

RESEARCH ARTICLE

Prediction of perturbed proton transfer networks

Marco Reidelbach¹, Marcus Weber², Petra Imhof^{1*}

1 Institute for Theoretical Physics, Freie Universität Berlin, Berlin, Germany, **2** Mathematics for Life and Materials Sciences, Zuse Institute Berlin, Berlin, Germany

* petra.imhof@fu-berlin.de



Abstract

The transfer of protons through proton translocating channels is a complex process, for which direct samplings of different protonation states and side chain conformations in a transition network calculation provide an efficient, bias-free description. In principle, a new transition network calculation is required for every unsampled change in the system of interest, e.g. an unsampled protonation state change, which is associated with significant computational costs. Transition networks void of or including an unsampled change are termed unperturbed or perturbed, respectively. Here, we present a prediction method, which is based on an extensive coarse-graining of the underlying transition networks to speed up the calculations. It uses the minimum spanning tree and a corresponding sensitivity analysis of an unperturbed transition network as initial guess and refinement parameter for the determination of an unknown, perturbed transition network. Thereby, the minimum spanning tree defines a sub-network connecting all nodes without cycles and minimal edge weight sum, while the sensitivity analysis analyzes the stability of the minimum spanning tree towards individual edge weight reductions. Using the prediction method, we are able to reduce the calculation costs in a model system by up to 80%, while important network properties are maintained in most predictions.

OPEN ACCESS

Citation: Reidelbach M, Weber M, Imhof P (2018) Prediction of perturbed proton transfer networks. PLoS ONE 13(12): e0207718. <https://doi.org/10.1371/journal.pone.0207718>

Editor: Xuhui Huang, Hong Kong University of Science and Technology, HONG KONG

Received: August 8, 2018

Accepted: November 5, 2018

Published: December 12, 2018

Copyright: © 2018 Reidelbach et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying this study have been uploaded to the Harvard Dataverse and are accessible using the following doi: <https://doi.org/10.7910/DVN/YTOKHX>.

Funding: This work was supported by Project C5 in CRC 1078, Deutsche Forschungsgemeinschaft, www.dfg.de (PI). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The translocation of protons from one side of a biological membrane, e.g. the inner mitochondrial membrane, to the other is an exceptionally important process in nature [1]. To a small extent, protons are able to permeate through membranes on their own [2, 3]. The bulk of such translocations, however, occurs via proton-permeable channels [4]. One of the simplest channels is provided by *gramicidine* [5], a cation-selective, water-filled pore, which allows the translocation of protons at high rates [6]. In contrast to the translocation of other cations no translational motion of the water molecules within the channel is required [7], instead the proton translocation is believed to occur grotthus-like [8], i.e. proton “hops” between neighboring water molecules by exchanging covalent and hydrogen bonds to the donor and acceptor oxygen, along a single-file water chain [5]. Hence, a pore and a water chain (or a network of water molecules) is enough to facilitate the proton transfer. Interestingly, however, in another type

of water-filled channels, the *aquaporines* [9], no proton permeation was observed [10]. Molecular dynamics (MD) simulations provided an explanation for these observations, revealing an interruption of the required water chain at an arginine, whose positive charge is repelling protons and thus preventing their translocation [11]. A more sophisticated regulation of the proton translocation was proposed for the *viral proton channel M2* [12]. Here, the side chains of four unprotonated histidine residues occlude the channel [13]. However, the protonation of at least two of them results in a conformational change, which allows the formation of a continuous water chain [13] and thus proton transfer. Such gates were also proposed for other proton transfer channels, e.g. the D-channel of *Cytochrome c Oxidase (CcO)* [14]. The proposed D-channel gate, however, is formed by an asparagine, rendering conformational changes due to a direct protonation hardly probable. Instead, MD simulations revealed a correlation between the conformation of the gate and the protonation state of other residues inside the channel [15, 16]. Furthermore, MD simulations revealed a correlation between the channel hydration and the protonation state of individual residues [17, 18]. Hence, protons seem to be able to alter important channel properties, while they are being transferred. To receive a comprehensive picture of the proton translocation through individual proton transfer channels, the well orchestrated interplay of local effects, e.g. conformational changes upon protonation, single proton “hops” with subsequent water shell re-orientations or proton induced water chain formations [18–21], and distant effects, e.g. conformational or hydrational changes upon protonation of distant residues, proton translocations along water chains, needs to be elucidated.

Due to the rareness of transition events, straight forward molecular simulations are not able to sample transition paths efficiently. Steered molecular simulations are only efficient and applicable, if it is possible to pre-define the relevant reaction coordinates of the described complex mechanisms. If it is not possible to anticipate molecular mechanisms, then path optimization methods are applicable [22]. Instead of trying to find the minimum energy path (MEP) by sampling, one can solve a corresponding optimization problem. On the basis of some initial guess-path a local optimization method determines the MEP in a very efficient way.

