, 2003.

[153] L Onsager. Possible mechanism of ion transit. In *Physical Principles of Biological Membranes*, pages 137–139. Gordon and Breach, 1970.

[154] O Haapanen and V Sharma. Role of water and protein dynamics in proton pumping by complex i. *Scientific Reports*, 7, 2017.

[155] Y Peng, JMJ Swanson, SG Kang, R Zhou, and GA Voth. Hydrated excess protons can create their own water wires. *Journal of Physical Chemistry B*, 119:9212–9218, 2015.

[156] RM Henry, C-H Yu, T Rodinger, and R Pommès. Functional hydration and conformational gating of proton uptake in cytochrome c oxidase. *Journal of Molecular Biology*, 387:1165–1185, 2009.

[157] MM Pereira, M Santana, and MA Teixeira. A novel-scenario for the evolution of haem-copper oxygen reductases. *Biochimica et Biophysica Acta - Bioenergetics*, 1505:185–208, 2001.

[158] M Wikstrom, K Krab, and V Sharma. Oxygen activation and energy conservation by cytochrome c oxidase. *Chemical Review*, 118:2469–2490, 2018.

[159] J Simon, RMJ van Spanning, and DJ Richardson. The organisation of proton motive and non-protive motive redox loops in prokaryotic respiratory systems. *Biochimica et Biophysica Acta*, 1777:1480–1490, 2008.

[160] T Tsukihara, H Aoyama, E Yamashita, T Tomizaki, H Yamaguchi, K Shinzawa-Ithoh, R Nakashima, R Yaono, and S Yoshikawa. The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 a. *Science*, 272:1136–1144, 1996.

[161] M Svensson-Ek, J Abramson, G Larsson, S Törnroth, P Brzezinski, and S Iwata. The x-rayy crystal structures of wild-type and eq(i-286) mutant ctochrome c oxidase from rhodobacter sphaeroides. *Journal of Molecular Biology*, 321:329–339, 2002.

[162] DA Mills, L Geren, C Hiser, B Schmidt, B Durham, F Millett, and S Ferguson-Miller. An arginine to lysine mutation in the vicinit of the heme propionates affects the redox potentials of the hemes and associated electron and proton transfer pathways in cyytochrome c oxidase. *Biochemistry*, 44:10457–10465, 2005.

[163] C von Ballmoos, N Gronska, P Lachmann, RB Gennis, P Ädelroth, and P Brzezinski. Mutation of a single residue in the ba3 oxidase specifically impairs protonation of the pump site. *PNAS*, 112:3397–3402, 2015.

[164] M Wikstrom and V Sharma. Proton pumping by cytochrome c oxidase - a 40 year anniversary. *Biochimica et Biophysica Acta - Bioenergetics*, 1859:692–698, 2018.

[165] JG Lindsay and DF Wilson. Reaction of cytochrome c oxidase with co: Involvementof the invisible copper. *FEBS Letters*, 48:45–49, 1974.

[166] T Kitagawa and T Ogura. *Oxygen Activation Mechanism at the binuclear site of heme-copper oxidase superfamily as revealed by time-resolved resonance raman spectroscopy.* John Wiley and Sons, 1996.

[167] MI Verkhovsky, JE Morgan, and M Wikström. Oxygen binding and activation: early steps in the reaction of oxygen with cytochrome c oxidase. *Biochemistry*, 33:3079–3086, 1994.

[168] M Wikström. Energy-dependent reversal of the cytochrome c oxidase reaction. *PNAS*, 78:4051–4054, 1981.

[169] EA Gorbikova, I Belevich, M Wikström, and MI Verkhovsky. The proton donor for oo bond scission by cytochrome c oxidase. *PNAS*, 105:10733–10737, 2008.

[170] P Rich. Towards an understading of the chemistry oxygen reduction. *Australian Journal of Plant Physiology*, 22:479–486, 1995.

[171] I Belevich, MI Verkhovsky, and M Wikström. Proton-coupled electron transfer drives the proton pump of cytochrome c oxidase. *Nature*, 440:829–832, 2006.

[172] C Tommos and GT Babcock. Proton and hydrogen currents in photosynthetic water oxidation. *Biochimica et Biophysica Acta - Bioenergetics*, 1458:199–219, 2000.

[173] S Han, S Takahashi, and DL Rousseau. Time dependence of the catalytic intermediate in cytochrome c oxidase. *Journal of Biological Chemistry*, 275:1910–1919, 2000.

[174] V Sharma, KD Karlin, and M Wikström. Computational study of the activated o_h state in the catalytic mechanism of cytochrome c oxidase. *PNAS*, 110:16844–16849, 2013.

[175] VR Kailla, MP Johansson, D Sundholm, L Laakkonen, and M Wikström. The chemistry of the cu_b site in cytochrome c oxidase and the importance of its unique his-tyr bond. *Biochimica et Biophysica Acta - Bioenergetics*, 1787:221–233, 2009.

[176] MR Blomberg. Mechanism of oxygen reduction in cytochrome c oxidase and the role of th active site tyrosine. *Biochimica et Biophysica Acta - Bioenergetics*, 1787:221–233, 2009.

[177] RM Cordeiro. Reactive oxygen species at phospholipid bilayers: Distribution, mobility and permeation. *Biochimica et Biophysica Acta - Biomembranes*, 1838:438–444, 2014.

[178] MN Möller, Q Li, M Chinnaraj, HC Cheung, JR Lancaster Jr., and A Denicola. Solubility and diffuion of oxygen in phospholipid membranes. *Biochimica et Biophysica Acta - Biomembranes*, 1858:2923–2030, 2016.

[179] S Riistama, A Puustinen, A Garcia-Horsman, S Iwata, H Michel, and M Wikström. Channeling of dioxygen into the respiratory enzyme. *Biochimica et Biophysica Acta - Bioenergetics*, 1275:1–4, 1996.

[180] VM Luna, Y Chen, JA Fee, and CD Stout. Crystallographc studies of xe and kr binding with the large internal cavity of cytochrome ba_3 from thermus thermophilus: Structural analysis and role of oxygen transport channels in the heme-cu oxidases. *Biochemistry*, 47:4657–4665, 2008.

[181] VM Luna, JA Fee, Deniz AA, and CD Stout. Mobility of xe atoms within the oxygen diffusion channel of cytochrome ba_3 oxidase. *Biochemistry*, 51:4669–4676, 2012.

[182] B Schmidt, J McCracken, SA Ferguson-Miller. Discrete water exit pathway in the membrane protein cytochrome c oxidase. *PNAS*, 100:15539–15542, 2003.

[183] M Wikström, C Ribacka, M Molin, L Laakkonen, and M Verkhovky. A gating of proton and water transfer in the respiratory enzyme cytochrome c oxidase. *PNAS*, 102:10478–10481, 2005.

[184] S Shimada et al. Complex structure of cytochrome c - cytochrome c oxidase reveals a novel protein-protein interaction mode. *EMBO Journal*, 36:291–300, 2017.

[185] DA Mills, L Florens, C Hiser, J Qian, and S Ferguson-Miller. Where is 'outside' in cytochrome c oxidase and how and when do protons get there? *Biochimica et Biophysica Acta*, 1458:180–187, 2000.

[186] L Qin, C Hiser, A Mulichak, RM Garavita and S Ferguson-Miller. Identification of conserved lipid/detergent-binding in a high resolution structure of the membrane protein cytochrome c oxidase. *PNAS*, 103:16117–16122, 2006.

[187] E Olkhova, MC Hutter, MA Lill, V Helms, and H Michel. Dynamic water networks in ctochrome c oxidase from paracoccus denitrificans investigated by molecular dynamics simulations. *Biophysical Journal*, 86:1873–1889, 2004.

[188] T Tsukihara, K Shimokata, Y Katayama, H Shimada, K Muramoto, and H Aoyama. The low-spin heme of cytochrome c oxidase as the driving element of the proton-pumping process. *PNAS*, 100:15304–15309, 2003.

[189] K Shinzawa-Itoh et al. Structures and physiological roles of 13 integral lipids on bovine heart cytochrome c oxidase. *EMBO Journal*, 26:1713–1725, 2007.

[190] JA Garcia-Horsman, A Puustinen, RB Gennis, and M Wikström. Proton transfer in cytochrome bo3 ubiquinol oxidase of eschrichia coli: second-site

mutations in subunit i that restore proton pumping in the mutant asp135. *Biochemistry*, 34:4428–4433, 1995.

[191] G Brändén, AS Pawate, RB Gennis, and P Brzezinski. Controlled uncoupling and recoupling of proton pumping cytochrome c oxidase. *PNAS*, 103:317–322, 2006.

[192] AS Pawate, J Morgan, A Namslauer, D Mills, P Brzezinski, S Ferguson-Miller, and RB Gennis. A mutation of subunit i in cytochrome c oxidase from rhodobacter sphaeroides results in an increase in steady state activity but completely eliminates proton pumping. *Biochemistry*, 41:13417–13423, 2002.

[193] U Pfitzner, K Hoffmeier, A Harrenga, A Kannt, H Michel, and E Bamberg. Tracing the d-pathway in reconstituted site-directed mutants of cytochrome c oxidase from paracoccus dentrificans. *Biochemistry*, 39:6756–6762, 2000.

[194] H Lepp, L Salomonsson, JP Zhu, RB Gennis, and P Brzezinski. Impaired proton pumping in cytochrome c oxidase upon structural alteration of the d pathway. *Biochimica et Biophysica Acta*, 1777:897–903, 2008.

[195] R Liang, JMJ Swansson, M Wikström, and GA Voth. Understaning the essential proton-pumping kinetic gates and decoupling mutations in cytochrome c oxidase. *PNAS*, 114:5924–5929, 2017.

[196] D Han, A Namslauer, AS Pawate, JE Morgan, S Nagy, AS Vakkasoglu, P Brzezinski, and RB Gennis. Replacing asn207 by aspartate at the neck of the d-channel in the aa_3 cytochrome c oxidase from rhodobacter sphaeroides results in decoupling of the proton pump. *Biochemistry*, 45:14064–14074, 2006.

[197] I Belevich, E Gorbikova, NP Belevich, V Rauhamäki, M Wikström, and MI Verkhovsky. Initiation of the proton pump of cytochrome c oxidase. *PNAS*, 107:18469–18474, 2010.

[198] C Backgren, G Hummer, M Wikström, and A Puustinen. Proton translocatin by cytochrome c oxidase can take place without the conserved glutamic acid in subunit i. *Biochemistry*, 39:7863–7867, 2000.

[199] R Pòmes, G Hummer, and M Wikström. Structure and dynamics of a proton shuttle in cytochrome c oxidase. *Biochimica et Biphysica Acta*, 1365:255–260, 1998.

[200] S Yang and Q Cui. Glu-286 rotation and water wire reorientation are unlikely the gating elements for proton pumping in cytochrome c oxidase. *Biophysical Journal*, 101:61–69, 2011.

[201] A Maréchal and PR Rich. Water molecule reorganization in cytochrome c oxidase revealed by ftir spectroscopy. *PNAS*, 108:8634–8638, 2011.

[202] V Sharma, G Enkavi, I Vattulainen. Proton-coupled electron transfer and the role of water molecules in proton pumping by cytochrome c oxidase. *PNAS*, 112:2040–2045, 2015.

[203] J Xu and GA Voth. Computer simulation of explicit proton translocation in cytochrome c oxidase: the d-pathway. *PNAS*, 102:6795–6800, 2005.

[204] J Xu, MA Sharpe, L Qin, S Ferguson-Miller, and GA Voth. Storage of an excess proton in the hydrogen-bonded network of the d-pathway of cytochrome c oxidase: identification of a protonated water cluster. *Journal of th American Chemical Society*, 129:2910–2913, 2007.

[205] J Xu and GA Voth. Free energy profiles for h^+ conduction in the d-pathway of cytochrome c oxidase: a study of the wild type and n98d mutant enzyme. *Biochimica et Biophysica Acta*, 1757:852–859, 2006.

[206] R Liang, JMJ Swansson, Y Peng, M Wikström, and GA Voth. Multiscale simulations reveal key-features of the proton-pumping mechanism in cytochrome c oxidase. *PNAS*, 113:7420–7425, 2016.

[207] FL Tomson, JE Morgan, G Gu, B Barquera, T Vygodina, and RB Gennis. Substitutions of glutamate 101 in subunit ii of cytochrome c oxidase from rhodobacter sphaeroides result in blocking the proton-conducting k-channel. *Biochemistry*, 42:1711–1717, 2003.

[208] OM Richter, KL DÃŒrr, A Kannt, B Ludwig, FM Scandurra, A Giuffre, P Sarti, and P Hellwig. Probing the access of protons to th k pathway in the paracoccus denitrificans cytochrome c oxidase. *FEBS Journal*, 272:404–412, 2005.

[209] J Koepke, E Olkhova, H Angerer, H Müller, G Peng, and H Michel. High resolutin crystal structure of paracoccus denitrificans cytochrome c oxidase: New insights into the active site and the proton transfer pathways. *Biochimica et Biophysica Acta*, 1787:635–645, 2009.

[210] S Supekar and VRI Kailla. Dewetting transitions coupled to k-channel activation in cytochrome c oxidase. *Chemical Science*, 9:6703–6710, 2018.

[211] AL Woelke, G Galstyan, E-W Knapp. Lysine 362 in cytochrome c oxidase regulates opening of the k-channel via changes in pk_a and conformation. *Biochimica et Biophysica Acta - Bioenergetics*, 1837:1998–2003, 2014.

[212] V Sharma and M Wikström. The role of the k-channel and the active-site tyrosine in the catalytic mechanism of cytochrome c oxidase. *Biochimica et Biophysica Acta - Bioenergetics*, 1857:1111–1115, 2016.

[213] S Jünemann, B Meunier, RB Gennis, and PR Rich. Effects of the mutation of the conserved lysine-362 in cytochrome c oxidase from rhodobacter sphaeroides. *Biochemistry*, 36:14456–14464, 1997.

[214] V Sharma, PG Jambrina, M Kaukonen, E Rosta, and PR Rich. Insights into functions of the h channel of cytochrome c oxidase from atomistic molecular dynamics simulations. *PNAS*, 114, 2017.

[215] S Yoshikawa and A Shimada. Reaction mechanism of cytochrome c oxidase. *Chemical Reviews*, 115:1936–1989, 2015.

[216] S Yoshikawa, K Muramoto, and K Shizawa-Itoh. Proton-pumping mechanism of cytochrome c oxidase. *Annual Review of Biophysics*, 40:205–223, 2011.

[217] AA Shimada. Nanosecond time-resolved xfel analysis of structural changes associated with co release from cytochrome c oxidase. *Science Advances*, 3, 2017.

[218] HM Lee, TK Das, DL Rousseau, D Mills, S Ferguson-Miller, and RB Gennis. Mutations in the putative h-channelin the cytochrome c oxidase of rhodobacter sphaeroides shw that this channel is not important for proton conduction but reveal modulation of the properties of heme a. *Biochemistry*, 39:2989–2996, 2000.

[219] J Salje, B Ludwig, and OMH Richter. Is a third proton-conducting pathway operative in bacterial cytochrome c oxidase? *Biochemical Society Transactions*, 33:829–831, 2005.

[220] A Maréchal, B Meunier, D Lee, C Orengo, and PR Rich. Yeast cytochrome c oxidase: A model system to study mitochondrial forms of the haem-copper oxidase superfamily. *Biochimica et Biophysica Acta*, 1817:620–628, 2012.

[221] RJ Dodia. *Structure-function relationship of mitochondrial cytochrome c oxidase: Redox centres, proton pathways and isozymes.* PhD thesis, University College London.

[222] A Maréchal, F Haraux, B Meunier, and PR Rich. Determination of h^+/e^- ratios in mitochondrial yeast cytochrome c oxidase. *Biochimica et Biophysica Acta - Bioenergetics*, 1837, 2014.

[223] PR Rich and A Maréchal. Functions of the hydrophilic channels in protonmotive cytochrome c oxidase. *Journal of The Royal Society Interface*, 10, 2013.

[224] M Wikström and K Krab. Proton-pumping cytochrome c oxidase. *Biochimica et Biophysica Acta - Bioenergetics*, 549:177–222, 1979.

[225] S Yoshikawa, K Muramoto, K Shinzawa-Itoh, and M Mochizuki. Structural studies on bovine heart cytochrome c oxidase. *Biochimica et Biophysica Acta - Bioenergetics*, 1817:579–589, 2012.

[226] N Yano, K Muramoto, A Shimada, S Takemura, J Baba, H Fujisawa, M Mochizuki, K Shinzawa-Itoh, E Yamashita, T Tsukihara, and S Yoshikawa. The $mg2^+$-containing water cluster of mammalian cytochrome c oxidase collects four pumping proton equivalents in each catalytic cycle. *Journal of Biological Chemistry*, 291:23882–23894, 2016.

[227] I Belevich, DA Bloch, N Belevich, M Wikström, and MI Verkhovsky. Exploring the proton pump mechanism of cytochrome c oxidase in real time. *PNAS*, 104:2685–2690, 2007.

[228] SA Siletsky and AA Konstatinov. Cytochrome c oxidase: Charge translocation coupled to single-electron partial steps of the catalytic cycle. *Biochimica et Biophysica Acta - Bienergetics*, 1817:476–488, 2012.

[229] SA Siletsky, AS Pawate, K Weiss, RB Gennis, and AA Konstatinov. Transmembrane charge separation during the ferryl-oxo to oxidized transition in a nonpumping mutant of cytochrome c oxidase. *Journal of Biological Chemistry*, 279:52558–52565, 2004.

[230] M Wikström and MI Verkhovsky. Towards the mechanism of proton pumping by haem-copper oxidases. *Biochimica et Biophysica Acta - Bioenergetics*, 1757:1047–1051, 2006.

[231] JE Morgan, JE Verkhovsky, and MI Wikström. The histidine cycle: A new model for proton translocation in the respiratory heme-copper oxidases. *Journal of Bioenergetics and Biomembranes*, 26:599–608, 1994.

[232] J Quenneville, DM Popović, and AA Stuchebrukhov. Redox-dependent pk_a of of cu_b histidine ligand in cytochrome c oxidase. *Journal of Physical Chemistry*, 108:18383–18389, 2004.

[233] M Wikström and MI Verkhovsky. Mechanism and energetics of proton translocation by the respiratory heme-copper oxidases. *Biochimica et Biophysica Acta - Bienergetics*, 1767:1200–1214, 2007.