Over time a broad range of path optimization methods has been developed, e.g. the Nudged Elastic Band method (NEB) [23] with several modifications [24–26], the Conjugate Peak Refinement method (CPR) [27], the Ridge method [28], the DHS method [29], and the Dimer method (transition with unknown final state) [30]. While all these methods provide an estimate of the transition state, the NEB and CPR method provide a more global view on the energy surface [31]. Both methods allowed the determination of complex re-arrangements in proteins [32–35] as well as proton transfer events [36–38]. Another promising method is the string method [39], in which intrinsically parametrized curves evolve to the most probable transition pathway by following their dynamics, e.g. applied to the *RNA* backbone cleavage by *ribonuclease H* in which proton transfer reactions play a key role [40]. All these methods, however, tend to fail to represent the multitude of co-existing transition mechanisms [22], providing only the MEP whose transition states are closest to the initial guess-path [31]. To achieve a comprehensive description of the transition process nonetheless the state space can be partitioned into different substates, thus translating a complex reaction into a network of simpler transitions [22]. This Transition Network (TN) approach was successfully used to study small atom clusters or glasses [41, 42], peptides [43, 44] and complex protein transitions [45, 46]. Furthermore, we showed recently that TNs are also suitable to identify different proton transfer mechanisms, i.e. concerted/stepwise proton transfer, in a channel-like proton transfer system [47].

The TN approach characterizes the dynamical behavior of molecular systems by transition barriers, and thus, following the transition state theory [48], by transition rates. Other ways to express the same information are transition probabilities, like it is done in Markov State

modeling [49], or mean first passages times [50]. However, all three types of matrices, i.e. rate matrices, transition probability matrices and mean first passage time matrices, can be transformed into each other [50]. Hence, although the TN approach seems to focus on one special minimum energy pathway it includes, in principle, the information for all other paths as well.

A challenging aspect in calculating TNs is the exponential increase of stationary points (and thus MEP calculations) on the energy surface with increasing system size [51]. In the case of proton translocations through proton transfer channels, the number of degrees of freedom (DOFs) to sample, corresponding to critical channel residues and intra-channel water molecules, is already problematic. However, investigations in *CcO* as well as *NADH dehydrogenase* furthermore suggest, that the environment of the proton transfer channel is able to affect intra-channel properties [15, 17, 18]. Hence, thorough investigations of the proton translocation should also include the channel environment. Coupling the TN approach with an MD-based sampling of the positions and orientations of the water molecules inside the channel [47] reduces the TN calculation costs significantly. Still, a direct inclusion of the DOFs of the channel environment in the sampling of proton transfer pathways is infeasible. Therefore, the channel environment needs to be considered indirectly, i.e. several TN calculations need to be performed for varying configurations of the channel environment, to gain a comprehensive understanding of this exceptionally important process.

Here, we present a method, which characterizes TNs for different configurations of the unsampled channel environment by determining important graph theoretical properties, i.e. the minimum spanning tree (MST) and minimax best pathway (MBP), defined in the methods section, using an extensive coarse-graining of the underlying transition networks and the MST of an existing, complete TN calculation as initial guess. Thereby the calculation costs are reduced significantly, while important network properties are maintained, e.g. the maximal barrier of the transition [52]. To validate the novel method we used the same channel-like proton transfer system as in *Reidelbach et al* [47]. Here, however, we introduced a fixed point charge in the vicinity of the channel, representing the unsampled channel environment. Several MSTs or MBPs were determined for various translocations, increases or decreases of the point charge, and compared to the MSTs or MBPs of the respective complete TN calculations. To distinguish the MST or MBP characterization of unknown TNs (as introduced before) from complete TN calculations, the approach presented in this work is termed TN prediction.

Materials and methods

Recently, we investigated the proton transfer through a cylinder filled with thirteen water molecules and one excess proton. Top and bottom of the cylinder were formed by two stationary *t*-butyl structures, while a harmonic potential, setting on at 3.0 Å away from the cylindrical axis with a force constant of 500 kcal/mol/Å², modeled the walls. Attached to the central carbon atom of either *t*-butyl structure was a carboxyl group pointing inside the cylinder. Hence, the model system resembled a water-filled channel in a protein connecting two aspartate or glutamate-like side chains [47]. In the current study we extended our model by introducing a fixed point charge in the vicinity of the cylinder, 6 Å away from the midpoint of the cylindrical axis (cf Supplementary Information for a detailed description of the initial position). Value and position of the point charge were constant within individual TN calculations, corresponding to one TN for each point charge position or value. The initial value of the point charge was set to $q = 0.050$. Other configurations of the point charge were achieved by circular translocations around the initial position with $|\mathbf{r}| = 0.5 \text{ \AA}$, 1.0 Å, or 2.0 Å and $\phi = 0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$, or 315°, parallel to a plane containing the cylindrical axis (cf Fig 1), charge decreases down to $q = 0.000$ (actually $q = 0.000001$ to keep the point charge included in the

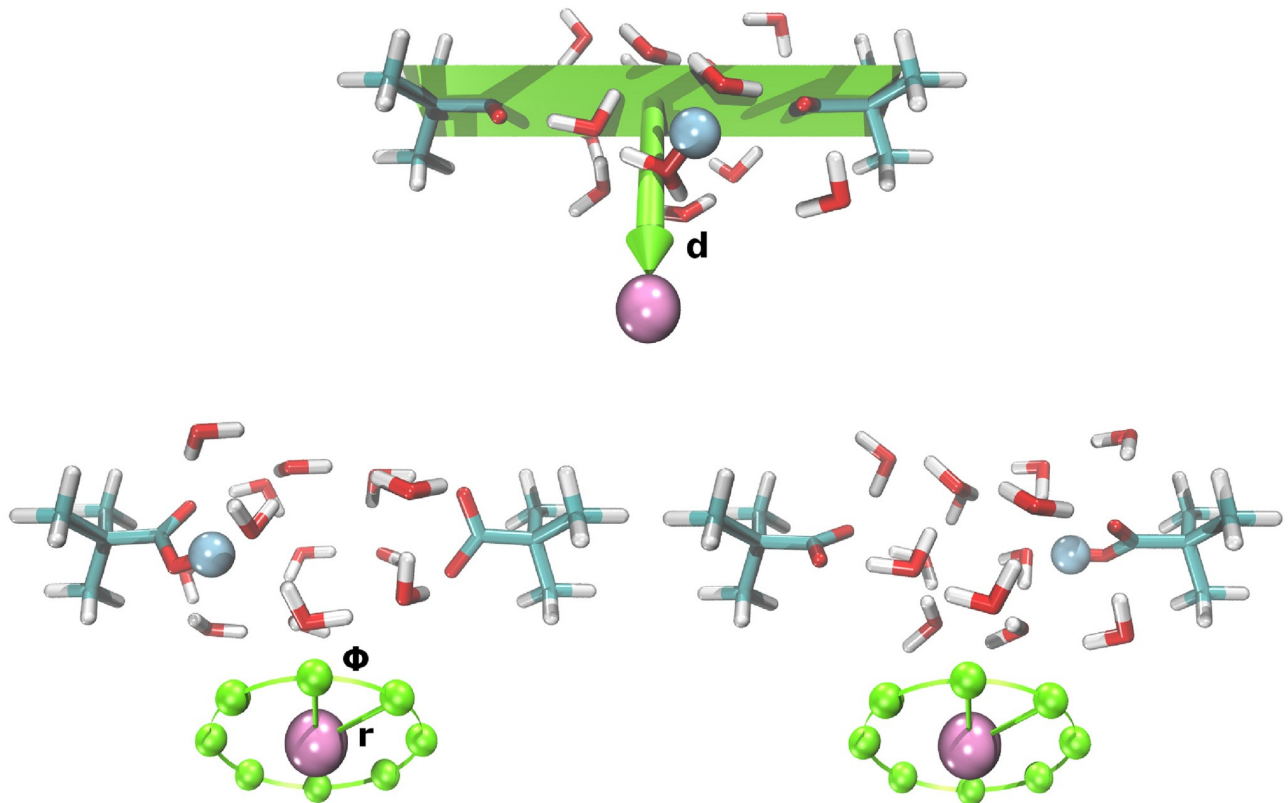


Fig 1. Model system for proton transfer. Model system for calculating proton transfer pathways with an additional point charge (pink sphere). Top: plane containing the cylindrical axis and the orthogonal vector \mathbf{d} ($|\mathbf{d}| = 6.0 \text{ \AA}$) to locate the initial position of the additional 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 proton transfer reaction. Circular translocations of the additional point charge, parallel to the plane with $|\mathbf{r}| = 0.5 \text{ \AA}$, 1.0 \AA , or 2.0 \AA and $\phi = 0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$, or 315° , are indicated by green spheres. Charge increases or decreases are not depicted.

<https://doi.org/10.1371/journal.pone.0207718.g001>

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 $q = 0.100$, giving 51 configurations overall.

In a first step, we performed 51 complete TN calculations, starting from two configurations with the excess proton located at either of the carboxyl groups, i.e. the reactant and product state of the overall proton transfer reaction. To construct the TNs we sampled different protonation states and side chain conformations of the carboxylated t-butyl structures. Therefore, we placed the excess proton on either water molecule or on one of the carboxyl groups and rotated the side chain dihedral angles in 45° steps, giving 1088 initial states. The initial state set was then subjected to a quantum mechanics/molecular mechanics (QM/MM) energy minimization with a convergence criterion of $0.001 \text{ kcal/mol/\AA}$, in which the water-filled cylinder and the additional point charge represented the QM and MM part, respectively. As energy functions we used the semi-empirical OM2 method [53] for the QM part and CHARMM27 [54] for the MM part. Both parts were coupled electrostatically, thus allowed the MM point charge to polarize the QM electron density, while van der Waals interactions between the QM and MM part were modeled by the MM force field. Following the minimization, the optimized states were classified with respect to the overall reactant state, regarding their side chain dihedral angles, protonation state and water pattern, which gave the nodes of the TN, i.e. a set of unique states representing the branching points of the TN. The side chain dihedral angles and