[234] VRI Kailla, V Sharma, and M Wikström. The identity of the transient proton loading site of the proton-pumping mechanism of cytochrome c oxidase. *Biochimica et Biophysica Acta - Bienergetics*, 1807:80–84, 2011.

[235] PE Siegbahn, MR Blomberg, and ML Blomberg. Theoretical study of the energetics of proton pumping and oxygen reduction in cytochrome c oxidase. *Journal of Physical Chemistry*, 107:10946–10955, 2003.

[236] S Supekar, AP Gamiz-Hernandez, and VRI Kailla. A protonated water-cluster as a transient proton-loading site in cytochrome c oxidase. *Angewandte Chemie International Edition*, 55:11940–11944, 2016.

[237] R Mitchell and PR Rich. Proton uptake by cytochrome c oxidase on reduction and on ligand binding. *Biochimica et Biophysica Acta - Bioenergetics*, 1186:19–26, 1994.

[238] D Bloch, I Belevich, A Jasaitis, C Ribacka, A Puustinen, MI Verkhovsky, M Wikström. The catalytic cycle of cytochrome c oxidase is not the sum of its two halves. *PNAS*, 101:529–533, 2004.

[239] K Fax2́n, G Gilderson, P Ädelroth, and PA Brzezinski. Mechanistic principle for proton pumping by cytochrome c oxidase. *Nature*, 437:286–289, 2005.

[240] M Wikström, MI Verkhovsky, and G Hummer. Water-gated mechanism of proton translocation by cytochrome c oxidase. *Biochimica et Biophysica Acta*, 1604:61–65, 2003.

[241] PE Siegbahn and MR Blomberg. Energy diagrams and mechanism for proton pumping in cytochrome c oxidase. *Biochimica et Biophysica Acta - Bioenergetics*, 1767:1143–1156, 2007.

[242] AL Woelke, G Galstyan, A Galstyan, T Meyer, J Heberle, and E-W Knapp. Exploring the possible role of glu286 in cco by electrostatic energy computations combined with molecular dynamics. *Journal of Physical Chemistry B*, 117:12432–12441, 2013.

[243] VRI Kailla, MI Verkhovsky, G Hummer, and M Wikström. Glutamic acid 242 is a valve in the proton pump of cytochrome c oxidase. *PNAS*, 105:6255–6259, 2008.

[244] W Weber. *Ein neues semiempirisches NDDO-Verfahren mit Orthogonalisierungskorrekturen: Entwicklung des Modells, Implementierung, Parametrisierung und Anwendungen*. PhD thesis, Universität Zürich, 1996.

[245] W Weber and W Thiel. Orthogonalization corrections for semiempirical methods. *Theoretical Chemistry Accounts*, 103:495–506, 2000.

[246] M Scholten. *Semiempirische Verfahren mit Orthogonalisierungskorrekturen: Die OM3 Methode*. PhD thesis, Heinrich-Heine-Universität Düsseldorf, 2003.

[247] M Born and JR Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 20:457–484, 1927.

[248] DJ Wales. The born-oppenheimer approximation and normal modes. In *Energy Landscapes - With Applications to Clusters, Biomolecules and Glasses*, chapter 2, pages 119–157. Cambridge University Press, 2003.

[249] V Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift für Physik*, 61:126–148, 1930.

[250] JC Slater. Note on hartree's method. *Physical Review*, 35, 1930.

[251] DR Hartree. The wave mechanics of an atom with a non-coulomb central field. part i- theory and methods. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24:89–110, 1928.

[252] R Pariser and RG Parr. A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. *International Journal of Chemical Physics*, 21, 1953.

[253] JA Pople, DP Santry, and GA Segal. Approximate self-consistent molecular orbital theory. i. invariant procedures. *Journal of Chemical Physics*, 43, 1965.

[254] W Heisenbeg. Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. *Zeitschrift für Physik*, 43:172–198, 1927.

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

[256] IN Levine. The schrödinger equation. In *Quantum Chemistry, 7th Edition*, chapter 1, pages 1–19. PEARSON, 2012.

[257] R Bertlmann. Time-independent schrödinger equation, 2008.

[258] IN Levine. Electronic structure of diatomic molecules. In *Quantum Chemistry, 7th Edition*, chapter 13, pages 344–398. PEARSON, 2012.

[259] IN Levine. Many-electron atoms. In *Quantum Chemistry, 7th Edition*, chapter 11, pages 289–324. PEARSON, 2012.

[260] P Tulip. *Dielectric and lattice dynamical properties of molecular crystals via density functional perturbation theory: Implementation within a first principle code*. PhD thesis, University of Durham, 2004.

[261] W Pauli. Über den einfluss der geschwindigkeitsabhängigkeit der elektronenmasse auf den zeemaneffekt. *Zeitschrift für Physik*, 31:373–385, 1925.

[262] A Williamson. *Quantum Monte Carlo Calculations of Electronic Excitations*. PhD thesis, Cambridge University, 1996.

[263] GG Hall. The molecular orbital theory of chemical valency viii. a method of calculating ionization potentials. *Proceedings of the Royal Society A*, 205:541–552, 1951.

[264] CCJ Roothaan. New developments in molecular orbital theory. *Reviews of Modern Physics*, 23, 1951.

[265] NT Anh. Atomic and molecular orbitals. In *Frontir Orbitals: A Practical Manual*, chapter 2. Wiley, 2007.

[266] DL Strout and GE Scuseria. A quantitative study of the scaling properties of the hartree-fock method. *The Journal of Chemical Physics*, 102, 1995.

[267] F Jensen. Semi-empirical methods. In *Introduction to computational chemistry*, chapter 7. John Wiley and Sons, 1999.

[268] R Hoffmann. An extended hückel theory. i. hydrocarbons. *The Journal of Chemical Physics*, 39:1397–1412, 1963.

[269] MO Steinhauser. Näherungsmethoden in der quantenmechnaik. In *Quantenmechanik für Naturwissenschaftler: Ein Lehr- und Übungsbuch mit zahlreichen Aufgaben und Lösungen*, chapter 7. Springer, 2017.

[270] JA Pople and DL Beveridge. Approximate molecular orbital theories. In *Approxmate Molecular Orbital Theory*, chapter 3. McGraw-Hill Book Company, 1970.

[271] JA Pople and DL Beveridge. Approximate self-consistent molecular-orbital theory. v. intermediate neglect of differential overlap. *The Journal of Chemical Physics*, 47, 1967.

[272] MJS Dewar and NC Baird. Ground states of sigma-bonded molecules. iv. the mindo method and its application to hydrocarbons. *The Journal of Chemical Physics*, 50:1262–1274, 1969.

[273] MJS Dewar and W Thiel. Ground states of molecules. 38. the mndo method. approximations and parameters. *Journal of the American Chemical Society*, 99:4899–4906, 1977.

[274] MJS Dewar and W Thiel. Ground states of molecules. 39. mndo results for molecules containing hydrogen, carbon, nitrogen, and oxygen. *Journal of the American Chemical Society*, 99:4907–4917, 1977.

[275] J Ridley and M Zerner. An intermediate neglect of differential overlap technique for spectroscopy: Pyrrole and the azines. *Theoretica chimica acta*, 32:111–134, 1973.

[276] MJS Dewar. Quantum mechanical molecular models. *The Journal of Physical Chemistry*, 89:2145–2150, 1985.

[277] W Thiel. Semiempirical methods: Current status and perspectives. *Tetrahedron*, 44:7393–7408, 1988.

[278] JJP Stewart. Mopac: A semiempirical molecular orbital program. *Journal of computer-aided molecular design*, 4:1–105, 1990.

[279] JJP Stewart. Semiempirical molecular orbital methods. In *Reviews in Computational Chemistry*. Verlag Chemie, 1990.

[280] MC Zerner. Semiempirical molecular orbital methods. In *Reviews in Computational Chemistry*. Verlag Chemie, 1990.

[281] MJS Dewar, EG Zobbisch, EF Healy, JP Stewart. Am1: A new general purpose quantum mechanica molcular model. *Journal of the american chemical society*, 107:3902–3909, 1985.

[282] KR Roby. On the justifiability of neglect of differential overlap molecular orbital methods. *Chemical Physics Letters*, 11:6–10, 1971.

[283] KR Roby. Fundamentals of an orthonormal basis set molecular orbital theory. *Chemical Physics Letters*, 12:579–582, 1972.

[284] M Kolb. *Ein neues semiempirisches Verfahren auf Grundlage der NDDO-Näherung: Entwicklung der Methode, Parametrisierung und Anwandungen.* PhD thesis, Bergische Universität - Gesamthochschule Wuppertal, 1991.

[285] M Kolb and W Thiel. Beyond the mndo model: Methodical considerations and numerical results. *Journal of Computational Chemistry*, 14:775–789, 1993.

[286] EG Lewars. The concept of the potential energy surface. In *Computational Chemistry*, chapter 2, pages 9–49. Springer, Cham, 2016.

[287] J Nocedal and SJ Wright. Fundamentals of unconstrained optimization. In *Numerical Optimization*, chapter 2, pages 10–30. Springer, 1999.

[288] JE Sinclair and R Fletcher. A new method of saddle-point location for the calculation of defect migration energies. *Journal of Physics C: Solid State Physics*, 7:864–870, 1974.

[289] R Fletcher and CM Reeves. Function minimization by conjugate gradients. *The Computer Journal*, 7:149–154, 1964.

[290] MJD Powell. An iterative method for finding stationary values of a function of several variables. *The Computer Journal*, 5:147–151, 1962.

[291] A Cauchy. Methodes generales pour la resolution des syst'emes dequations simultanees. *Comptes Rendus Mathematique Academie des Sciences Paris*, 25:536–538, 1847.

[292] J Nocedal and SJ Wright. Line search methods. In *Numerical Optimization*, chapter 3, pages 34–62. Springer, 1999.

[293] J-W Chu, BL Trout and BR Brooks. A super-linear minimization scheme for the nudged elastic band method. *The Journal of Chemical Physics*, 119, 2003.

[294] J Barzilai and J Borwein. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8:141–148, 1988.

[295] G Henkelman, G Jóhannesson, and H Jónsson. Methods for finding saddle points and minimum energy pathways. In *Theoretical Methods in Condensed Phase Chemistry*, chapter 10, pages 269–300. Springer, 2002.

[296] V Lasrado, D Alhat, and Y Wang. A review of recent phase transition simulation methods: Transition path search. *34th Design Automation Conference*, 1:93–101, 2008.

[297] W Quapp and D Heidrich. Analysis of the concept of minimum energy path on the potetial energy surface of chemically reacting systems. *Theoretical Chemistry Accounts*, 66:245–260, 1984.

[298] P Pechukas. On simple saddle points of a potential surface, the conservation of nuclear symmetry along paths of steepest descent, and the symmetry of transition states. *Journal of Chemical Physics*, 64:1516–1521, 1976.

[299] JD Duniz. Chemical reaction paths. *Philosophical Transactions of the Royal Society B*, 272, 1975.

[300] J Panciř. Calculation of the least energy path on the energy hypersurface. *Collection of Czechoslovak Chemical Communications*, 40:1112–1118, 1975.

[301] D Alhat, V Lasrado, and Y Wang. A review of recent phase transition simulation methods: Saddle point search. *34th Design Automation Conference*, 1, 2008.

[302] J Simons, P Jorgensen, H Taylor, and J Ozment. Walking on potential energy surfaces. *Journal of Physical Chemistry*, 87:2745–2753, 1983.

[303] MJS Dewar, EF Healy, and JJP Stewart. Location of transition states in reaction mechanisms. *Journal of the Chemical Society Faraday Transactions 2: Molecular and Chemical Physics*, 80:227–233, 1984.

[304] A Banerjee, N Adams, J Simons, and R Shepard. Search for stationary points on surfaces. *Journal of Physical Chemistry*, 89:52–57, 1985.

[305] IV Ionova and EA Carter. Ridge method for finding saddle points on potential energy surfaces. *Journal of Chemical Physics*, 98:6377–6386, 1993.

[306] N Mousseau and GT Barkema. Traveling through potential energy landscapes of disordered matrials: the activation-relaxation technique. *Physical Review E*, 57:2419–2424, 1998.

[307] A Heyden, AT Bell, and FJ Keil. Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method. *Journal of Chemical Physics*, 123:224101–224114, 2005.

[308] RA Miron and KA Fichthorn. The step and slide method for finding saddle points on multidimensional potential surfaces. *Journal of Chemical Physics*, 115:8742–8747, 2001.

[309] D Passerone, M Ceccarelli, and M Parrinello. A concerted variational strategy for investigating rare events. *Journal of Chemical Physics*, 118:2025–2032, 2003.

[310] TA Halgren and WN Lipscomb. The synchronous-transit method for determining reaction pathways and locating molecular transition states. *Chemical Physics Letters*, 49:225–232, 1977.

[311] W Quapp, M Hirsch, O Imig, and D Heidrich. Searching for saddle points of a potential energy surface by following a reduced gradient. *Journal of Computational Chemistry*, 19:1087–1100, 1998.

[312] M Hirsch and W Quapp. Improved rgf method to find saddle points. *Journal of Computational Chemistry*, 23:887–894, 2002.

[313] JM Anglada, E Besal/'u, JM Bofill, and R Crehuet. On the quadratic reaction path evaluated in a reduced potential energy surface model and the problem to locate transition states. *Journal of Computational Chemistry*, 22:387–406, 2001.

[314] G Henkelman and H Jónsson. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *Journal of Chemical Physics*, 113:9978–9985, 2000.

[315] G Henkelman, B Uberuaga, and H Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics*, 113:9901–9904, 2000.

[316] S Trygubenko and D Wales. A doubly nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics*, 120:2082–2094, 2004.

[317] I Galvan and M Field. Improving the efficiency of the neb reaction path finding algorithm. *Journal of Computational Chemistry*, 29:139–143, 2008.

[318] E Weinan, W Ren, and E Vanden-Eijnden. A simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *Journal of Chemical Physics*, 126, 2007.

[319] W Ren. Higher order string method for finding minimum energy path. *Communication in Mathematical Science*, 1, 2003.

[320] E Weinan, W Ren and E Vanden-Eijnden. String method for the study of rare events. *Physical Review B*, 66, 2002.

[321] B Peters, A Heyden, A Bell, and A Chakraborty. A growing string method for determining trasition states: Comparison to the nudged elastic band and string methods. *Journal of Chemical Physics*, 120:7877–7886, 2004.

[322] S Burger and W Yang. Quadratic string methodfor determining the minimum-energy path based on multiobjective optimization. *Journal of Chemical Physics*, 124, 2006.

[323] L Chen, C Ying, and T Ala-Nissila. Finding transition paths and rate coefficients through accelerated langevin dynamics. *Physical Review E*, 65, 2002.

[324] B Dey and P Ayers. A hamilton-jacobi type equation for computing minimum potential energy paths. *Molecular Physics*, 104:541–558, 2006.

[325] P Imhof. A network approach to modeling enzymatic reactions. In GA Voth, editor, *Computational Approaches for studying enzyme mechanism Part B*, chapter 11, pages 249–272. Academic Press, 2016.

[326] D Dörrich. Random state selection of proton-transfer transition networks, 2016.

[327] RC Hahn. Conversion of alkyl chloride to a bromide (or vice-versa) by homogeneous nucleophile attack. *Journal of Chemical Education*, 74:836–838, 1997.

[328] Y Qin, L Wang, and D Zhong. Dynamics and mechanism of ultrafast water-protein interactions. *Proceedings of the National Academy of Science*, 113:8424–8429, 2016.

[329] R Pòmes and B Roux. Structure and dynamics of a proton wire: A theoretical study of h^+ translocation along the single-file water chain in the gramicidin a channel. *Biophysical Journal*, 71:19–39, 1996.

[330] LJ Holsinger, D Nichani, LH Pinto, and RA Lamb. Influenza a virus m2 ion channel protein: structure-function analysis. *Journal of Virology*, 68:1551–1563, 1994.

[331] VRI Kaila and G Hummer. Energetics and dynamics of proton transfer reactions along short water wires. *Physical Chemistry Chemical Physics*, 13:13207–13215, 2011.

[332] GA Olah and GK Surya Prakash. Protonated (protosolvated) onium ions (onium dications). *Chemical Intermediates*, 12:141–159, 1989.

[333] MT Henzl, CL Trevino, L Dvoráka, and JM Boschi. Evidence that deprotonation of serine-55 is responsible for the ph-dependence of the parvalbumin $eu^3+ \ ^7f_0 \rightarrow \ ^5d_0$ spectrum. *FEBS Letter*, 314:130–134, 1992.

[334] C Zhao, B Ling, L Dong, and Y Liu. Theoretical insights into the protonation states of active site cysteine and citrullination mechanism of *Porphyromona gingivalis* peptidylarginine deiminase. *Proteins Structure, Function Bioinformatics*, 85:1518–1528, 2017.

[335] HW Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2:83–97, 1955.

[336] EW Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[337] L Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical Review*, 159:98–103, 1967.

[338] D Frenkel and B Smit. *Understanding Molecular Simulation*. Academic Press, 2001.

[339] KH Rosen. Graphs. In *Discrete Matheatics and its Applications, 7th Edition*, chapter 10, pages 641–735. McGraw-Hill Company, 2012.

[340] M Newman. In *Networks: An Introduction*. Oxford University Press, 2010.

[341] D Easley and J Kleinberg. In *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010.

[342] B Hayes. Graph theory in practice: Part 1. *American Scientist*, 1:9–13, 2000.

[343] S Catanese, P De Meo, E Ferrara and G Fiumara. Analyzing the facebook friendship graph. *Computing Research Repository - CORR*, 685, 2010.

[344] AJ Schwenk. Acquaintance graph party problem. *The American Mathematical Monthly*, 79:1113–1117, 1972.

[345] DM Romero, W Galuba, S Asur, and BA Huberman. Influence and passivity in social media. In *Machine Learning and Knowledge Discovery in Databases*, pages 18–33. Springer, 2011.

[346] Jerrold W. Grossman. The evolution of the mathematical research collaboration graph.

[347] J Chan, A Holmes, and R Rabadan. Network analysis of global influenza spread. *PLOS - Computational Biology*, 2010.

[348] Q Ye, T Zhu, D Hu, B Wu, N Du, and B Wang. Cell phone mini challenge award: Social network accuracy - exploring temporal communication in mobile call graphs. *IEEE Symposium on Visual Analytics Science and Technology*, 2008.

[349] P Wang, MC González, CA Hidalgo, and A-L Barabási. Understanding the spreading patterns of mobile phone viruses. *Science*, 324:1071–1076, 2009.

[350] JM Kleinberg, R Kumar, P Raghavan, S Rajagopalan, AS Tomkins. The web as a graph: Measurements, models, and methods. In *Computing and Combinatorics*, pages 1–17. Springer, 1999.

[351] W Lu, J Janssen, E Milios, N Japkowicz, and Y Zhang. Node similarity in the citation graph. *Knowledge and Information Systems*, 11:105–129, 2007.

[352] T Steiner, R Verborgh, R Troncy, J Gabarro, and R van de Walle. Adding realtime coverage to the google knowledge graph. In *11th International Semantic Web Conference*, pages 65–68. 2012.

[353] I Ross. An information processing theory of consumer choice. *Journal of Marketing*, 43:124–126, 1979.

[354] S Mancoridis, BS Mitchell, C Rorres, Y Chen, and ER Gansner. Using automatic clustering to produce high-level system organizations of source code. *Proceedings. 6th International Workshop on Program Comprehension*, 1998.

[355] LSH de Mello and AC Sanderson. And/or graph representation of assembly plans. *IEEE Transactions on Robotics and Automation*, 6:188–199, 1990.

[356] P De, EJ Dunne, JB Ghosh, and CE Wells. The discrete time-cost tradeoff problem revisited. *European Journal of Operational Research*, 81:225–238, 1995.

[357] C Li-Ping, W Ru, S Hang, X Xin-Ping, Z Jin-Song, L Wei, and C Xu. Structural properties of us flight network. *Chinese Physics Letters*, 20, 2003.

[358] U Brandes and D Wagner. Using graph layout to visualize train interconnection data. *Graph Algorithms and Applications*, 2:169–189, 2004.

[359] AI Mees. Railway scheduling by network optimization. *Mathematical and Computer Modeling*, 15:33–42, 1991.

[360] K Lee and HY Ryu. Automatic circuity and accessibility extraction by road graph network and its application with high-resolution satelite imagery. *IEEE International Geoscience and Remote Sensing Symposium*, 2004.

[361] LR Foulds and JW Gifffn. A graph-theoretic heuristic for minimizing total transport cost in facilities layout. *International Journal of Production Research*, 2007.

[362] J Vlasblom and SJ Wodak. Markov clustering versus affinity propagation for the partitioning of protein interaction graphs. *BMC Bioinformatics*, 2009.

[363] S-H Yook, ZN Oltavi, and A-L Barabási. Functional and topological characterization of protein interaction networks. *Proteomics*, 2004.

[364] JE Cohen and ZJ Palka. A stochastic theory of community food webs. v. intervality and triangulation in the trophic-niche overlap graph. *The American Naturalist*, 1990.

[365] Z Sjoerds, SM Stufflebeam, DJ Veltman, W van den Brink, BWJH Penninx, and L Douw. Loss of brain graph network efficiency in alcohol dependence. *Addiction Biology*, 2015.

[366] D Conte, P Foggia, C Sansone, and M Vento. Graph matching applications in pattern recognition and image processing. *International Conference of Image Processing*, 2003.

[367] S Barua. *A Generic Framework for the Application of Graph Theory to Image Processing*. PhD thesis, University of North Carolina at Charlotte, 2007.

[368] R Singh, J Xu, and B Berger. Global alignment of multiple protein interaction networks with application to functional orthology detection. *PNAS*, 105:12763–12768, 2008.

[369] C Clark and J Kalita. A comparison of algorithms for the pairwise alignment of biological networks. *Bioinformatics*, 30:2351–2359, 2014.

[370] A Elmsallati, C Clark, and J Kalita. Global alignment of protein-protein interaction networks: A survey. *IEEE/ACM Transactions Computanional Biology and Bioinformatics*, 2016.

[371] EM Clarke, R Enclers, T Filkorn, and S Jha. Exploiting symmetry in temporal logic model checking. *Formal Methods in System Design*, 1996.

[372] LP Cordella, P Foggia, C Sansone, and M Vento. Performance evaluation of the vf graph matching algorithm. *Conference Image Analysis and Processing*, 1999.

[373] RP Singh. Application of graph theory in computer science and engineering. *International Journal of Computer Applications*, 104:10–13, 2014.

[374] AT Balaban. Applications of graph theory in chemistry. *Journal of Chemical Information and Modeling*, 25:334–343, 1985.

[375] R Somkunwar and VM Vaze. A comparative study of graph isomorphism applications. *International Journal of Computer Applications*, 162:34–37, 2017.

[376] JA Bondy and USR Murty. Connectivity. In *Graph Theory with Applications*, chapter 42-51, pages 9–49. The Macmillan Press, 1982.

[377] SR Gallagher. *Graph Connectivity: Approximation Algorithms and Applications to Protein-Protein Interaction Networks*. PhD thesis, University of Colorado, 2010.

[378] BCM van Wijk, CJ Stam, and A Daffertshofer. Comparing brain networks of different size and connectivity density using graph theory. *PLOS One*, 2010.

[379] AK Baruah. Traffic control problems using graph connectivity. *International Journal of Computer Applications*, 86:1–3, 2014.

[380] A Ben-Dor, R Shamir, Z Yakhini. Clustering gene expression patterns. *Journal of Computational Biology*, 6:281–297, 1999.

[381] G Sugihara. Graph theory, homology and food webs. In *Population Biology*. W.H. Freeman, 1984.

[382] R Samudrala and J Moult. A graph-theoretic algorithm for comparative modeling of protein structure. *Journal of Molecular Biology*, 279:287–302, 1998.

[383] FS Kuhl, GM Crippen, and DK Friesen. A combinatorial algorithm for calculating ligand binding. *Journal of Computational Biology*, 5:24–34, 1983.

[384] MR Garay and DS Johnson. In *Computers and Intractibility: A Guide to the Theory of NP-completeness*. W.H. Freeman, 1990.

[385] L Euler. Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 8:128–140, 1741.

[386] WR Hamilton. Memorandum respecting a new system of roots of unity. *Philosophical Magazine*, 12:446, 1856.

[387] WR Hamilton. Account of the icosian calculus. *Proceedings of the Royal Irish Academy*, 6:415–416, 1858.

[388] RW Floyd. Algorithm 97. *Communications of the ACM*, 5:345, 1962.

[389] S Warshall. A theorem on boolean matrices. *Journal of the ACM*, 9:11–12, 1962.

[390] RJ Gould. Cycles and circuits. In *Graph Theory*, chapter 5. Dover Publications Inc., 2012.

[391] PA Pevzner, H Tang, and MS Waterman. An eulerian path approach to dna fragment assembly. *PNAS*, 98:9748–9753, 2001.

[392] J Kaptcianos. A graph-theoretical approach to dna fragment assembly. *American Journal of undergraduate Research*, 7, 2008.

[393] C Hierholzer. Über die möglichkeit einen linienzug ohne wiederholung und ohne unterbrechung zu umfahren. *Mathematische Annalen*, 6:30–32, 1873.

[394] O Veblen. An application of modular equations in analysis situs. *Annals of Mathematics*, 14:86–94, 1912.

[395] G Laporte. The traveling salesman problem: An overview of exact and approximate algorithms. *European Journal of Operational Research*, 59:231–247, 1992.

[396] G Ertl. Shortest path calculation in large road networks. *Operations-Research-Spektrum*, 20:15–20, 1998.

[397] RZ Farahani, E Miandoabchi, WY Szeto, and H Rashidi. A review of urban transportation network design problems. *European Journal of Operational Research*, 229:281–302, 2013.

[398] O Berman and GY Handler. Optimal minimax path of a single service unit on a network to nonservice destinations. *Transportation Science*, 21:63–129, 1987.

[399] A Shimbel. Structural parameters of communication networks. *Bulletin of Mathematical Biophysic*, 15:501–507, 1953.

[400] R Bellman. On a routing problem. *Quaterly of applied Mathematics*, 16:87–90, 1958.

[401] LR Ford Jr. Network flow theory. 1953.

[402] EF Moore. The shortest path through a maze. *Procedings of an International Symposium on the Theory of Switching*, pages 285–292, 1959.

[403] HN Gabow. Scaling algorithms for network problems. *Journal of Computer and System Sciences*, 31:148–168, 1985.

[404] HN Gabow and RE Tarjan. Faster scaling algorithms for network problems. *SIAM Journal on Computing*, pages 1013–1036, 1989.

[405] AV Goldberg. An efficient implementation of a scaling minimum-cost flow algorithm. 1992.

[406] P Sankowski. Shortest paths in matrix multiplication time. *European Symposium on Algorithms*, pages 770–778, 2005.

[407] R Yuster and U Zwick. Answering distance queries in directed graphs using fast matrix multiplication. *Symposium on Foundations of Computer Science*, pages 389–396, 2005.

[408] K Mehlhorn and P Sanders. Shortest paths. In *Algorithms and Data Structures: The Basic Toolbox*, chapter 10. Springer, 2008.

[409] D Medhi and K Ramasamy. Routing algorithms: Shortest path, widest path, and spanning tree. In *Network Routing: Algorithms, Protocols, and Architectures*, chapter 2. Morgan Kaufman Publisher, 2017.

[410] J Bang-Jensen and G Gutin. The floyd-warshall algorithm. In *Diagraphs: Theory, Algorithms and Applications*, chapter 2.3.5. Springer, 2006.

[411] A Opiz. Der floyd-warshall-algorhitmus. pages 1–5, 2000.

[412] KH Rosen. Trees. In *Discrete Matheatics and its Applications, 7th Edition*, chapter 11, pages 745–803. McGraw-Hill Company, 2012.

[413] S Even. Depth-first search. In *Graph Algorithms*, chapter 3, pages 46–65. Cambridge University Press, 2011.

[414] CY Lee. An algorithm for path connections and its applications. *IRE Transactions on Electronic Computers*, 10:346–365, 1961.

[415] V Jarnik. O jistém problému minimálním. *Práce Moravské Prírodovedecké Spolecnosti*, 6:57–63, 1930.

[416] JB Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7:48–50, 1956.

[417] J Nesetril, E Milkova, and H Nesetrilova. Otakar boruvka on minimum spanning tree problem: translation of both the 1926 papers, comments, history. *Discrete Mathematics*, 233:3–36, 2001.

[418] DR Karger, PN Klein, and RE Tarjan. A randomized linear-time algorithm to find minimum spanning trees. *Journal of the Association for Computing Machinery*, 42:328–341, 1995.

[419] B Chazelle. A minimum spanning tree algorithm with inverse-ackermann type complexity. *Journal of the Association for Computing Machinery*, 47:1028–1047, 2000.

[420] S Pettie and V Ramachandran. An optimal minimum spanning tree algorithm. *Journal of the Association for Computing Machinery*, 49:16–34, 2002.

[421] A Caylay. On the analytic forms called trees, with applications to the theory of chemical combinations. *Report of the British Association for the Advancement of Science*, 45:257–305, 1875.

[422] T Heckmann, W Schwanghart, and JD Phillips. Graph theory - recent developments of its application in geomorphology. *Geomorphology*, 243:130–146, 2015.

[423] S Tillem, J von Dongen, IA Brazil, and A Baskin-Sommers. Psychopathic traits are differentially associated with efficiency of neural communication. *Psychophysiology*, 55, 2018.

[424] N Patel and KM Patel. A survey on: Enhancement of minimum spanning tree. *International Journal of Engineering Research and Applications*, 5:6–10, 2015.

[425] A Kasperski, P Kobylanski, M Kulej, and P Zielinski. Minimizing maximal regret in discrete optimization problems with interval data. 2004.

[426] P Kouvelis and G Yu. In *Robust discrete optimization and its applications*. Kluwer Academic Publishers, 1997.

[427] B Roy. Robustness in operational research and decision aiding: A multifaceted issue. *European Journal of Operational Research*, 200:629–638, 2010.

[428] H Yaman, OE Karasan, and M Pinar. The robust spanning tree problem with interval data. *Operations Research Letters*, 29:31–40, 2001.

[429] I Aron and van Hentenryck P. A constraint satisfaction approach to the robust spanning tree with interval data. *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, 2002.

[430] R Montemanni and LM Gambardella. A branch and bound algorithm for the robust spanning tree problem with interval data. *European Journal of Operational Research*, 161:771–779, 2005.

[431] OE Karasan, MC Pinar, and H Yamen. The robust shortest path problem with interval data. *Computers and Operations Research*, 2002.

[432] R Montemanni and LM Gambardella. An exact algorithm for the robust shortest path problem with interval data. *Computers and Operations Research*, 31:1667–1680, 2004.

[433] R Montemanni, LM Gambardella, and AV Donati. A branch and bound algorithm for the robust shortest path problem with interval data. *Operations Research Letters*, 32:224–232, 2004.

[434] A Kaspersky and P Zielinski. Minimizing maximal regret in linear assignment problems with interval data. 2004.

[435] RE Tarjan. Sensitivity analysis of minimum spanning trees and shortest path trees. *Information Processing Letters*, 14:30–33, 1982.

[436] B Dixon, M Rauch, and RE Tarjan. Verification and sensitivity analysis of minimum spanning trees in linear time. *SIAM Journal on Computing*, 21:1184–1192, 1992.

[437] S Pettie. Sensitivity analysis of minimum spanning trees in sub-inverse-ackermann time. *Algorithms and Computation. ISAAC 2005. Lecture Notes in Computer Science*, 3827, 2005.

[438] S Butler and F Chung. Spectral graph theory. In *Handbook of linear algebra*. CRC Press, 2017.

[439] F Chung. Eigenvalues and the laplacian of a graph. In *Spectral Graph Theory*, chapter 1. American Mathematical Society, 2009.

[440] NMM de Abreu. Old and new results on algebraic connectivity of graphs. *Linear Algebra and its Applications*, 423:53–73, 2007.

[441] M Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23:298–305, 1973.

[442] M Reidelbach, F Betz, RM Mäusle, and Petra Imhof. Proton transfer pathways in an aspartate-water cluster sampled by a network of discrete states. *Chemical Physics Letters*, 659:169–175, 2016.

[443] M Reidelbach, M Weber, and P Imhof. Prediction of perturbed proton transfer networks. *PLOS ONE*, 13:e0207718, 2018.

[444] BR Brooks, CL Brooks III, AD Mackerell, L Nilsson, RJ Petrella, B Roux, Y Won, G Archontis, C Bartels, S Boresch, A Caflisch, L Caves, Q Cui, A R Dinner, M Feig, S Fischer, J Gao, M Hodoscek, W Im, K Kuczera, T Lazaridis, J Ma, V Ovchinnikov, E Paci, R W Pastor, C B Post, J Z Pu, M Schaefer, B Tidor, R M Venable, H L Woodcock, X Wu, W Yang, D M York, and M Karplus. Charmm: The biomolecular simulation program. *Journal of Computational Chemistry*, 30:1545–1516, 2009.

[445] W Thiel. Mndo version 6.1. *Max-Planck Institut für Kohlenforschung*, Mühlheim a. d. Ruhr, 2004.

[446] F Noé, D Krachtus, JC Smith and S Fischer. Transition networks for the comprehensive characterization of complex conformational change in proteins. *Journal of Chemical Theory and Computation*, 2:840–857, 2006.

[447] S Nosé. A unified formulation of the constant temperature molecular-dynamics methods. *Journal of Chemical Physics*, 81:511–519, 1984.

[448] WG Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31:1695–1697, 1985.

[449] AD MacKerell Jr et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*, 102:3586–3616, 1998.

[450] D agnella and G Voth. Structure and dynamics of hydronium in the ion channel gramicidin a. *Biophysical Journal*, 70:2043–2051, 1996.

[451] WL Jorgensen, J Chandrasekhar, JD Madura, RW Impey, and ML Klein. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79:926–935, 1983.

[452] A Hassanali, F Giberti, J Cuny, TD ühne, and M Parinello. Proton transfer through the water gossamer. *PNAS*, 110:13723–13728, 2013.

[453] L Martini, A Kells, R Covino, G Hummer, N-V Buchete, and E Rosta. Variational identification of markovian transition states. *Physical Review X*, 7, 2017.

[454] X Wu, W Thiel, S Pezeshki, and H Lin. Specific reaction path hamiltonian for proton transfer in water: Reparametrized semiempirical models. *Journal of Chemical Theory and Computation*, 9:2672–2686, 2013.

[455] M Bagherpoor Helabad, T Ghane, M Reidelbach, AL Woelke, EW Knapp, and P Imhof. Protonation-state-dependent communication in cytochrome c oxidase. *Biophysical Journal*, 113:817–828, 2017.

[456] HM Berman, J Westbrook, F Zeng, G Gilliland, TN Bhat, H Weissig, IN Shindiyalov, and PE Bourne. The protein databank. *Nucleic Acid Research*, 28:283–422, 2000.

[457] A MacKerell, J Wiorkiewicz-Kuczera, and M Karplus. An all atom empirical energy function for the simulation of nucleic acids. *Journal of the American Chemical Society*, 117:11946–11975, 1995.

[458] JB Klauda, RM Venable, JA Freitas, JW O'Connor, DJ Tobias, C Mondragon-Ramirez, I Vorobyov, A MacKerell, and RW Pastor. Update of the charmm all-atom additive force field for lipids: validation on six lipid types. *The Journal of Physical Chemistry B*, 114:7830–7843, 2010.

[459] T Ghane, RF Gorriz, S Wrzalek, S Volkenandt, F Dalatieh, M Reidelbach, and P Imhof. Hydrogen-bonded network and water dynamics in the d-channel of cytochrome c oxidase. *Journal of Membrane Biology*, 251:299–314, 2018.

[460] L Varanasi and J Hosler. Subunit iii-depleted cytochrome c oxidase provides insight into the process of proton uptake by proteins. *Biochimica et Biophysica Acta*, 1817:545–551, 2012.

[461] BR Brooks, RE Bruccoleri, BD Olafson, DJ States, S Swaminathan, and M Karplus. Charmm: A program for macromlecular energy minimization and dynamics calculation. *Journal of Computational Chemistry*, 4:187–217, 1983.

[462] JC Meza. Steepest descent. *WIREs Computational Statistics*, 2:719–722, 2010.