the protonation state were assigned according to the sampling of the initial states, e.g. a 45° rotation of either dihedral angle compared to the overall reactant state or a change in the protonation state compared to the overall reactant state results in a node distinct from the overall reactant state. The motion of the water molecules was not considered in the initial sampling. Still, the pattern of all water molecules in the model channel, defined by their positions and orientations, was included in the state assignment to account for the highly dynamic nature of the water molecules. The minimal difference between two unique water patterns was the translation of a single water molecule by 2 Å in x, y, or z-direction compared to the overall reactant state or the rotation of a single water molecule by 22.5° around the x, y, or z-axis compared to the overall reactant state. To avoid chemically equal nodes, which only differ in their atom labels, e.g. due to a 180° rotation of a deprotonated carboxyl group or due to water molecules switching positions, an excessive atom re-labeling is performed with respect to the overall reactant state using an implementation of the *Hungarian algorithm* [46]. Once the node set of the TN was determined, transitions between selected pairs of nodes, i.e. nodes which differed in each DOF by at most one step, were computed using the CPR method [27], which provided the edges of the TN, i.e. the node connections of the TN. Note that this still allows a transition of a proton along several water molecules simultaneously since the difference in DOFs is evaluated for the proton position to have changed but not by how much. As edge weights we used the energy of the highest transition state in between adjacent nodes relative to the overall reactant state. Nodes and edges were then combined to form a weighted, undirected graph. Finally, the MBP, i.e. the pathway with the lowest highest transition state energy (termed ω^* in this work), connecting the reactant and product state of the overall proton transfer reaction was computed using Dijkstra's algorithm [55].

In a second step, we performed 50 TN predictions using the previously determined complete TN with the point charge in its initial configuration, the initial TN, as starting point for the TN predictions of all other configurations, termed here the perturbed TNs. To characterize the initial TN we determined its MST, i.e. edge subset connecting all nodes without cycles and minimal edge weight sum [56], using Kruskal's [57] algorithm. The MST provides the MBPs between all pairs of nodes and is unique for TNs with distinct edge weights. In principle, each non-MST edge could be part of the MST if its weight is reduced appropriately. Non-MST edges for which a small weight reduction is already sufficient to alter the MST are called high sensitive edges, low sensitive edges are identified accordingly (cf Fig 2, top left). We calculated the sensitivity of all non-MST edges, which allowed us to order the non-MST edges from high to low sensitivity. The MST of the initial TN is later on used as initial guess for the MST or MBP of the perturbed TNs, while the ordered initial non-MST edge lists are used for the refinement of the initial guess.

Once the initial MST and ordered initial non-MST edge list were determined, we coarse-grained the initial TN (initial coarse-graining). In contrast to other coarse-graining techniques, the performed clustering is based on transition barriers, using ω^* as upper bound, instead of a (usual) structure-related clustering [58]. Therefore, we determined the eigenvectors of the *Laplacian matrix*, \mathbf{L} , corresponding to the initial TN,

$$\mathbf{L} = \mathbf{D} - \mathbf{A}^*, \tag{1}$$

where \mathbf{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

$$a_{ij} = \begin{cases} 0 & \text{if } \omega_{ij} > \omega^* \\ 1 & \text{if } \omega_{ij} < \omega^* \end{cases}, \tag{2}$$