[463] WC Davidon. Variable metric method for minimisation. *A.E.C. Research and Development Report ANL-5990*, 1952.

[464] R Fletcher and MJD Powell. A rapidly convergent descent method for minimization. *The Computer Journal*, 6:163–168, 1963.

[465] WC Davidon. Variable metric method for minimisation. *SIAM journal on Optimization*, 1:1–17, 1991.

[466] MR Hestenes and E Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49:409–436, 1952.

[467] JJ Sylvester. A demonstration of the theorem that every homogeneous quadratic polynomial is reducible by real orthogonal substitutions to the form of a sum of positive and negative squares. *Philosophical Magazine*, 4:138–142, 1852.

[468] EML Beale. A derivation of conjugate gradients. In J Abadie, editor, *Numerical methods for non-linear optimization*. Academic Press, New York, 1972.

[469] FJ Gisdon, M Culka, and GM Ullmann. Pycpr - a python-based implementation of the conjugate peak refinement (cpr) algorithm for finding transition state structures. *Journal of Molecular Modeling*, 22:242, 2016.

[470] C Choi and R Elber. Reaction path study of helix formation in tetrapeptides: Effect of side chains. *Journal of Chemical Physics*, 94:751, 1991.

[471] R Elber and M Karplus. A method for determining reaction paths in large molecules: Application to myoglobin. *Chemical Physics Letters*, 139:375–380, 1987.

[472] R Czerminski and R Elber. Self-avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems. *International Journal of Quantum Chemistry*, 38:167–185, 1990.

[473] A Ulitsky and R Elber. A new technique to calculate steepest descent paths in flexible polyatomic systems. *Journal of Chemical Physics*, 92:1510–1511, 1990.

[474] D Sheppard, R Terrell, and G Henkelman. Optimization methods for finding minimum energy paths. *Journal of Chemical Physics*, 128, 2008.

[475] RE Gillian and KR Wilson. Shadowing, rare events, and rubber bands. a variational verlet algorithm for molecular dynamics. *Journal of Chemical Physics*, 97, 1992.

[476] AD Becke. Density-functional exchange-energy approximation with correct asymptotic behaviour. *Physical Reviews A*, 38:3098–3100, 1988.

[477] C Lee, B Yang, and RG Par. Development of the colle-salvetti correlation energy formula into a functional of the electron density. *Physical Reviews B*, 37:735–739, 1988.

[478] AD Becke. A new mixing of hartree-fock and local density-functional theories. *Journal of Chemical Physics*, 98:1372–1377, 1993.

[479] AD Becke. Density-functional thermochemistry. iii. the role of exact exchange. *Journal of Chemical Physics*, 98:5648–5652, 1993.

[480] A Hellweg, C Hättig, S Höfener, and W Klopper. Optimized accurate auxiliary basis sets for ri-mp2 and ri-cc2 calculations for the atoms rb to rn. *Theoretical Chemical Accounts*, 117:587–597, 2007.

[481] R Ahlrichs, M Bar, M Haser, H Horn, and C Kolmel. Electronicstructurecalculations on workstation computers: The program system turbomole. *Chemical Physics Letters*, 162:165–169, 1989.

[482] S Sadhukhan, D Munoz, C Adamo, and GE Scuseria. Predicting proton transfer barriers with density functional methods. *Chemical Physics Letters*, 306:83–87, 1999.

List of Figures

List of Tables

Appendix A

Minimization Algorithms

Appendix A provides a detailed introduction of the steepest descent (*SD*) minimization algorithm [291], the conjugate gradient (*CG*) minimization algorithm [289], and the adopted basis Newton-Raphson (ABNR) minimization algorithm [461].

Steepest Descent Minimization

The *SD* minimization, developed in 1847 by *Cauchy* [291], is an algorithm which allows the determination of local potential energy minima $\boldsymbol{x_m}$ as the limit of sequences of states $\boldsymbol{x_i}$ (with $i = 0, 1, ..., m$) on the *PES*, using only the gradient $\boldsymbol{g}$ of V. Here, $\boldsymbol{x_0}$ represents some initial approximation of $\boldsymbol{x_m}$. The basic idea behind the algorithm is, that a step from some position $\boldsymbol{x_i}$ to the next position $\boldsymbol{x_{i+1}}$, in the direction of $-\boldsymbol{g_i}$, reduces the potential energy. The only difficulty now is to define a proper size α_i of the individual steps.

The *SD* minimization algorithm starts at $\boldsymbol{x_0}$. Each succeeding position is then defined as

$$\boldsymbol{x_{i+1}} = \boldsymbol{x_i} - \alpha_i \boldsymbol{g_i} \quad \text{with} \quad i = 0, 1, ..., m-1 \quad \text{and} \quad \boldsymbol{g_i} = \boldsymbol{g}(\boldsymbol{x_i}), \tag{A.1}$$

where α_i is defined by a line search

$$\alpha_i = \arg\min_{\alpha} V(\boldsymbol{x_i} - \alpha_i \boldsymbol{g_i}) \tag{A.2}$$

until $\boldsymbol{x_m}$ is reached within some tolerance ϵ.

Thus, the main work of the *SD* minimization algorithm lies in the determination of $\boldsymbol{g}$ and α. Overall, the algorithm is easy to implement and the storage requirements are low. However, the convergence of the algorithm is rather slow, when considering

real-life problems [462].

At the end of the 20th century, *Barzilai and Borwein* [294], proposed a modification of the *SD* minimization algorithm. In their procedure $\boldsymbol{x_{i+1}}$ is defined as

$$\boldsymbol{x_{i+1}} = \boldsymbol{x_i} - \frac{1}{\alpha_i}\boldsymbol{g_i} \quad \text{with} \quad i = 0, 1, ..., m-1 \quad \text{and} \quad \boldsymbol{g_i} = \boldsymbol{g}(\boldsymbol{x_i}), \tag{A.3}$$

where α_i is determined as

$$\alpha_i = \frac{\boldsymbol{s_{i-1}^T y_{i-1}}}{\boldsymbol{s_{i-1}^T s_{i-1}}} \tag{A.4}$$

with

$$\boldsymbol{s_{i-1}} = \boldsymbol{x_i} - \boldsymbol{x_{i-1}} \quad \text{its transpose} \quad \boldsymbol{s_{i-1}^T} \quad \text{and} \quad \boldsymbol{y_{i-1}} = \boldsymbol{g_i} - \boldsymbol{g_{i-1}}. \tag{A.5}$$

For $i = 0$, $\boldsymbol{x_{i-1}}$ and $\boldsymbol{g_{i-1}}$ are not available. Therefore, the original *SD* minimization algorithm is used for the first step. Overall, the modifications introduced by *Barzilai and Borwein* improve the performance of the *SD* minimization algorithm significantly.

Conjugate Gradient Minimization

The identification of local potential energy minima $\boldsymbol{x_m}$ on the *PES* using conjugate directions, or to be more precise **H**-conjugate directions, received a great deal of attention [463, 290, 464, 465]. In principle, the procedure followed by the *CG* minimization algorithm is equal to the procedure followed by the *SD* minimization algorithm, i.e. $\boldsymbol{x_m}$ is determined as the limit of a sequence of states $\boldsymbol{x_i}$ (with $i = 0, 1, ..., m$) derived by successive line minimizations on the *PES*. The only difference between both minimization algorithms is the definition of the search direction, $-\boldsymbol{g}$ or **H**-conjugate for the *SD* minimization algorithm or the *CG* minimization algorithm, respectively.

Two directions, $\boldsymbol{s_i}$ and $\boldsymbol{s_j}$ ($i \neq j$), are said to be **H**-conjugate if they satisfy the following condition:

$$\boldsymbol{s_i^T H s_j} = 0. \tag{A.6}$$

In 1952 *Hestenes and Stiefel* derived a formula for the determination of **H**-conjugate gradients, without a direct evaluation of **H** [466]. Here, the conjugate directions $\boldsymbol{s_{i+1}}$ (with $i = 0, 1, ..., m-2$) are defined as linear combination of $\boldsymbol{g_{i+1}}$

and the preceding direction $\boldsymbol{s_i}$

$$\boldsymbol{s_{i+1}} = -\boldsymbol{g_{i+1}} + \beta_i \boldsymbol{s_i} \quad \text{with} \quad i = 0, 1, ..., m-2 \quad \text{and} \quad \boldsymbol{g_{i+1}} = \boldsymbol{g}(\boldsymbol{x_{i+1}}), \tag{A.7}$$

where β_i is defined as

$$\beta_i = \frac{\boldsymbol{g_{i+1}^2}}{\boldsymbol{g_i^2}}. \tag{A.8}$$

Using eqs. A.7 and A.8 *Fletcher and Reeves* formulated a minimization algorithm. Here, the initial direction $\boldsymbol{s_0}$ is defined by the steepest descent ($\boldsymbol{s_0} = -\boldsymbol{g}$), while the directions $\boldsymbol{s_1}$ till $\boldsymbol{s_{m-1}}$ are derived using eqs. A.7 and A.8. The update of $\boldsymbol{x_i}$ is performed by

$$\boldsymbol{x_{i+1}} = \boldsymbol{x_i} - \alpha_i \boldsymbol{s_i} \quad \text{with} \quad i = 0, 1, ..., m-1. \tag{A.9}$$

Overall, the *CG* minimization algorithm is easy to implement, only requires the storage of three vectors ($\boldsymbol{g_i}$, $\boldsymbol{g_{i+1}}$, and $\boldsymbol{s_i}$) at a time, and displays quadratic convergence, i.e. on a D-dimensional, quadratic *PES*, the algorithm identifies the potential energy minimum in $\leq D$ steps. For other *PES*, the process is iterative and a convergence test is required [289].

Adopted Basis Newton-Raphson Minimization

The *ABNR* minimization algorithm, developed in 1983 by *States et al* [461] as a combination of the Newton-Raphson (*NR*) minimization algorithm and the *SD* minimization algorithm, is a widely used algorithm for the identification of local potential energy minima $\boldsymbol{x_m}$ on the *PES*. As before, $\boldsymbol{x_m}$ is determined as the limit of a sequence of states $\boldsymbol{x_i}$ (with $i = 0, 1, ..., m$).

At first, the most important aspects of the *NR* minimization algorithm are introduced. Starting from a Taylor expansion of the *PES* around the current state $\boldsymbol{x_i}$

$$V(\boldsymbol{x_{i+1}}) \approx V(\boldsymbol{x_i}) + (\boldsymbol{x_{i+1}} - \boldsymbol{x_i})^T \boldsymbol{g_i} + \frac{1}{2}(\boldsymbol{x_{i+1}} - \boldsymbol{x_i})^T \mathbf{H_i} (\boldsymbol{x_{i+1}} - \boldsymbol{x_i}) \tag{A.10}$$

with $\boldsymbol{g_i} = \boldsymbol{g}(\boldsymbol{x_i})$ and $\mathbf{H_i} = \mathbf{H}(\boldsymbol{x_i})$, the gradient at the succeeding state, $\boldsymbol{g_{i+1}}$, is defined as

$$\boldsymbol{g_{i+1}} \approx \boldsymbol{g_i} + \mathbf{H_i}(\boldsymbol{x_{i+1}} - \boldsymbol{x_i}). \tag{A.11}$$

Assuming a local potential energy minimum of the *PES* at $\boldsymbol{x_{i+1}}$, i.e. $\boldsymbol{g_{i+1}} = \mathbf{0}$, reduces eq. A.11 to a linear equation system,

$$\boldsymbol{x_{i+1}} = \boldsymbol{x_i} - \mathbf{H_i^{-1}}\boldsymbol{g_i} \quad \text{with} \quad \mathbf{H}_i^{-1} \quad \text{the inverse of} \quad \mathbf{H}_i \tag{A.12}$$

Thereby, the *NR* minimization algorithm is able to converge towards $\boldsymbol{x_m}$ within a few iteration steps, at least if eq. A.10 is valid. However, despite the fast convergence, in terms of iteration steps, the calculation and handling of $\mathbf{H}$ in every iteration step is rather time consuming [292].

The *ABNR* minimization algorithm attenuates this drawback, applying the *NR* minimization algorithm only to a subspace of the system and combining it with an *SD* step

$$\boldsymbol{x_{i+1}} = \boldsymbol{x_i} + \Delta\boldsymbol{x_i^{NR}} + \Delta\boldsymbol{x_i^{SD}}. \tag{A.13}$$

Here, $\Delta\boldsymbol{x_i^{SD}}$ is defined as

$$\Delta\boldsymbol{x_i^{SD}} = \frac{\boldsymbol{g_i^*}}{|\boldsymbol{g_i^*}|}\alpha_i \quad \text{with} \quad \boldsymbol{g_i^*} = -\boldsymbol{g}(\boldsymbol{x_i} + \Delta\boldsymbol{x_i^{NR}}) \tag{A.14}$$

and α_i, the stepsize, which is adjusted by the sign of the dot product

$$\boldsymbol{g}(\boldsymbol{x_i} + \Delta\boldsymbol{x_i^{NR}}) \cdot \boldsymbol{g}(\boldsymbol{x_i} + \Delta\boldsymbol{x_i^{NR}} + \Delta\boldsymbol{x_i^{SD}}) \quad \begin{cases} < 0 & \alpha_i \text{ is decreased} \\ > 0 & \alpha_i \text{ is increased} \end{cases}. \tag{A.15}$$

In the first steps of the *ABNR* minimization algorithm only the *SD* minimization algorithm is applied, i.e. $\Delta\boldsymbol{x_i^{NR}} = \mathbf{0}$, thus reducing $\boldsymbol{g_i^*}$ to $\boldsymbol{g_i}$. Following a number of *SD* steps, the last n displacements on the *PES* are used as basis vectors, to derive $\Delta\boldsymbol{x_i^{NR}}$,

$$\Delta\boldsymbol{x_i^{NR}} = \sum_{l=1}^{n} \Delta\boldsymbol{x_{li}} c_{li} \quad \text{with} \quad \Delta\boldsymbol{x_{li}} = \boldsymbol{x_{i-l}} - \boldsymbol{x_i} \tag{A.16}$$

A Taylor expansion of eq. A.11 around $\boldsymbol{x_i}$ with $\Delta\boldsymbol{x_i^{NR}}$ as displacement vector gives

$$\sum_{l'=1}^{n}\sum_{l=1}^{n} \Delta\boldsymbol{x_{il'}} \left[\boldsymbol{g_{i-l}} - \boldsymbol{g_i}\right] c_{li} = -\sum_{l'=1}^{n} \Delta\boldsymbol{x_{il'}}\boldsymbol{g_k}. \tag{A.17}$$

Eq. A.17 is a set of n equations, which can be solved for the coefficients c_{li} by diagonalizing the small $n \times n$ matrix. The determination of $\Delta\boldsymbol{x_i^{NR}}$ then occurs via

eq. A.16. An estimate of $\boldsymbol{g}_i^*$ is defined as follows:

$$\boldsymbol{g}_i^* \approx \boldsymbol{g}_i + \sum_{l=1}^{n} \left[\boldsymbol{g}_{i-l} - \boldsymbol{g}_i\right] c_{li}. \tag{A.18}$$

Finally, $\Delta \boldsymbol{x}_i^{SD}$ and $\boldsymbol{x}_{i+1}$ are determined using eqs. A.14 and A.13, respectively. Overall, the *ABNR* minimization algorithm requires roughly equal numbers of iteration steps compared to the *CG* minimization algorithm, however, the efficiency is much higher [461, 293].

Appendix B

Transition Pathway Search Algorithms

Appendix B provides a detailed introduction of the conjugate peak refinement (*CPR*) algorithm [92] and the nudged elastic band (*NEB*) algorithm [91] with the climbing image [315] and improved tangents [314] approach.

Conjugate Peak Refinement

The *CPR* algorithm, developed in 1992 by *Karplus et al* [92], is an algorithm which allows the determination of transition states of a D-dimensonal *PES* forming a continuous transition pathway connecting two pre-defined local potential energy minima. The method is based on and an extension of two procedures, published twenty and thirty years earlier [288, 289]. Apart from the structures of the local potential energy minima, the algorithm requires the evaluation of V and $\boldsymbol{g}$. Cost-intensive evaluations of $\mathbf{H}$ are not required. The algorithm is able to refine several transition states at once, returning a pathway along which every local potential energy maximum corresponds to a transition state [92].

$\mathbf{H}$ is a real symmetric matrix and thus diagonizable by a conjugate basis set. The eigenvectors of $\mathbf{H}$ form such a set. Therefore, it holds that

$$\mathbf{E}^T\mathbf{H}\mathbf{E} = \mathbf{D}, \tag{B.1}$$

where $\mathbf{E}$ and $\mathbf{E}^T$ are the matrix of eigenvectors of $\mathbf{H}$ and its transpose, respectively, and $\mathbf{D}$ is a matrix containing the eigenvalues of $\mathbf{H}$ on the diagonal. In principle, other basis sets $\mathbf{S}$ transforming $\mathbf{H}$ into different diagonalized matrices can be defined by using vectors $\boldsymbol{s_j}$ $(j = 0, ..., D-1)$ conjugate with respect to $\mathbf{H}$ (cf. eq A.6).

According to *Sylvester's theorem of inertia* the number of positive and negative values on the diagonal of matrices, diagonalized with different basis sets, is invariant [467]. As described before, at a transition state **H** has one negative and $D-1$ positive eigenvalues. Hence, in the vicinity of a transition state any set of **H**-conjugate vectors has one direction $\boldsymbol{s_0}$ along which there is a local potential energy maximum and $D-1$ directions $\boldsymbol{s_1}$ till $\boldsymbol{s_{D-1}}$ along which there are local potential energy minima. The set of conjugate vectors, required to converge to a transition state, can be built recursively using the formula of Beale [468], starting from an arbitrary vector $\boldsymbol{s_0}$ along which there is a local potential energy maximum. The vectors are then defined as follows:

$$\boldsymbol{s_0} \quad (arbitrary) \tag{B.2}$$

$$\boldsymbol{s_1} = -\boldsymbol{g_1} + \frac{\boldsymbol{g_1}^T(\boldsymbol{g_1}-\boldsymbol{g_0})}{\boldsymbol{s_0}^T(\boldsymbol{g_1}-\boldsymbol{g_0})}\boldsymbol{s_0} \tag{B.3}$$

$$\boldsymbol{s_j} = -\boldsymbol{g_j} + \frac{\boldsymbol{g_j}^T(\boldsymbol{g_1}-\boldsymbol{g_0})}{\boldsymbol{s_0}^T(\boldsymbol{g_1}-\boldsymbol{g_0})}\boldsymbol{s_0} + \frac{|\boldsymbol{g_j}|^2}{|\boldsymbol{g_{j-1}}|^2}\boldsymbol{s_{j-1}} \quad j>1 \tag{B.4}$$

with $\boldsymbol{g_j}$ the gradient at $\boldsymbol{y_j}$, the local potential energy maximum along $\boldsymbol{s_{j-1}}$. Setting $\boldsymbol{s_0} = -\boldsymbol{g_0}$ reduces eqs. B.2 - B.4 to eq. A.7 from the *CG* minimization algorithm (Appendix A) [92, 469, 288].