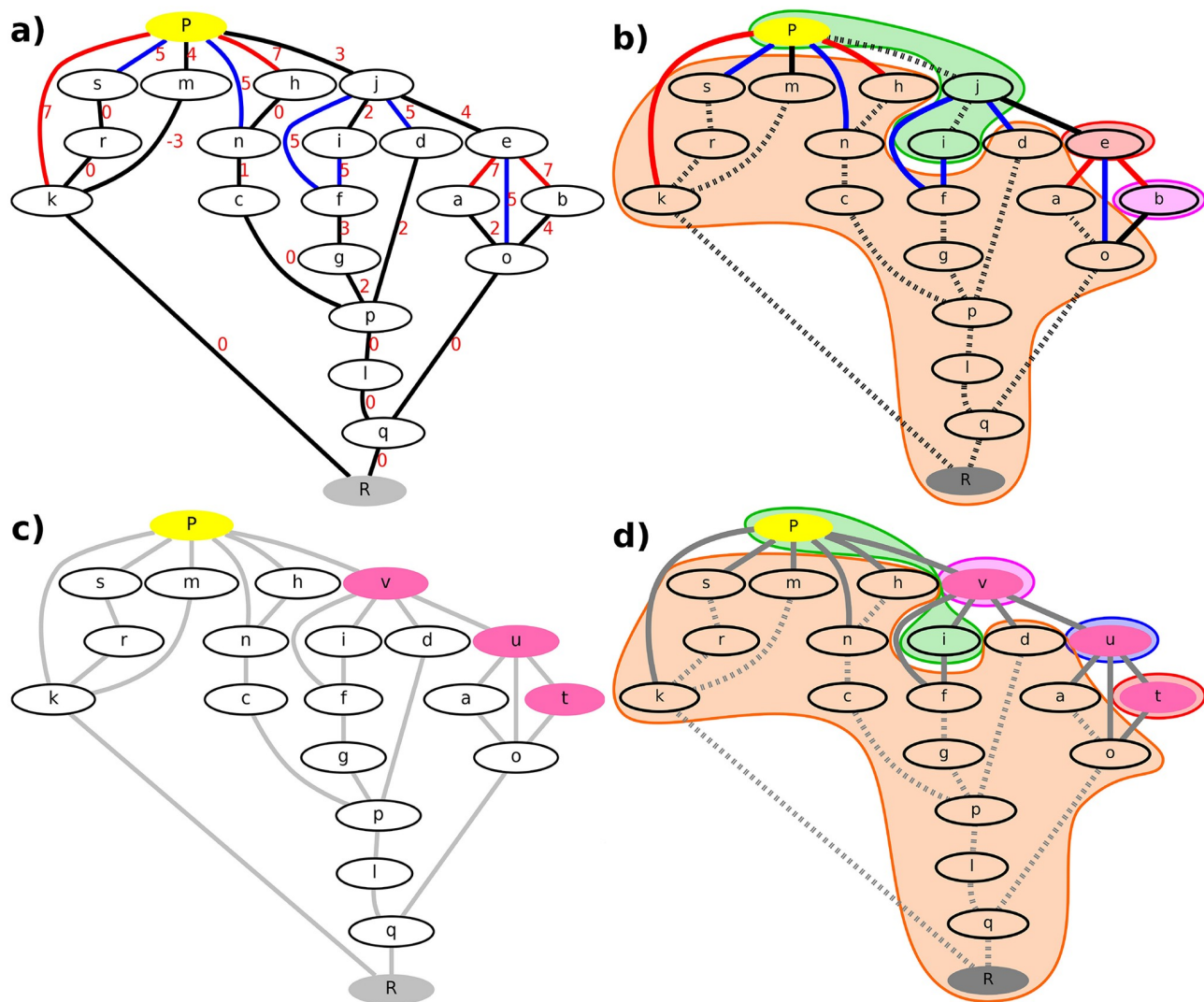


Fig 2. Steps of the TN prediction method. a) Initial TN (combined TN from *Reidelbach et al [47]*) containing 21 nodes and 29 edges. Nodes are shown as ellipses, edges are shown as lines. The reactant state is highlighted in gray, the product state in yellow, labelled by R and P. Intermediate nodes are labelled from a to s. A complete description of either node regarding the side chain dihedral angles of the carboxylated t-butyl structures, the water pattern, and the protonation state can be found in *S1 Table*. Red numbers represent the energy of the maximal transition state along the edge used as edge weight. All energies are in kcal/mol, relative to the overall reactant energy 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. 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.

<https://doi.org/10.1371/journal.pone.0207718.g002>

to determine the connected components and thus the coarse-grained nodes. By that, the initial MBP is for example reduced to a path with two coarse-grained nodes (each containing at least one original node) connected by an edge with ω^* as weight (cf Fig 2, top right). These coarse-grained nodes, however, cannot be understood as single representative conformations, e.g. average conformations of original nodes or original nodes most similar to the average conformations, because these representative conformations are in most cases no local minima

conformations, which is required by the CPR algorithm, or they suffer from the initial guess-path problem. Instead, they need to be interpreted as sets of original nodes between which barrier-free transitions (compared to ω^*) are possible. Hence, the coarse-graining step is not actually reducing the number of nodes to look at, but the number of edges, i.e. all edges connecting nodes within the same coarse-grained node can be excluded from further investigations. Thereby, we reduced the sets of edges representing the initial MST and ordered initial non-MST edge list to those edges connecting distinct coarse-grained nodes only.

Following the analysis of the initial TN, we redid the sampling of the initial states, the minimizations and the node pairing in the perturbed systems, according to the same criteria of changes in DOFs between pairs of nodes as in the initial TN. Thereby, we calculated all nodes of the perturbed node sets and identified all edges to be calculated for the perturbed edge sets, thus receiving a comprehensive picture of the topology of the perturbed TNs. Furthermore, we assumed that the initial coarse-graining is still valid, thus disregarding all edges connecting nodes within the same coarse-grained node in the perturbed edge sets (cf Fig 2, bottom). The coarse-grained initial MST and coarse-grained ordered initial non-MST edge list were then adjusted to the perturbed TNs: All edges present in the initial edge set but absent in the perturbed edge sets were removed from the edge sets representing the coarse-grained initial MST and coarse-grained ordered initial non-MST edge list, on the other hand, edges present in the perturbed edge sets but absent in the initial edge set were added to the edge set representing the coarse-grained ordered initial non-MST edge list. To integrate these edges properly into the sensitivity ranking 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. The sensitivity value for edges connecting nodes not present in the initial node set was set to zero. If several edges were assigned the same sensitivity value, they were ordered randomly. Hence, the coarse-grained initial non-MST edge list is ordered from high to low sensitivity values with a non-deterministic order for edges with equal sensitivity values.

The remaining coarse-grained initial MST edges were then calculated in the perturbed TNs. In few cases these calculations were already sufficient to obtain the coarse-grained MST or coarse-grained MBP of the perturbed TNs. If that was not the case we calculated additional non-MST edges from the coarse-grained ordered initial non-MST edge lists in the order from high to low sensitivity values. The calculation of additional edges allowed a further coarse-graining of the perturbed TNs (“on-the-fly” coarse-graining), i.e. merging coarse-grained nodes if connected by an additional edge with edge weight below ω^* , thus reducing the coarse-grained ordered initial non-MST edge lists even further along the edge calculation. The calculations 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 3 the TN prediction method is summarized.

In a third and final step, we tested our TN prediction method for an extended set of perturbations using each of the previously determined complete TNs as initial TN for the prediction of the MST or MBP of all other TNs, yielding 702 and 600 TN predictions for the increase or decrease of the additional point charge or the translocation around its initial position, respectively.

All energy minimizations and CPR calculations were performed using the CHARMM programme [59] interfaced to MNDO [60]. The generation of the initial states, the node assignment, the neighbor search as well as the compilation and analysis of the TNs were performed with our own java code and java libraries from *Noe et al* [45]. 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 our own python code.

Results and discussion

Complete transition network calculations

In a first step 51 complete TNs were calculated. These were the initial TN, 26 TNs with an increased or decreased value of the additional point charge and 24 TNs with charge translocations around the initial position. Here, only the results for the initial TN are presented in detail (a detailed description of all 51 TN calculations is given in [S2 Table](#) and [S1 to S9 Figs](#)). The initial TN calculation resulted in a network with 252 nodes connected by 20316 edges. The MBP connecting the reactant and product state of the overall proton transfer reaction contains three intermediate nodes and a maximal transition barrier of 5 kcal/mol. This pathway involves the re-arrangement of both side chain dihedral angles, the protonation state and the water pattern. Interestingly, the change in the side chain dihedral angle of the second carboxylated t-butyl structure and protonation state is not gradually. Instead, back transitions to the overall reactant state or previous intermediate node are observed, e.g. a counterclockwise rotation of $\approx 45^\circ$ (transition from 0 to 7) followed by a clockwise rotation of $\approx 90^\circ$ in two steps (transitions from 7 to 0 and 0 to 1) for the side chain dihedral angle of the second carboxylated t-butyl structure (cf. [Fig 4](#), indigo pathway). The next best proton transfer pathways have transition barriers of 8 and 10 kcal/mol, respectively. The direct TN calculation without an external point charge [47] gives the same maximal transition barrier for the MBP. However, the intermediate states involved differ, simply due to the fact that the conformation of the reactant and product states differ. Nevertheless, structural elements, in terms of visited protonation sites, of the second best proton transfer pathway resemble the pathway previously reported.

The increase, decrease and translocation of the external point charge affects the MBP properties. For point charge increases or decreases for example the maximal transition barrier of the MBP varies between 3 and 9 kcal/mol with an average maximal transition barrier of the MBPs of 5 ± 1.5 kcal/mol (5 ± 1.5 kcal/mol for charge translocations), while the number of intermediate nodes varies between 1 and 7 nodes with an average number of intermediate nodes of 5 ± 1 (5 ± 1 for charge translocations). Furthermore, the general network topology of either TN is affected to a large extent. In case of charge translocations for example 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 following the charge translocations (41 to 60% of the nodes and 58 to 80% of the edges for point charge increases or decreases). On the other hand, 30 to 70% of the nodes and 50 to 87% of the edges present in perturbed TNs do not exist in the initial TN (38 to 68% of the nodes and 56 to 82% of the edges for point charge increases or decreases). Hence, in most cases the perturbed TNs contain more “unknown” topological features than “known”. Thus, 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. Instead, further information about the perturbed TNs need to be acquired while calculating them. In our TN prediction method this fact is acknowledged by the “on-the-fly” coarse-graining step.

Transition network predictions

The TN prediction method proposed in this paper contains three fundamental steps. 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.