In Fig. B.1 the most important steps of the *CPR* algorithm are depicted on an exemplary *PES* (cf. Appendix C), spanned by two reaction coordinates, χ_1 and χ_2. $\boldsymbol{R}$ and $\boldsymbol{P}$ are representing two local potential energy minima of the *PES*. Furthermore, $\boldsymbol{S_1}$ and $\boldsymbol{S_2}$ are representing two transition states of the *PES* (Fig. B.1, top left). In a first round of the *CPR* algorithm an arbitrary direction $\boldsymbol{s_0}$ needs to be defined, along which $\boldsymbol{R}$ and $\boldsymbol{P}$ are representing local potential energy minima with some local potential energy maximum at an intermediate point. Therefore, a set of states is distributed along the straight line connecting $\boldsymbol{R}$ and $\boldsymbol{P}$, revealing an energy maximum in the vicinity of $\boldsymbol{y_0^1}$ (not depicted here). Hence, $\boldsymbol{s_0}$ can be defined as straight line connecting $\boldsymbol{R}$ and $\boldsymbol{P}$. A line maximization is carried out along $\boldsymbol{s_0}$, revealing an energy maximum along $\boldsymbol{s_0}$ at $\boldsymbol{y_1^1}$. According to eq. B.3, the gradients at $\boldsymbol{y_0^1}$ and $\boldsymbol{y_1^1}$ ($\boldsymbol{g_0}$ and $\boldsymbol{g_1}$) are used to define the conjugate direction $\boldsymbol{s_1}$. Following a line minimization along $\boldsymbol{s_1}$, the first intermediate point of the CPR algorithm $\boldsymbol{x_1}$ is identified, which is already close to one of the transition states (Fig. B.1, top right). In a second round of the *CPR* algorithm, a local potential energy maximum is searched along the straight line connecting $\boldsymbol{R}$ and $\boldsymbol{x_1}$ and found in the vicinity

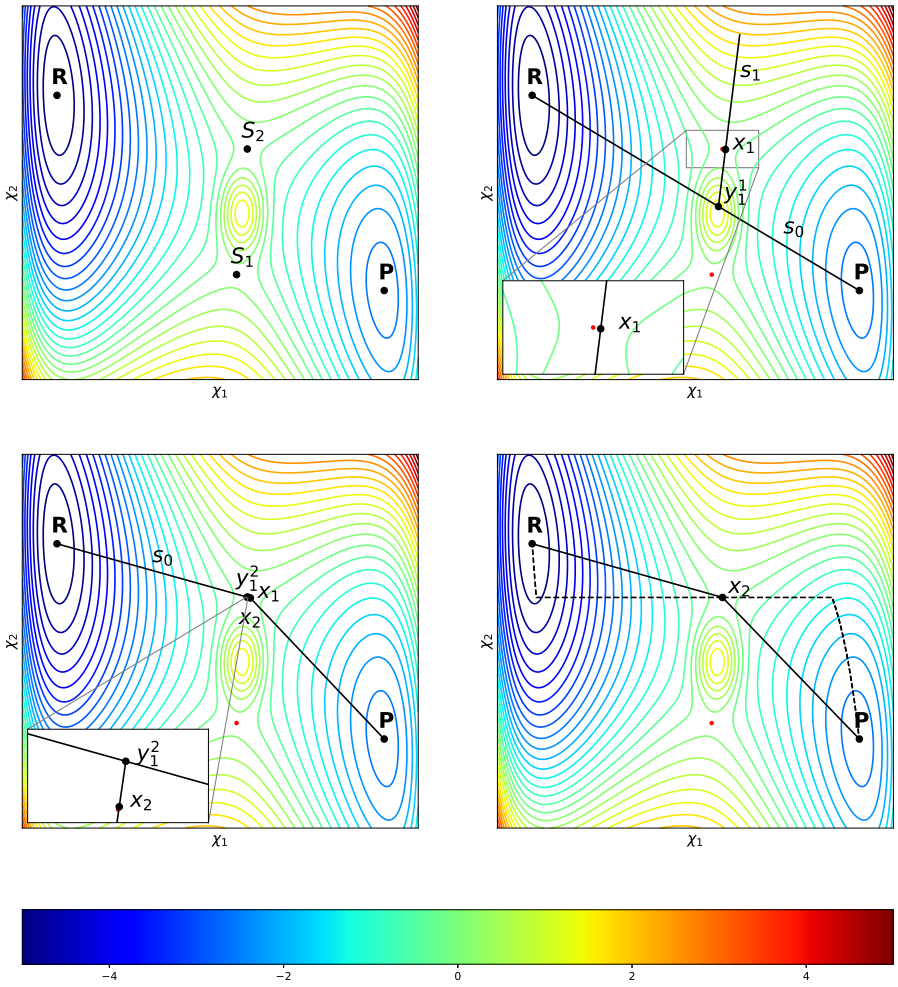


Figure B.1: Visualization of the *CPR* algorithm on an exemplary *PES* (high energies in red, low energies in blue) spanned by χ_1 and χ_2. Reactant, product and transition states are denoted by $\boldsymbol{R}$, $\boldsymbol{P}$, and $\boldsymbol{S_{1/2}}$, respectively (top left). Arbitrary direction $\boldsymbol{s_0}$ as straight line connecting $\boldsymbol{R}$ and $\boldsymbol{P}$, its maximum $\boldsymbol{y_1^1}$, and the conjugate direction $\boldsymbol{s_1}$ with the energy minimum $\boldsymbol{x_1}$ (top right). Arbitrary direction $\boldsymbol{s_0}$ as straight line connecting $\boldsymbol{R}$ and $\boldsymbol{x_1}$, its maximum $\boldsymbol{y_1^2}$, and the conjugate direction $\boldsymbol{s_1}$ with the energy minimum $\boldsymbol{x_2}$ (bottom left). *MEP* connecting $\boldsymbol{R}$ and $\boldsymbol{P}$ via $\boldsymbol{x_2}$ as dashed line (bottom right). Further information can be found in the text.

of $\boldsymbol{y_0^2}$ (not depicted here). Hence, $\boldsymbol{s_0}$ is defined as straight line connecting $\boldsymbol{R}$ and $\boldsymbol{x_1}$. Once again, a line maximization provides the energy maximum $\boldsymbol{y_1^2}$ along $\boldsymbol{s_0}$, which allows the definition of the conjugate direction $\boldsymbol{s_1}$. A line minimization along $\boldsymbol{s_1}$ provides the second intermediate point $\boldsymbol{x_2}$, which is representing a transition state of the reaction according to some gradient criterion $\boldsymbol{g^T(x_2)g(x_2)} < \epsilon$ (Fig. B.1, bottom left). An investigation of the straight line connecting $\boldsymbol{x_1}$ and $\boldsymbol{P}$ is not required, due to the absence of a local potential energy maximum in between. Finally, an *SD* minimization (Appendix A), starting at the transition state, provides the *MEP* connecting $\boldsymbol{R}$ and $\boldsymbol{P}$ (Fig. B.1, bottom right). In principle, it would also

be possible that $\boldsymbol{x_2}$ is not fulfilling the gradient criterion and is thus no transition state of the reaction, but still a peak along the segments. In such a situation, $\boldsymbol{s_0}$ for the next round of the *CPR* algorithm is defined as the tangent of the path segment [288, 92].

Nudged Elastic Band

The *NEB* algorithm, developed in 1998 by *Jónsson et al* [91] on top of various other algorithms [470, 471, 472, 473], is an algorithm which allows the determination of transition pathways, and corresponding transition states, on a D-dimensonal *PES* connecting pre-defined local potential energy minima. Each pathway is traced out by P images ($\boldsymbol{R_i}$ with $i = 0, ..., P-1$, initial state $\boldsymbol{R_0}$, final state $\boldsymbol{R_{P-1}}$) connected by $P-1$ springs. Starting from an initial pathway, usually a straight line formed by equidistant images connecting $\boldsymbol{R_0}$ and $\boldsymbol{R_{P-1}}$, each intermediate image is adjusted iteratively along the *NEB* algorithm, until it represents a potential energy minimum in all directions perpendicular to the pathway [474, 296]. Apart from the structures of $\boldsymbol{R_0}$ and $\boldsymbol{R_{P-1}}$, the number of images P to distribute and the spring constants k_i, the method requires the evaluation of V and $\boldsymbol{g}$. Cost-intensive evaluations of $\mathbf{H}$ are not required [91].

In a Plain Elastic Band calculation [91], the total force acting on each image $\boldsymbol{F_i}$ is defined as

$$\boldsymbol{F_i} = -\boldsymbol{g_i} + \boldsymbol{F_i^s} \quad \text{with} \quad \boldsymbol{g_i} = \boldsymbol{g}(\boldsymbol{R_i}) \tag{B.5}$$

and $\boldsymbol{F_i^s}$ the spring force

$$\boldsymbol{F_i^s} = k_{i+1}(\boldsymbol{R_{i+1}} - \boldsymbol{R_i}) - k_i(\boldsymbol{R_i} - \boldsymbol{R_{i-1}}). \tag{B.6}$$

Two problems arise when trying to determine the *MEP* by minimizing eq. B.5. These are [475, 91]:

1. The pathway cuts corners (avoids the transition state region) due to the component of $\boldsymbol{F_i^s}$ which is perpendicular to the pathway.

2. The images slide down the pathway (avoiding the transition state and thus reducing its resolution) due to the component of $-\boldsymbol{g_i}$ which is parallel to the pathway.

The *NEB* algorithm solves both problems by projecting out the perpendicular and

parallel component of $\boldsymbol{F_i^s}$ and $-\boldsymbol{g_i}$, respectively. The resulting force on image i is then defined as

$$\boldsymbol{F_i} = -\boldsymbol{g}_i|_\perp + \boldsymbol{F_i^s}|_\parallel \tag{B.7}$$

with $-\boldsymbol{g_i}|_\perp$ and $\boldsymbol{F_i^s}| \parallel$ the perpendicular and parallel component of the gradient and the spring force

$$\boldsymbol{g}_i|_\perp = \boldsymbol{g}_i - \boldsymbol{g}_i \cdot \hat{\boldsymbol{\tau}}_{\boldsymbol{i}} \tag{B.8}$$

$$\boldsymbol{F_i^s}|_\parallel = k_{i+1}(\boldsymbol{R_{i+1}} - \boldsymbol{R_i}) - k_i(\boldsymbol{R_i} - \boldsymbol{R_{i-1}}) \cdot \hat{\boldsymbol{\tau}}_{\boldsymbol{i}}. \tag{B.9}$$

Here, $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}$ represents the unit tangent vector

$$\boldsymbol{\tau_i} = \frac{\boldsymbol{R_i} - \boldsymbol{R_{i-1}}}{|\boldsymbol{R_i} - \boldsymbol{R_{i-1}}|} + \frac{\boldsymbol{R_{i+1}} - \boldsymbol{R_i}}{|\boldsymbol{R_{i+1}} - \boldsymbol{R_i}|} \tag{B.10}$$

in its normalized form

$$\hat{\boldsymbol{\tau}}_{\boldsymbol{i}} = \frac{\boldsymbol{\tau_i}}{|\boldsymbol{\tau_i}|}. \tag{B.11}$$

$\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}$ ensures equidistant images along the pathway. In certain situations, however, $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}$ introduces kinks along the pathway, thereby preventing the convergence towards the transition pathway. Hence, an improved tangent approach was introduced by *Henkelmann et al* [314]. Here, the tangent vector at image i is defined as the vector connecting image i with the neighboring image with higher energy

$$\boldsymbol{\tau_i} = \begin{cases} \boldsymbol{\tau_i^+} & \text{if } V_{i+1} > V_i > V_{i-1} \\ \boldsymbol{\tau_i^-} & \text{if } V_{i+1} < V_i < V_{i-1} \end{cases} \quad \text{with} \quad V_i = V(\boldsymbol{R_i}), \tag{B.12}$$

where

$$\boldsymbol{\tau_i^+} = \boldsymbol{R_{i+1}} - \boldsymbol{R_i} \text{ and } \boldsymbol{\tau_i^-} = \boldsymbol{R_i} - \boldsymbol{R_{i-1}}. \tag{B.13}$$

For situations in which image i is located close to a local potential energy minimum or maximum, $V_{i+1} > V_i < V_{i-1}$ or $V_{i+1} < V_i > V_{i-1}$, the tangent vector becomes an energy weighted average of $\boldsymbol{\tau_i^+}$ and $\boldsymbol{\tau_i^-}$

$$\boldsymbol{\tau_i} = \begin{cases} \boldsymbol{\tau_i^+}\Delta V_i^{max} + \boldsymbol{\tau_i^-}\Delta V_i^{min} & \text{if } V_{i+1} > V_{i-1} \\ \boldsymbol{\tau_i^+}\Delta V_i^{min} + \boldsymbol{\tau_i^-}\Delta V_i^{max} & \text{if } V_{i+1} < V_{i-1} \end{cases}, \tag{B.14}$$

where

$$\Delta V_i^{max} = \max(|V_{i+1} - V_i|, |V_i - V_{i-1}|) \tag{B.15}$$

$$\Delta V_i^{min} = \min(|V_{i+1} - V_i|, |V_i - V_{i-1}|) \tag{B.16}$$

to ensure a smooth switch of the tangent across the local potential energy extremum. As before the tangent vector needs to be normalized (cf eq. B.11).

The *NEB* algorithm can be further improved by using the climbing image approach formulated by *Henkelmann et al* [315]. Here, a separate force calculation is performed for the image with the highest energy i_{max} along the pathway,

$$\boldsymbol{F}_{\boldsymbol{i_{max}}} = -\boldsymbol{g}_{\boldsymbol{i_{max}}} + 2\boldsymbol{g}_{\boldsymbol{i_{max}}} \cdot \hat{\boldsymbol{\tau}}_{\boldsymbol{i_{max}}} \hat{\boldsymbol{\tau}}_{\boldsymbol{i_{max}}}, \tag{B.17}$$

which allows the image i_{max} to move towards the transition state. Thereby, an overall higher resolution of the transition state is achieved. Other improvements of the *NEB* algorithm are for example the Doubly Nudged Elastic Band algorithm [316] or the Cubic Spline approach [317].

In Fig. B.2 individual steps of the *NEB* algorithm, in conjunction with the improved tangent and climbing image approach, are depicted on an exemplary *PES* (cf Appendix C), spanned by two reaction coordinates, χ_1 and χ_2. Two local potential energy minima of the *PES* are represented by $\boldsymbol{R_0}$ and $\boldsymbol{R_3}$. Furthermore, two transition states of the *PES* are represented by $\boldsymbol{S_1}$ and $\boldsymbol{S_2}$. In a first step of the *NEB* algorithm, an initial pathway is defined as straight line connecting $\boldsymbol{R_0}$ and $\boldsymbol{R_3}$. Here, two images, $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$, are distributed along the pathway (Fig. B.2, top left). Following 6 and 732 rounds of the *NEB* algorithm (Fig. B.2, top right and bottom left) one of the transition states is refined with an accuracy of $\epsilon < 1$ and 10^{-4}, respectively. However, despite the accurate refinement of the transition state, two images are not sufficient to refine a transition pathway equal to the *MEP* (Fig. B.2, bottom right).

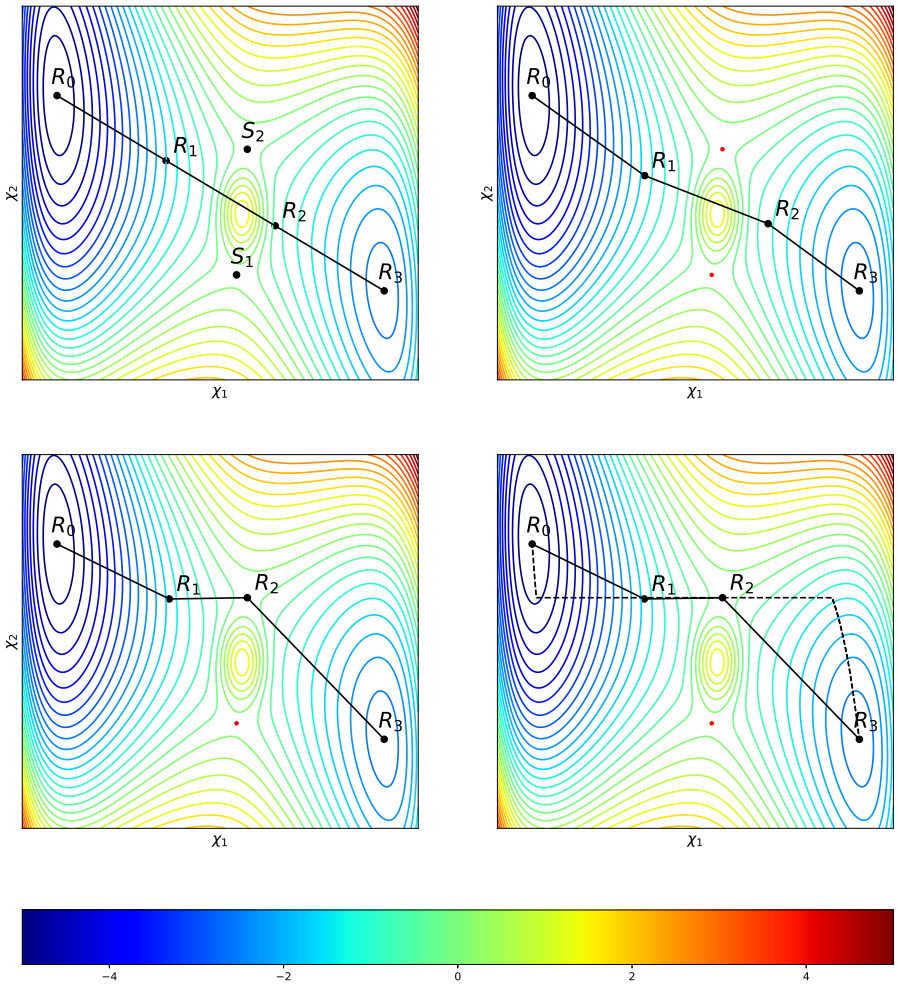


Figure B.2: Visualization of the *NEB* algorithm on an exemplary *PES* (high energies in red, low energies in blue) spanned by χ_1 and χ_2. Reactant, product and transition states are denoted by $\boldsymbol{R_0}$, $\boldsymbol{R_3}$, and $\boldsymbol{S_{1/2}}$, respectively. The initial guess path is depicted as black line, connecting $\boldsymbol{R_0}$ and $\boldsymbol{R_3}$ via the intermediate states $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$ (top left). Pathway following 6 rounds of the *NEB* algorithm, no transition state refined (top right). Pathway following 732 rounds of the *NEB* algorithm, transition state refined (bottom left). Comparison of *NEB* and *SD* pathway (bottom right).

Appendix C

Exemplary Potential Energy Surface

Appendix C provides the basis for the *PES* depicted in Figs B.1 and B.2 to illustrate the *CPR* and *NEB* algorithm. The *PES* is taken from *Henkelmann et al* [100] and mimics a reaction of the form

$$AB + C \rightarrow A + BC \tag{C.1}$$

along a straight line and is defined as:

$$\begin{aligned} V^{LEPS}(r_{AB}, r_{BC}) &= \frac{Q_{AB}}{1+a} + \frac{Q_{BC}}{1+b} + \frac{Q_{AC}}{1+c} \\ &- \left[\frac{J_{AB}^2}{(1+a)^2} + \frac{J_{BC}^2}{(1+b)^2} + \frac{J_{AC}^2}{(1+c)^2} \right. \\ &\left. - \frac{J_{AB}J_{BC}}{(1+a)(1+b)} - \frac{J_{BC}J_{AC}}{(1+b)(1+c)} - \frac{J_{AB}J_{AC}}{(1+a)(1+c)} \right]^{1/2} \end{aligned} \tag{C.2}$$

with $a = 0.05$, $b = 0.80$, and $c = 0.05$.

The Q functions

$$Q(r) = \frac{d}{2}\left(\frac{3}{2}e^{-2\alpha(r-r_0)} - e^{-\alpha(r-r_0)}\right) \tag{C.3}$$

represent the coulomb interactions between the nuclei and the electron clouds, while the quantum mechanical exchange interactions are denoted by J

$$J(r) = \frac{d}{4}\left(e^{-2\alpha(r-r_0)} - 6e^{-\alpha(r-r_0)}\right) \tag{C.4}$$

with $d_{AB} = 4.746$, $d_{BC} = 4.746$, and $d_{AC} = 3.445$.

From the three atoms, atom B is the only one allowed to move along the straight line, thereby reducing the number of variables. Furthermore, a *condensed phase environment* is added using a harmonic oscillator *DOF* coupled to atom B. The resulting potential is defined as

$$V(r_{AB}, x) = V^{LEPS}(r_{AB}, r_{AC} - r_{AB}) + 2k_c \left(r_{AB} - \left(\frac{r_{AC}}{2} - \frac{x}{e}\right)\right)^2 \tag{C.5}$$

with $r_{AC} = 3.742$, $k_c = 0.2025$, and $e = 1.154$. Eq. C.5 is frequently used to model a chemical reaction in a liquid or solid matrix.

In order to achieve a *PES* with two individual saddle points, a Gaussian function

$$G(f, g) = exp\left(-0.5\left(\left(\frac{f}{0.1}\right)^2 + \left(\frac{g}{0.35}\right)^2\right)\right) \tag{C.6}$$

is added to $V(r_{AB}, x)$, resulting in $V^{tot}(r_{AB}, x)$

$$V^{tot}(r_{AB}, x) = V(r_{AB}, x) + 1.5G(r_{AB} - 2.02083, x + 0.272881) \tag{C.7}$$

the *PES* used in Figs B.1 and B.2.

Appendix D

Shortest Pathway Algorithms

Appendix D provides a detailed introduction of the Dijkstra [336] and Floyd-Warshall algorithm [388, 389] used to determine optimal pathways, i.e. the shortest pathway (SP), the widest pathway (WP), and the minimax best pathway (MBP), in simple graphs.

Dijkstra's Algorithm

The Dijkstra algorithm [336], published in 1959, is an algorithm which provides a solution to the single-source SP problem [408]. In the following, the most important steps of the original Dijkstra algorithm are introduced in detail.

Apart from a connected, weighted, positive graph, the algorithm only requires the definition of a source (s) vertex. In a first step the distance to the source vertex, $dist(s)$, is set to zero, while the distance to all other vertices, $dist(v)$ for $v \, \epsilon \, V - s$, is set to infinity. Furthermore, a set of visited vertices S and a queue of vertices Q is defined. S is initially empty, while Q initially contains V. As long as Q is not empty, the algorithm carries out the following steps:

1. Choose a vertex u from Q, add u to S and remove u from Q.
2. For all neighbors of u, $v \, \epsilon \, neighbor(u)$, check if the distance to v, $dist(v)$, is higher than the sum of $dist(u)$ and the weight $\omega(u, v)$.
3. If the condition in point two is fulfilled set $dist(v)$ to $dist(u) + \omega(u, v)$.

Once, Q is empty, the algorithm is concluded and the shortest distance from s to $V - s$ is determined. Instead of performing a blind search of SPs, it is also possible to define a target (t) vertex. Here, the algorithm is concluded once t is removed from Q. An exemplary calculation is depicted in Fig D.1.

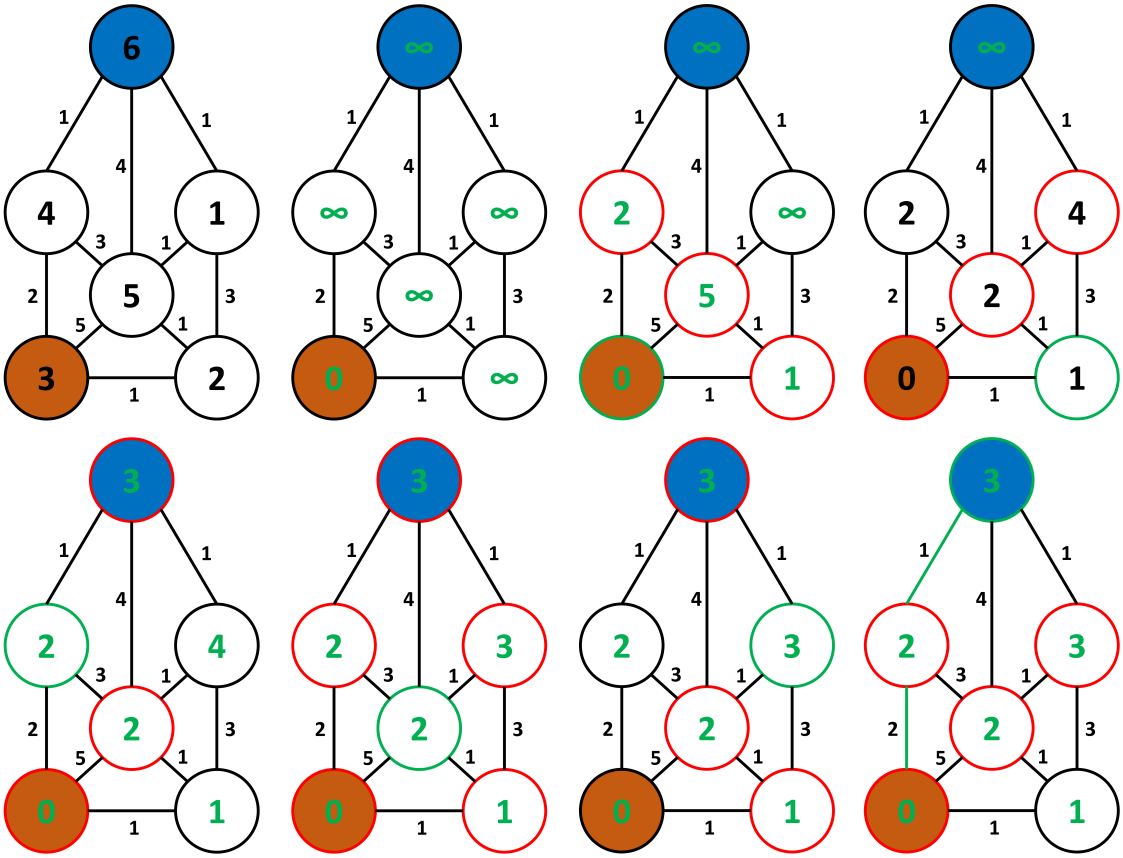


Figure D.1: Dijkstra algorithm for the determination of the SP in a weighted, simple graph with $|V| = 6$ vertices and $|E| = 10$ edges. Top from left to right: 1) Vertices 1, 2, 3, 4, 5, and 6. Source vertex in orange, target vertex in blue. 2) Vertices contain distances, indicated by green numbers. Set $dist(1) = 0$ and $dist(2) = dist(3) = dist(4) = dist(5) = dist(6) = \infty$. The set of visited vertices, $S = \emptyset$, and the queue of vertices $Q = 1, 2, 3, 4, 5, 6$. 3) Set u to vertex with minimal distance, $u = 3$ (green ring), with three neighbors, $v = 2, 4, 5$ (red rings). Adjust distances of vertices 2, 4, and 5. $S = 3$ and $Q = 1, 2, 4, 5, 6$. 4) Set $u = 2$ with three neighbors, $v = 1, 3, 5$. Adjust distances of vertices 1, 3, and 5. $S = 2, 3$ and $Q = 1, 4, 5, 6$. Bottom from left to right: 1) Set $u = 4$ with three neighbors, $v = 3, 5, 6$. Adjust distances of vertices 3, 5, and 6. $S = 2, 3, 4$ and $Q = 1, 5, 6$. 2) Set $u = 5$ with five neighbors, $v = 1, 2, 3, 4, 6$. Adjust distances of vertices 1, 2, 3, 4, and 6. $S = 2, 3, 4, 5$ and $Q = 1, 6$. 3) Set $u = 1$ with three neighbors, $v = 2, 5, 6$. Adjust distances of vertices 2, 5, and 6. $S = 1, 2, 3, 4, 5$ and $Q = 6$. 4) Set $u = 6$ with three neighbors, $v = 1, 4, 5$. Adjust distances of vertices 1, 4, and 5. $S = 1, 2, 3, 4, 5, 6$ and $Q = \emptyset$. The SP is 3-4-6 with a distance of 3 (green path). Target vertex is no longer in Q (and $Q = \emptyset$), the algorithm is concluded.

Due to the condition and distance update in points two and three of the algorithm

$$dist(v) > dist(u) + \omega(u, v) \tag{D.1}$$

$$dist(v) = dist(u) + \omega(u, v) \tag{D.2}$$

the original Dijkstra algorithm provides the *SP*. Modifying the condition and distance update to

$$dist(v) > max\left(dist(u), \omega(u, v)\right) \tag{D.3}$$

$$dist(v) = max\left(dist(u), \omega(u, v)\right) \tag{D.4}$$

provides the *MBP*, while a modification to

$$dist(v) < min\left(dist(u), \omega(u, v)\right) \tag{D.5}$$

$$dist(v) = min\left(dist(u), \omega(u, v)\right) \tag{D.6}$$

provides the *WP*. Furthermore, the modifications in eqs. D.3 to D.6 allow the investigation of graphs with negative edge weights [336, 339, 409].

Floyd-Warshall algorithm

The Floyd-Warshall algorithm [388, 389], published in 1962, is an algorithm which provides a solution to the all-pairs *SP* problem in weighted, simple graphs with positive or negative edge weights, but without negative cycles. Originally, the algorithm only provided the distance between all vertices. However, by the introduction of slight modifications the algorithm also provides the actual *SPs* [410, 411]. In the following, the most important steps of the Floyd-Warshall algorithm are introduced in detail.

Apart from a weighted graph without negative cycles, the algorithm requires no further pre-settings. The underlying idea of the algorithm is that an *SP* connecting the vertices u and v via vertex w, implies that the sub-pathways connecting the vertices u and w and the vertices w and v are also *SPs*. Imagine a graph with $|V|$ vertices labeled $1, 2, ..., |V|$ and a function $sp(i, j, k)$ which determines the *SP* connecting the vertices i and j in the k-th iteration and only considers vertices from the set $1, 2, ..., k$ as intermediate vertices. Consequently, in the $k+1$-th iteration the *SP* connecting the vertices i and j only contains intermediate vertices from the set $1, 2, ..., k+1$. Thus, it holds for each vertex pair i and j in the $k+1$-th iteration, that

the connecting *SP* is either containing intermediate vertices from the set $1, 2, ..., k$ only or it is an *SP* which is additionally containing vertex $k + 1$. Hence, the *SP* in the $k + 1$-th iteration could either stay the same as in the k-th iteration or it could be improved by adding vertex $k + 1$. Overall, the algorithm boils down to the equation,

$$sp(i, j, k) = min\left(sp(i, j, k), sp(i, k + 1, k) + sp(k + 1, j, k)\right), \tag{D.7}$$

and an iteration over i, j, and k. By the introduction of several modifications (eqs. D.3 to D.6) the Floyd-Warshall algorithm also allows the determination of *MBPs* and *WPs* [410, 411], as well as the determination of the corresponding pathways.

Appendix E

Minimum Spanning Tree Algorithms

Appendix E provides a detailed introduction of Boruvka's [417] and Kruskal's algorithm [416] used to determine the *MST* in simple graphs.

Boruvka's algorithm

Boruvka's algorithm [417], published in 1926 and frequently called Sollin's algorithm, allows the determination of an *MST* in a weighted graph, as well as the determination of minimum spanning forests in disconnected, weighted graphs. The algorithm is based on three distinct steps:

1. Identify the edges with minimal edge weights incident to either vertex.
2. Add all edges identified in point one to the *MST* and compress the graph along the identified edges to obtain a pseudograph.
3. Remove edge loops from the compressed graph. If multiple edges are present in the compressed graph, keep those with minimal edge weight.

These steps are repeated until the *MST* (or minimum spanning forest) of the graph is determined. An exemplary determination of an *MST* using Boruvka's algorithm is depicted in Fig. E.1.

Kruskal's algorithm

Kruskal's algorithm [416], published in 1956, allows the determination of an *MST* in a connected, weighted graph, as well as the determination of minimum spanning

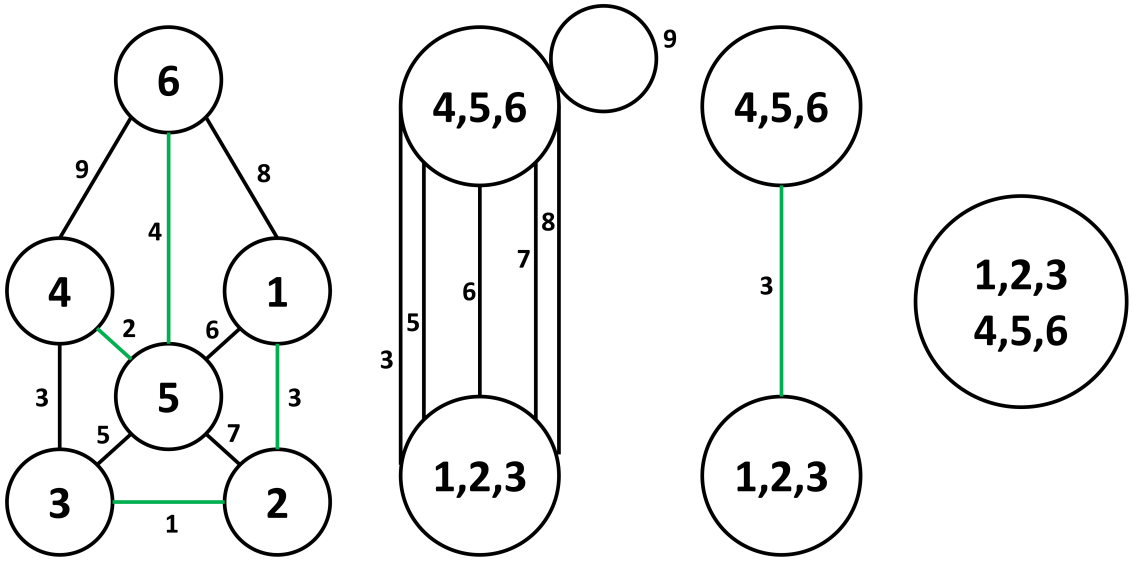


Figure E.1: Boruvka's algorithm for the determination of an *MST* in a weighted, simple graph with $|V| = 6$ vertices and $|E| = 10$ edges. From left to right: 1) Identification of the edges with minimal edge weight incident to either vertex (green). 2) Compressed pseudograph with $|V| = 2$ vertices and $|E| = 6$ edges, i.e. five multiple edges and an edge loop. 3) Identification of the edges with minimal edge weight incident to either vertex (green) following the removal of the edge loop and multiple edges. 4) Compressed graph with $|V| = 1$ vertices and $|E| = 0$ edges. The *MST* contains the following edges $1-2$, $2-3$, $3-4$, $4-5$, and $5-6$.

forests in disconnected, weighted graphs. The algorithm is based on two distinct steps:

1. Separation of the graph into vertices and edges, providing a set of $|V|$ one-vertex trees and a set of $|E|$ edges.

2. Removal of the edge with minimal weight from the edge set. If the edge connects two distinct trees, they are combined.

The second point is repeated until the *MST* (or minimum spanning forest) of the graph is determined. An exemplary determination of a *MST* using Kruskal's algorithm is depicted in Fig. E.2.