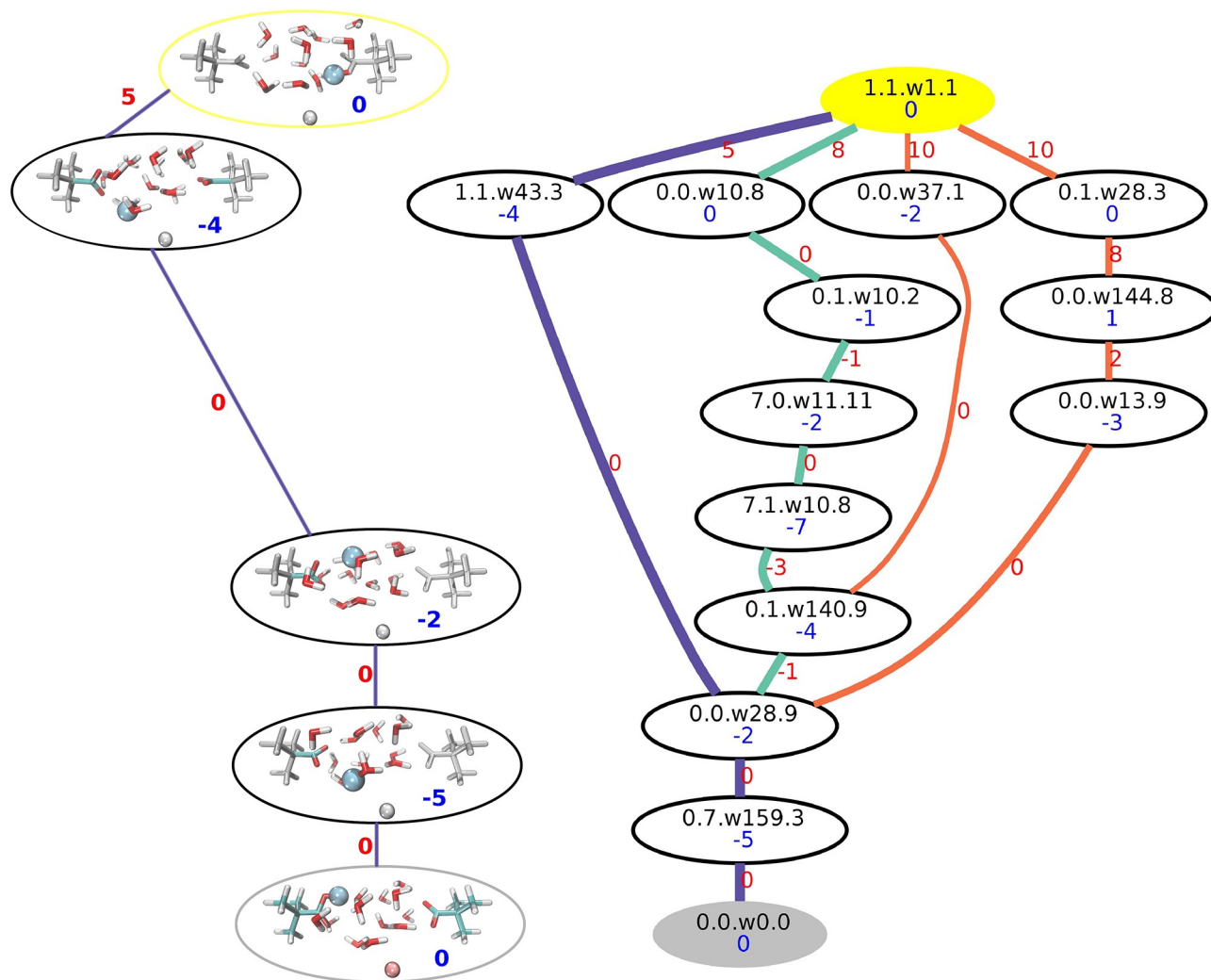


Fig 4. Initial TN. Right: Initial TN of a 13 water model system with additional point charge from a complete TN calculation. The nodes are shown as ellipses, labelled according to the side chain dihedral angles of the carboxylated t-butyl structures, the water pattern and the protonation state (SC0.SC1.Wi.P). The reactant state is shown in gray, the product state in yellow. Edges are shown as lines. Blue numbers represent the energy of the nodes, red numbers the energy of the maximal transition state along the edge used as edge weight. All energies are in kcal/mol, relative to the overall reactant energy and rounded to integer values. The MBP is highlighted in indigo, the next best paths are shown in turquoise and orange. Left: Detailed representation of the best pathway. The reactant state is shown in gray, the product state in yellow. Edges are shown as lines. Blue numbers represent the energy of the nodes, red numbers the energy of the maximal transition state along the edge used as edge weight. All energies are in kcal/mol, relative to the overall reactant energy and rounded to integer values. Colored structure elements indicate structural changes along the transition according to the assignment of the TN calculation.

<https://doi.org/10.1371/journal.pone.0207718.g004>

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

To demonstrate the cost reductive effect of either step Fig 5 displays the costs of the edge calculations, i.e. the number of essential edge calculations compared to the size of the perturbed edge sets in %, for an increase or decrease of the point charge or its translocation around the initial position, when only 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. All calculations were performed 1000 times per perturbation scenario to study the effect of the non-deterministically ordered edges (due to equal sensitivity values)