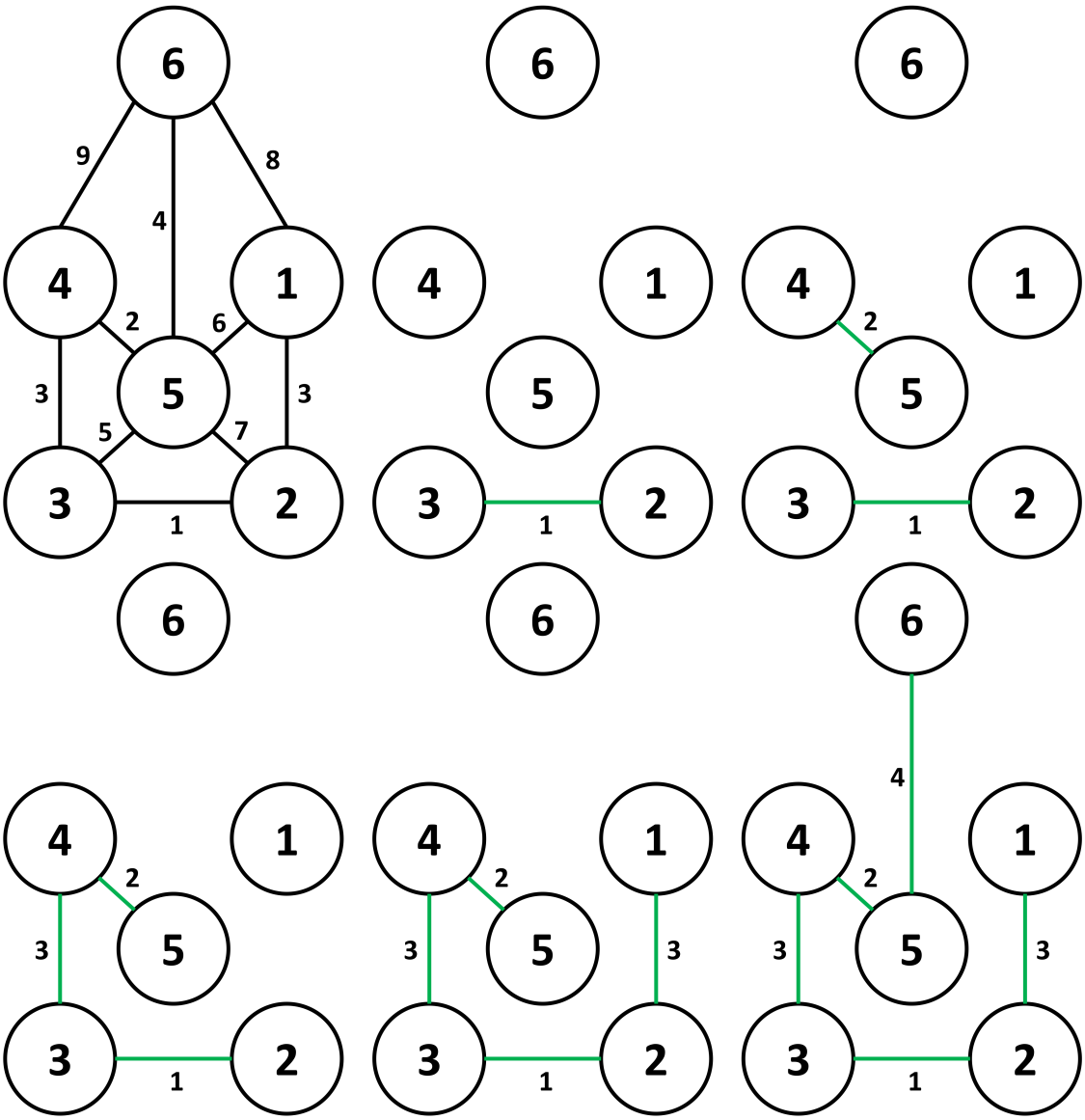


Figure E.2: Kruskal's algorithm for the determination of an *MST* in a weighted, simple graph with $|V| = 6$ vertices and $|E| = 10$ edges. Top from left to right: 1) Initial Graph. 2) Four one-vertex trees and one two-vertex tree connected by the edge $2 - 3$ with minimal edge weight. 3) Two one-vertex trees and two two-vertex tree connected by the edges $2 - 3$ and $4 - 5$ with minimal edge weights. Bottom from left to right: 1) Two one-vertex trees and one four-vertex tree connected by the edges $2 - 3$, $3 - 4$, and $4 - 5$ with minimal edge weights. 2) One one-vertex tree and one five-vertex tree connected by the edges $1 - 2$, $2 - 3$, $3 - 4$, and $4 - 5$ with minimal edge weights. 3) One six-vertex tree connected by the edges $1 - 2$, $2 - 3$, $3 - 4$, $4 - 5$, and $5 - 6$ with minimal edge weights, representing the *MST* of the graph.

Appendix F

Evaluation of the OM2 Method

Appendix F provides an evaluation of the *OM2* method for the calculation of proton transfer processes.

Small Hydrogen Bond Chain

The relative potential energies of three stationary points, i.e. the product state (P), the transition state (TS), and a state with an *HBC*, observed during the translocation of a proton along a small model chain (cf. Fig F.1), were compared following three distinct *QM* potential energy minimizations using the *OM2* method [244, 245] and the *DFT* functionals *BLYP* [476, 477] or *B3LYP* [478, 479] with the *def2-TZVP* basis set [480]. Furthermore, two single-point calculations using the *B3LYP* method with the *def2-TZVP* basis set were performed on top of the *QM* potential energy minimization using the *OM2* and *BLYP* method. Therefore, the CHARMM programme [444] was interfaced to MNDO [445] or Turbomole [481], respectively.

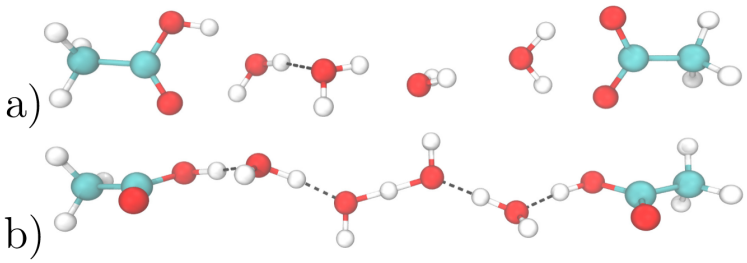


Figure F.1: Model system containing two side chains of aspartate-like molecules, four water molecules, and one excess proton. Two states depicted: a) reactant state and b) intermediate state with a continuous *HBC* preceding the transfer of the excess proton.

The relative potential energies of all three states investigated are comparable

for computations employing the *OM2* and *BLYP* method (cf. Tab. F.1). Both methods, however, deviate from the *B3LYP* method by ≈ 2 kcal/mol, at least for the *HBC* and *TS*, which is in agreement with previous results suggesting an underestimation of proton transfer barriers by *BLYP* [482]. Single-point calculations with the *B3LYP* method on top of the states minimized by the *OM2* and *BLYP*, indicate a structural origin for these deviations. The relative potential energy of the product state is in all cases zero, which is a necessity due to the symmetry of the system.

Table F.1: Relative potential energies (in kcal/mol) for three stationary points observed during the translocation of a proton along a small model chain (cf. Fig. F.1) using the *OM2*, *BLYP/def2-TVZP*, and *B3LYP/def2-TVZP* method. *B3LYP/def2-TVZP//OM2* and *B3LYP/def2-TVZP//BLYP/def2-TVZP* indicate single-point potential energy calculations using the *B3LYP/def2-TVZP* on top of the states minimized by the *OM2* or *BLYP/def2-TVZP* method.

Method	*HBC*	*TS*	*P*
OM2	0.69	7.31	0.00
BLYP/def2-TVZP	0.20	7.30	0.00
B3LYP/def2-TVZP	2.61	9.84	0.00
B3LYP/def2-TVZP//OM2	1.00	7.27	0.00
B3LYP/def2-TVZP//BLYP/def2-TVZP	1.00	7.68	0.01

Transition Network States

The nodes of the six best pathways from Fig. 4.6 were re-minimized using the *BLYP/def2-TVZP* method, re-assigned and subsequently compared to the original nodes. The results of this examination are summarized in Tab. F.2.

Table F.2: Potential energies and assigned states of nodes constituting the six best pathways from the *TN* following the direct sampling approach with *MD* simulations and the direct sampling approach with hydronium ion motions (cf. Fig. 4.6), calculated with the *OM2* and *BLYP/def2-TZVP* method, respectively. Differences in the state assignment are set in bold for the *BLYP*-optimized structures. The potential energies are in kcal/mol, while *RMSD* values are in Å unless stated otherwise.

OM2		*BLYP/def2-TVZP*		Comparison	
States	Rel. ΔE	States	Rel. ΔE	*RMSD*	ΔΔE
0.0.w0.0	0.00	0.0.w0.0	0.00	0.51	0.00
0.0.w1.1	-0.36	0.0.w**2**.1	0.37	0.62	-0.73
0.2.w231.3	-7.49	0.2.w**7**.3	-8.16	0.81	0.67
0.0.w17.3	-4.28	0.**2**.w**27**.3	-8.16	0.81	3.88
0.0.w4.0	-0.14	0.0.w**3**.0	0.00	0.48	-0.14
0.0.w47.2	-7.11	0.0.w**23**.2	-6.58	1.14	-0.53
0.0.w6.14	-6.34	0.0.w**8**.14	-5.06	0.72	-1.28
0.0.w49.11	-2.21	0.0.w**16**.11	-1.47	1.30	-0.73
1.0.w30.11	-0.19	1.0.w**12**.11	0.29	0.43	-0.48
0.1.w39.8	-7.35	0.1.w**24**.8	-4.90	0.36	-2.44
0.1.w40.8	-7.38	0.1.w**14**.8	-4.97	0.35	-2.41
1.0.w21.11	1.47	1.0.w**22**.11	1.44	0.80	0.04
1.0.w24.3	-0.96	1.0.w**11**.3	-3.69	0.56	2.73
0.1.w43.4	-0.06	0.1.w**13**.4	1.87	1.11	-1.93
0.2.w95.7	-6.27	0.2.w**5**.7	-4.54	0.57	-1.72
0.2.w47.3	-7.47	0.2.w**19**.3	-5.73	0.43	-1.73
MSD in kcal/mol				0.69	-0.43
MAD in kcal/mol				0.23	1.22
RMSD in kcal/mol				0.48	1.10

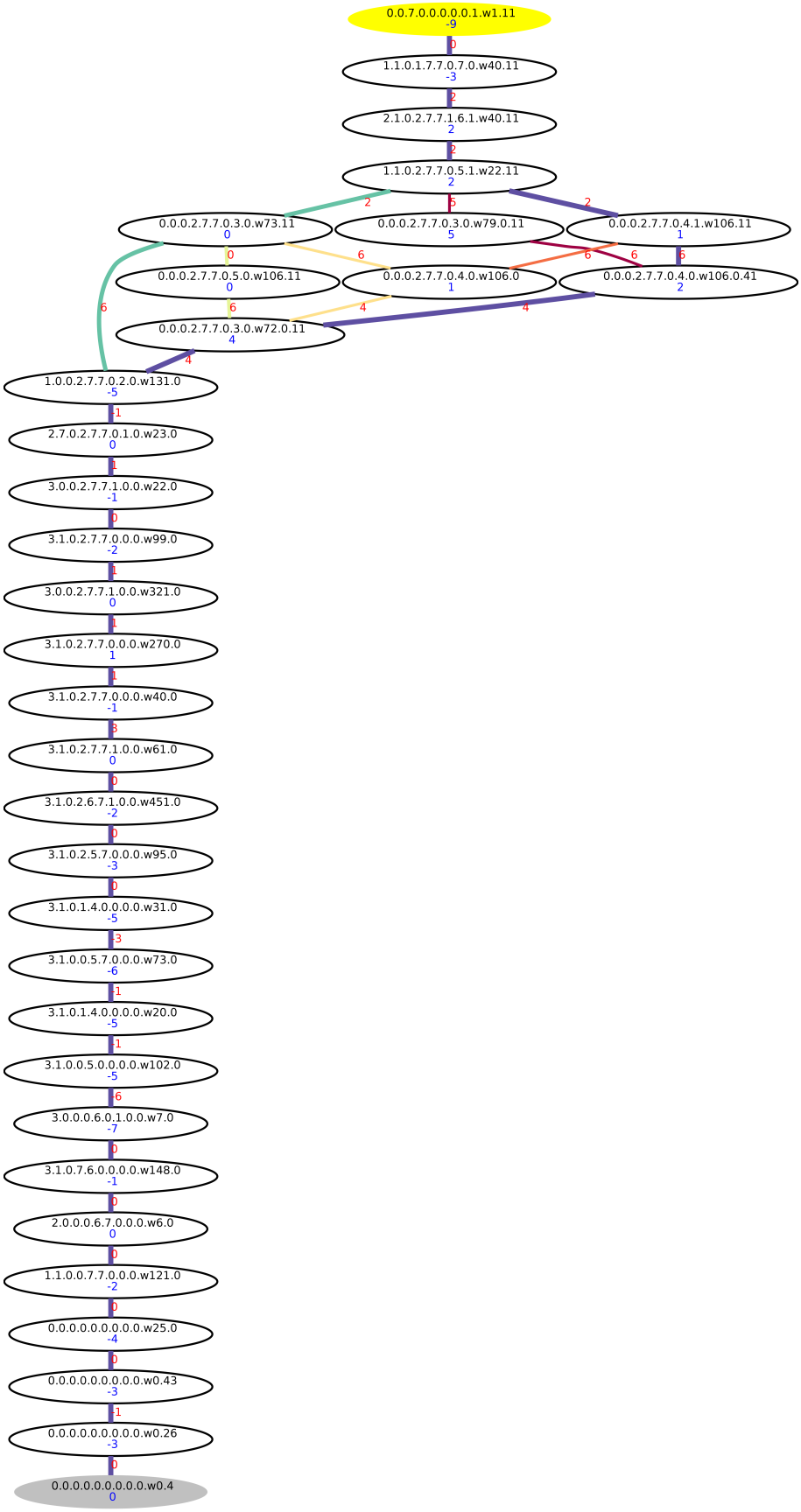
The re-minimizations and re-assignments of the 16 nodes, constituting the six best pathways from Fig. 4.6, using the *OM2* and *BLYP/def2-TVZP* method, revealed similar potential energies, similar protonation states, and similar side chain dihedral angle conformations. Only the side chain dihedral angle of the right aspartate-like molecule of node 0.0.w17.3 is altered along the re-minimizations and re-assignments from SC1=0 to SC1=2. The water patterns, on the other hand, are altered for every state. However, it should be remembered that the translation of a single water molecule by 2 Å or the rotation of a single water molecule by 22.5° leads to a different water pattern. Hence, the node sets minimized by the *OM2* and *BLYP/def2-TVZP* method are rather similar.

Appendix G

D-Channel Proton Transfer

Appendix G provides the complete *TN* for the proton transfer through the D-channel of *CcO*.

Figure G.1: TN describing the translocation of a proton along a model system of the D-channel of CcO, from D132 at the D-channel entrance to E286 at the D-channel exit, following five individual TN calculations using the TN-MD method, i.e. direct sampling of side chain dihedral angles and protonation states and MD simulations for the motion of the water molecules (cf. Tab. 5.2 for a detailed description of either TN calculation). The nodes are shown as ellipses, labeled after a subset of DOFs ($\chi_{1.Y33}$·$\chi_{2.N121}$·$\chi_{2.D132}$·$\chi_{1.N139}$·$\chi_{2.N139}$·$\chi_{3.N139}$·$\chi_{2.S200}$·$\chi_{2.S201}$·$\chi_{3.E286}$·wi.P). The proton poitions follow the residue numbering from Fig 5.7. The ellipses filled in gray and yellow color correspond to the reactant and product state of the overall proton translocation, respectively. Edges are shown as lines. The dashed edge represents 18 subtransitions (shown in Appendix ...). Blue numbers are the potential energies of the respective nodes, red numbers are the potential energies of the highest transition states along the edges used as edge weight. All energies are in kcal/mol, relative to the reactant state, and rounded to integer values for better readability. The *MBP* is bold and highlighted in indigo, the 2nd, 3rd, 4th, 5th, and 6th best pathways are bold and highlighted in turquoise, green, yellow, orange, and red, respectively.


0.0.7.0.0.0.0.0.1.w1.11
-9
0
1.1.0.1.7.7.0.7.0.w40.11
-3
2
2.1.0.2.7.7.1.6.1.w40.11
2
2
1.1.0.2.7.7.0.5.1.w22.11
2
2
5
2
0.0.0.2.7.7.0.3.0.w73.11
0
0.0.0.2.7.7.0.3.0.w79.0.11
5
0.0.0.2.7.7.0.4.1.w106.11
1
0
6
6
6
6
0.0.0.2.7.7.0.5.0.w106.11
0
0.0.0.2.7.7.0.4.0.w106.0
1
0.0.0.2.7.7.0.4.0.w106.0.41
2
6
6
4
4
0.0.0.2.7.7.0.3.0.w72.0.11
4
4
1.0.0.2.7.7.0.2.0.w131.0
-5
1
2.7.0.2.7.7.0.1.0.w23.0
0
1
3.0.0.2.7.7.1.0.0.w22.0
-1
0
3.1.0.2.7.7.0.0.0.w99.0
-2
1
3.0.0.2.7.7.1.0.0.w321.0
0
1
3.1.0.2.7.7.0.0.0.w270.0
1
1
3.1.0.2.7.7.0.0.0.w40.0
-1
3
3.1.0.2.7.7.1.0.0.w61.0
0
0
3.1.0.2.6.7.1.0.0.w451.0
-2
0
3.1.0.2.5.7.0.0.0.w95.0
-3
0
3.1.0.1.4.0.0.0.0.w31.0
-5
3
3.1.0.0.5.7.0.0.0.w73.0
-6
1
3.1.0.1.4.0.0.0.0.w20.0
-5
1
3.1.0.0.5.0.0.0.0.w102.0
-5
6
3.0.0.0.6.0.1.0.0.w7.0
-7
0
3.1.0.7.6.0.0.0.0.w148.0
-1
0
2.0.0.0.6.7.0.0.0.w6.0
0
0
1.1.0.0.7.7.0.0.0.w121.0
-2
0
0.0.0.0.0.0.0.0.0.w25.0
-4
0
0.0.0.0.0.0.0.0.0.w0.43
-3
1
0.0.0.0.0.0.0.0.0.w0.26
-3
0
0.0.0.0.0.0.0.0.0.w0.4
0

Appendix H

TN Prediction Code

Appendix H provides the TN prediction code.

try.py Thu Dec 13 21:05:36 2018 1

```
import re
import numpy as np
from collections import defaultdict
from heapq import *
from functools import cmp_to_key
import sys
import os

### Input standard: 0/1 for MST or MBP prediction
### (Standard files names are e_init, e_pert, rp_init, rp_pert)
### e_init / e_pert n1_n2 [bar] (edges by node name and barrier known)
### rp_init / rp_pert [r;p]
### User defined input: 0/1 [e_init] [e_pert] [rp_init] [rp_pert]

if len(sys.argv) == 2:
    method=int(sys.argv[1])
    e1='e_init'
    e2='e_pert'
    if1='rp_init'
    if2='rp_pert'
elif len(sys.argv) == 6:
    method=int(sys.argv[1])
    e1=sys.argv[2]
    e2=sys.argv[3]
    if1=sys.argv[4]
    if2=sys.argv[5]
else:
    print("Check yourInput.")
    exit()

# Read in edge lists, r and p (init and pert)

def readin(e1,e2,if1,if2):
    # unperturbed
    with open(e1) as f:
        E_un = f.readlines()
    E_un = [x.strip('\n') for x in E_un]
    E_un = [re.split(' |_',x) for x in E_un]
    # get unperturbed init, final, bar
    init_un = [item[0] for item in E_un]
    final_un = [item[1] for item in E_un]
    bar_un = [item[2] for item in E_un]
    # unperturbed nodes
    all_un=init_un[:]
    all_un.extend(final_un)
    nodes_un = [*{*all_un}]
    #perturbed
    with open(e2) as f:
        E_per = f.readlines()
    E_per = [x.strip('\n') for x in E_per]
    E_per = [re.split(' |_',x) for x in E_per]
    # get perturbed init, final
    init_per = [item[0] for item in E_per]
    final_per = [item[1] for item in E_per]
    bar_per = np.array([999 if x=='' else x for x in [item[2] for item in E_per]])
    # perturbed nodes
    all_per=init_per[:]
    all_per.extend(final_per)
    nodes_per = [*{*all_per}]
    # all nodes
    nodes_all=nodes_un[:]
    nodes_all.extend(nodes_per)
    nodes_all = [*{*nodes_all}]
    nodes_all.sort()
    with open(if1) as f:
        RP_un = f.readlines()
    RP_un = [x.strip('\n') for x in RP_un]
    R1=nodes_all.index(RP_un[1])
```

```
try.py       Thu Dec 13 21:05:36 2018       2
    P1=nodes_all.index(RP_un[0])
    with open(if2) as f:
        RP_per = f.readlines()
    RP_per = [x.strip('\n') for x in RP_per]
    R2=nodes_all.index(RP_per[1])
    P2=nodes_all.index(RP_per[0])
    return(E_un,E_per,R1,P1,R2,P2,init_un,final_un,bar_un,init_per,final_per,bar_per,nodes_
un,nodes_per,nodes_all)

E_un,E_per,R1,P1,R2,P2,init_un,final_un,bar_un,init_per,final_per,bar_per,nodes_un,nodes_pe
r,nodes_all=readin(e1,e2,if1,if2)

# Get initial Transition Matrix

D1=np.full((len(nodes_all),len(nodes_all)),np.inf)
for i in range(len(init_un)):
    if nodes_all.index(init_un[i])!=nodes_all.index(final_un[i]):
        D1[nodes_all.index(init_un[i]),nodes_all.index(final_un[i])]=bar_un[i]
        D1[nodes_all.index(final_un[i]),nodes_all.index(init_un[i])]=bar_un[i]
D2=np.full((len(nodes_all),len(nodes_all)),np.inf)
for i in range(len(init_per)):
    if nodes_all.index(init_per[i])!=nodes_all.index(final_per[i]):
        D2[nodes_all.index(init_per[i]),nodes_all.index(final_per[i])]=bar_per[i]
        D2[nodes_all.index(final_per[i]),nodes_all.index(init_per[i])]=bar_per[i]

def adj(X):
    # Converts X to A with A(i,j)=0 if X(i,j)=inf or 1 if X(i,j)!=inf..
    A=np.zeros_like((X));
    A[X!=np.inf]=1;
    return A

# Determine Edge changes over perturbation.

DIFF=adj(D1)-adj(D2)
r1,c1=np.where(DIFF==1)
r2,c2=np.where(DIFF==-1)

if len(r1)!=0:
    S=np.unique(np.sort(np.array([r1,c1]).T),axis=0)
    SUB=len(S)
else:
    S=np.array([])
    SUB=0
if len(r2)!=0:
    A=np.unique(np.sort(np.array([r2,c2]).T),axis=0)
    A=A[np.lexsort(A[:,::-1].T)]
    ADD=len(A)
else:
    A=np.array([])
    ADD=0

# Check if initial Transition Network is connected

EVD1=np.sort(np.round(np.linalg.eigvals(np.diag(np.sum(adj(D1),axis=1))-adj(D1)).real,5))
if EVD1[len(D1)-(len(nodes_un)-1)]<=0:
    print('Caution, your initial TN is not connected connected.')

# Determine the MST of D1 using Kruskal's algorithm.

def fnd(i,rep):
    while i != rep[i]:
        i = rep[i]
    return(i)

def kruskal(D1):
    # Convert logical adjacent matrix X to neighbors' matrix and get weights
    r,c=np.where(D1!=np.inf)
    ne=np.unique(np.sort(np.transpose(np.array([r,c]))),axis=0)
```

try.py Thu Dec 13 21:05:36 2018 3

```
    w=D1[ne[:,0],ne[:,1]]
    # Number of Nodes (N) and Edges (Ne)
    N=np.max(ne)+1
    NE=np.size(ne,axis=0)
    # edge index (one if edge in MST)
    lidx=np.zeros(NE)
    # Sort edges with respect to weight
    ind=np.argsort(w)
    w=w[ind]
    ne=ne[ind]
    # Initialize: Assign every Node to itself
    rep=np.arange(N)
    rnk=np.zeros(N)
    # Run Kruskal
    for k in range(NE):
        i = ne[k,0];
        j = ne[k,1];
        if fnd(i,rep) != fnd(j,rep):
            lidx[k] = 1;
            r_i=fnd(i,rep)
            r_j=fnd(j,rep)
            if rnk[r_i]>rnk[r_j]:
                rep[r_j]=r_i
            else:
                rep[r_i]=r_j
                if rnk[r_i]==rnk[r_j]:
                    rnk[r_j]+=1
    MST=ne[lidx==1]
    return MST

MST1=kruskal(D1)

# Output MST with "real" edges names

MST1_output=[None]*len(MST1)
print("Initial MST:")
for i in range(len(MST1)):
    xxx=str(nodes_all[MST1[i,0]])+'_'+str(nodes_all[MST1[i,1]])
    print(xxx)

# Determine MST sensitivity and coarse graining barrier by Dijkstra algorithm.

def dijkstra(edges, f, t):
    g = defaultdict(list)
    for l,r,c in edges:
        g[l].append((c,r))
    q, seen = [(0,f,())], set()
    while q:
        (cost,v1,path) = heappop(q)
        if v1 not in seen:
            seen.add(v1)
            path = (v1, path)
            if v1 == t: return (cost, path)
            for c, v2 in g.get(v1, ()):
                if v2 not in seen:
                    heappush(q, (max(cost,c), v2, path))
    return np.array([np.inf])

def dirMST(D,MST):
    MSTEDGES=np.zeros((len(MST)*2,3))
    MSTEDGES[:,[0,1]]=np.append(MST,np.flip(MST,axis=1)).reshape(2*len(MST),2)
    MSTEDGES[:,2]=D[MSTEDGES[:,0].astype(int),MSTEDGES[:,1].astype(int)]
    return(MSTEDGES)

def sensitivity(D,MST):
    # Determines sensitivity of MST towards all existing non-MST edges.
    [r,c]=np.where(D!=np.inf);
    ALL=np.unique(np.sort(np.append(r,c,axis=0).reshape((2,len(r))).T),axis=0);
```

```
try.py       Thu Dec 13 21:05:36 2018       4
    # Determine existing non-MST edges.
    e_list=np.setdiff1d(ALL.view([('', ALL.dtype)] * ALL.shape[1]), MST.view([('', MST.dtyp
e)] * MST.shape[1])).view(ALL.dtype).reshape(-1, ALL.shape[1])
    # Assign sensitivity values (best path value - the actual edge value).
    MSTEDGES=dirMST(D,MST)
    bar=np.zeros((len(e_list),1));
    for i in range(0,len(e_list)):
        bar[i,0]=dijkstra(MSTEDGES,e_list[i,0],e_list[i,1])[0]-D[e_list[i,0],e_list[i,1]];
    e_list=np.column_stack((e_list,bar));
    e_list=e_list[e_list[:,2].argsort()[::-1]];
    return (np.concatenate((e_list,np.hstack((MST,np.zeros((len(MST),1))))),axis=0),MSTEDGE
S)

# Sensitity of MST1 & directed MST edges.
[LIST,MSTEDGES1]=sensitivity(D1,MST1);

# Coarse graining barrier bar1.
bar1=dijkstra(MSTEDGES1,R1,P1)[0]

def SNtrue(D,M,criteria):
    # Determines coarse grained nodes.
    # Set up adjacency matrix A according to initial MST (D2(MST1)) and SN criteria.
    A=np.full((len(D),len(D)),np.inf)
    A[M[:,0].astype(int),M[:,1].astype(int)]=M[:,2]
    A[A>=criteria]=np.inf;
    A=adj(A)
    # Determine Eigenvalues, -vectors of the Laplacian of A.
    [v,l]=np.linalg.eig(np.diag(np.sum(A,axis=1))-A);
    # Sort v1 (and l1) in ascending order.
    idx=v.argsort()[::];
    v=v[idx];
    l=l[:,idx];
    # Loop through eigenvalues equal to 0. Add up (in f) corrsponding eigenvectors to seper
ate the nodes
    # Avoid multiplicity
    k=1;
    f=np.zeros_like((l[:,0]));
    while k<=len(l)-1:
        if np.round(abs(v[k]),5)==0.0:
            f=f+np.round(l[:,k],5)*np.sqrt(100*np.round(np.random.uniform(0,5.0),5)*k);
            k=k+1;
        else:
            break
    # Identify seperate node sets.
    [w,ic1]=np.unique(f,axis=0,return_inverse=True);
    L1=np.full((1+max(ic1).astype(int),len(D)-(max(ic1).astype(int)-1)),np.inf);
    for i in range(0,max(ic1)+1):
        L1[i,0:len(ic1[ic1==i])]=np.sort(np.where(ic1==i))[::-1];
    L1=L1[L1[:,0].argsort()];
    return L1

# Determine Coarse Grained Nodes in D1
L1=SNtrue(D1,MSTEDGES1,bar1)
# Compress D2 with coarse grained Nodes of D1. C is compressed Matrix, E edge dictionary.
def Mcomp(D,L):
    # Determines coarse grained matrix
    No=np.sum(adj(L),axis=1);
    # Coarse Graining
    E=[];
    C=np.full((len(L),len(L)),np.inf);
    for i in range(0,len(L)):
        s=D[L[i,0:No[i].astype(int)].astype(int),:];
        for k in range(i+1,len(L)):
            C[i,k]=s[:,L[k,0:No[k].astype(int)].astype(int)].min(axis=1).min(axis=0);
            C[k,i]=C[i,k];
            if C[i,k]!=np.inf:
                # Write out Edge dictionary
                x=np.argwhere(s[:,L[k,0:No[k].astype(int)].astype(int)]==s[:,L[k,0:No[k].as
```

```
try.py       Thu Dec 13 21:05:36 2018       5
type(int)].astype(int)].min(axis=1).min(axis=0));
                E.extend([i,k,L[i,x[0,0]],L[k,x[0,1]]]);
    E=np.asarray(E).reshape((len(E)//4,4));
    E[:,2:3]=np.sort(E[:,2:3]);
    return (E,C)

E,C=Mcomp(D2,L1)

# Delete and Add Edges to e_list according to Add and Sub

def mat_cmp(x, y):
    for i in [2, 0, 1]:
        if x[i] < y[i]:
            return -1
        elif x[i] > y[i]:
            return +1
    else:
        return 0

if len(S)!=0:
    # Delete S from LIST
    idx=[];
    for i in range(0,len(S)):
        x=np.where(np.all(LIST[:,:2].astype(int)==S[i,:],axis=1))
        if any(map(len,x)):
            idx.extend(x);
    idx=np.asarray(idx);
    LIST=np.delete(LIST,idx,axis=0);
if len(A)!=0:
    # Add A to LIST
    bar_xxx=np.zeros((len(A),1));
    for i in range(0,len(A)):
        bar_xxx[i,0]=1/(dijkstra(MSTEDGES1,A[i,0],A[i,1])[0]+0.0001);
    A=np.column_stack((A,bar_xxx));
    LIST=np.asarray(sorted(np.vstack((abs(LIST),A)),key=cmp_to_key(mat_cmp)))
    LIST2=np.copy(LIST);
    for i in range(0,len(L1)):
        idx=L1[i,np.isfinite(L1[i,:])]
        LIST2=np.delete(LIST2,np.where(np.sum(np.isin(LIST2[:,[0,1]],idx).astype(int),axis=
1)==2),axis=0)

# Delete intra coarse grained node edges from LIST
LIST3=np.copy(LIST2);
LIST3=np.column_stack((LIST3,D2[LIST3[:,0].astype(int),LIST3[:,1].astype(int)]));
ll=np.zeros((len(LIST3),2))
for i in range(0,len(LIST3)):
        [r1,c1]=np.where(L1==LIST3[i,0]);
        [r2,c2]=np.where(L1==LIST3[i,1]);
        if r2>=r1:
                ll[i,:]=[r1,r2];
        else:
                ll[i,:]=[r2,r1];
LIST3=np.column_stack((LIST3,ll));
LIST3_save=np.copy(LIST3)
DEL=np.column_stack((np.asarray(np.where(LIST3[:,3]<bar1)).T,LIST3[np.asarray(np.where(LIST
3[:,3]<bar1)).T,[4,5]]));
if len(DEL)>0:
        for i in range(0,len(DEL)):
                y=np.where(np.all(LIST3[DEL[i,0].astype(int)+1:,[4,5]]==DEL[i,[1,2]],axis=1
))
                if any(map(len,y)):
                    LIST3=np.delete(LIST3,y+(DEL[i,0].astype(int)+1),axis=0);

LIST3=LIST3[sum(LIST3[:,3]!=999):,:]

LIST3_output=[None]*len(LIST3)
if len(LIST3>0):
    print("Edges to be calculated:")
```

```
try.py        Thu Dec 13 21:05:36 2018        6
    for i in range(len(LIST3)):
        xxx=str(nodes_all[LIST3[i,0].astype(int)])+'_'+str(nodes_all[LIST3[i,1].astype(int)
])
        print(xxx)
else:
    # Determine perturbed barrier
    print("No further Edges to be calculated.")
    DC_pert=np.full((len(L1),len(L1)),np.inf)
    for i in range(len(LIST3_save)):
        if LIST3_save[i,3]<DC_pert[LIST3_save[i,4].astype(int),LIST3_save[i,5].astype(int)]
:
            DC_pert[LIST3_save[i,4].astype(int),LIST3_save[i,5].astype(int)]=LIST3_save[i,3
]
            DC_pert[LIST3_save[i,5].astype(int),LIST3_save[i,4].astype(int)]=LIST3_save[i,3
]
    R2c,c1=np.where(L1==R2)
    P2c,c2=np.where(L1==P2)
    bar_per=dijkstra(dirMST(DC_pert,kruskal(DC_pert)),int(R2c),int(P2c))[0]
    print("The maximal transition barrier of the optimal pathway in the perturbed TN is " +
 str(bar_per) + " kcal/mol.")
```

Zusammenfassung

Methoden wie *Molekulardynamik* (*MD*) Simulationen und Algorithmen zur Bestimmung optimaler Übergangswege liefern oft eine unzureichende Beschreibung quantenmechanischer Effekte und beeinflussen die erhaltenen Ergebnisse, wodurch nur eine limitierte Beschreibung komplexer Reaktion möglich ist. Die Methode der *Übergangsnetzwerke* (*TN*s) löst dieses Problem, in dem komplexe Reaktionen in simple Teilreaktionen aufgespalten werden. Jede Teilreaktion verbindet dabei zwei Energieminima der *potentiellen Energiefläche* über einen *minimalen Energieweg*. Die Summe aller Teilreaktionen bildet das *TN*, in dem der optimale Übergangsweg der komplexen Reaktion durch einfache Algorithmen bestimmt werden kann. Das große Problem der Methode der *TNs* ist der exponentielle Anstieg an Teilreaktion mit steigender Zahl an *Freiheitsgraden* der komplexen Reaktion. Hierdurch ist die Methode der *TNs* oftmals nicht zur Untersuchung komplexer Reaktionen geeignet.

Im Rahmen der vorliegenden Dissertation wird die Methode der *TNs* durch zwei zusätzliche Methoden optimiert. Diese sind: 1) die *TN-MD Methode*, die das diskrete Sampling der Methode der *TNs* nur für *Freiheitsgrade* verwendet, die durch substantielle Energiebarrieren getrennt sind, während *MD* Simulationen für das Sampling von *Freiheitsgraden* mit geringen Energiebarrieren verwendet werden und 2) die *TN Vohersage Methode*, die ein bekanntes, initiales *TN* zur Bestimmung unbekannter *TNs* in veränderten Umgebungen nutzt. Beide Methoden führen zu einer signifikanten Reduktion der Rechenkosten und erhalten wichtige Eigenschaften der *TNs*.

Zusätzlich wird die *TN-MD Methode* zur Bestimmung optimaler Protonen Transportwege in einem Modell des D-Kanals der *Cytochrom c Oxidase* verwendet. Hierdurch konnten mehrere Wege bestimmt werden, die zuvor erlangte Erkenntnisse reproduzieren und erweitern, z.B. maximale Übergangsbarrieren und korrespondierende Übergangszustände.

Eidesstattliche Erklärung

Hiermit erkläre ich alle Hilfsmittel und Hilfen angegeben und auf dieser Grundlage die Arbeit selbstständig verfasst zu haben. Zudem versichere ich, dass die vorliegende Arbeit in keinem früheren Promotionsverfahren eingereicht worden ist.

Ort, Datum Unterschrift