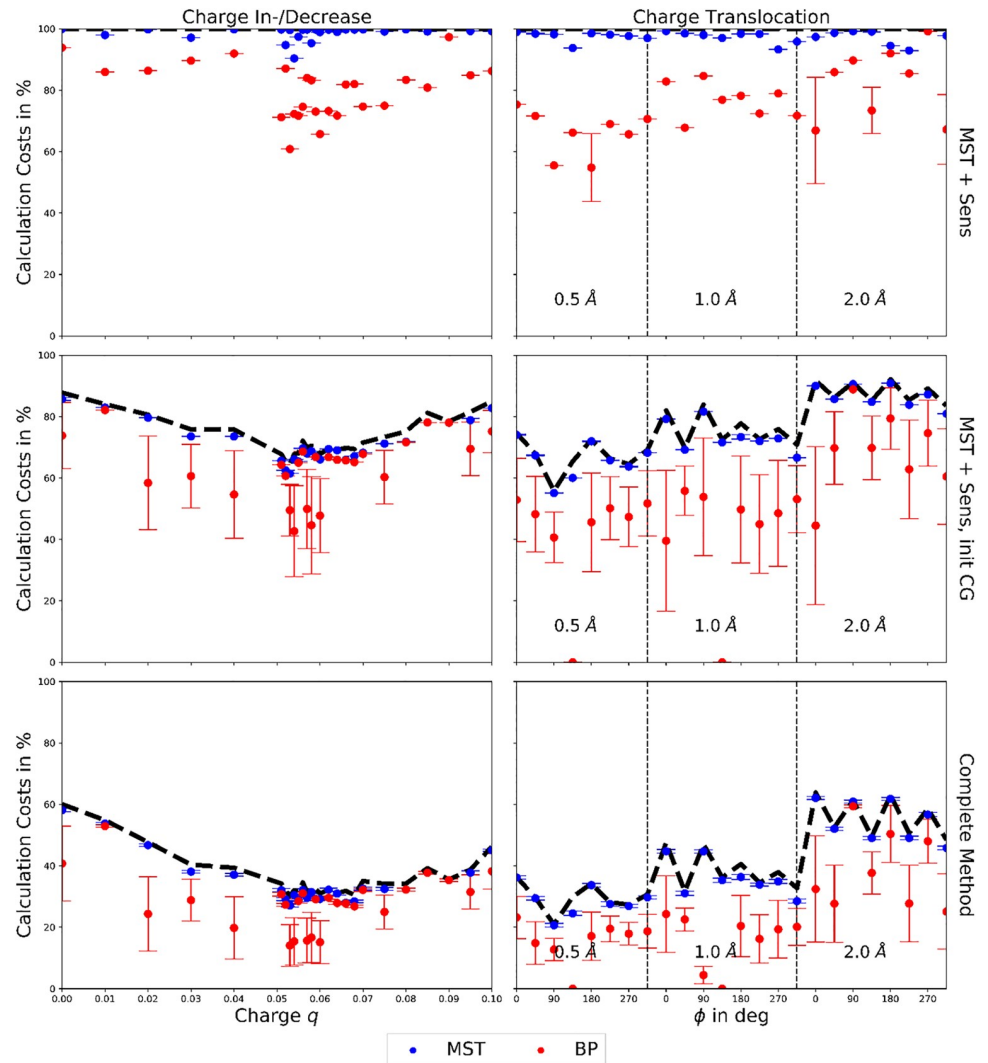


Fig 5. Average calculation cost of MST and MBP predictions. Average calculation costs (in % of the average length of the ordered non-MST edge lists without coarse-graining) for the (all-node or coarse-grained) prediction of the perturbed MST (blue) and MBP (red) from 1000 prediction runs per perturbation. Perturbations: charge in-/decrease (left column), charge translocation (right column). Top row: calculation by MST and sensitivity (all-node results), middle row: calculation by MST, sensitivity and 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.

<https://doi.org/10.1371/journal.pone.0207718.g005>

within the all-node or coarse-grained ordered non-MST edge lists and 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.

The determination of the perturbed MSTs required average calculations of 99%, 71%, or 35% of the non-MST edge lists for charge increases or decreases and 97%, 75%, or 40% for charge translocations, 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. The calculation costs for both types of perturbation do not depend on the severity of the perturbation, when performing calculations without coarse-graining.

However, when applying the initial coarse-graining or the initial and “on-the-fly” coarse-graining the calculation costs depend on the severity of the perturbation, e.g. costs for translocations of $0.5 \text{ \AA} < \text{costs for translocations of } 1.0 \text{ \AA} < \text{costs for translocations of } 2.0 \text{ \AA}$. Interestingly, the calculation costs are not symmetric for charge increases and decreases, revealing a calculation cost difference of 12% for $\Delta q = \pm 0.050$ when using the initial and “on-the-fly” coarse-graining. The prediction of the perturbed MSTs for charge increases and decreases or charge translocations around the initial position is based on the MST of the unperturbed TN, used as initial guess, giving rise to a possible bias of the predicted, perturbed MSTs towards the initial guess. However, a comparison of the initial guess MST with the predicted, perturbed MSTs reveals a maximal edge similarity of the MSTs of only 10% and 20% (on average $5 \pm 3\%$ and $10 \pm 6\%$) for charge increases or decreases and charge translocations around the initial position, respectively. On the other hand, the node similarity of the MSTs equals, obviously, the node similarity of the TNs stated before. Hence, the initial guess MST and predicted, perturbed MSTs are sufficiently different to indicate a bias-free 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 determination of the perturbed MBPs required average calculations of 80%, 64%, or 28% of the non-MST edge lists for charge increases or decreases and 75%, 51%, or 23% for charge translocations, 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. The calculation costs for the MBPs depend on the severity of the perturbation. For charge translocations of $|\mathbf{r}| = 0.5 \text{ \AA}$ or 1.0 \AA and $\phi = 135^\circ$ the calculation of the initial MST edges was already sufficient to determine the perturbed coarse-grained MBPs. In all cases the calculation of the perturbed MBPs required less edge calculations than the calculation 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 determination by coarse-graining, and thus reducing, the ordered non-MST edge lists (cf Fig 5 dashed black lines). A complete calculation of the coarse-grained non-MST edge lists guarantees the most accurate determination of the perturbed MSTs or MBPs (at least within the error related to the coarse-graining, discussed later on) and cost reductions of 40 up to 80% (compared to the edge calculation costs of the respective complete TN calculations). In principle, further cost reductions are possible, at least if one is only interested in the coarse-grained perturbed MBPs (cf Fig 5 red dots and previous paragraph). However, defining a lower edge calculation bound providing accurate coarse-grained MBPs is problematic due to high fluctuations (up to $\pm 17\%$) associated with the non-deterministic order of edges with equal sensitivity values within the coarse-grained ordered non-MST edge lists used for the predictions. Therefore, 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-set 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.

A single reaction pathway is often not enough to properly describe a chemical reaction [22], e.g. the transfer of protons. Therefore, we determined the perturbed second, third, fourth, and fifth best MBPs, next to the actual MBPs, with our TN prediction method. The results for charge increases or decreases and charge translocations around the initial position, using the MST and its sensitivity, the initial coarse-graining, and the “on-the-fly” coarse-graining, are presented in Fig 6.

The average calculation costs for the prediction of the perturbed second, third, fourth, or fifth best MBPs fluctuate, in most cases, around the calculation costs for the actual perturbed

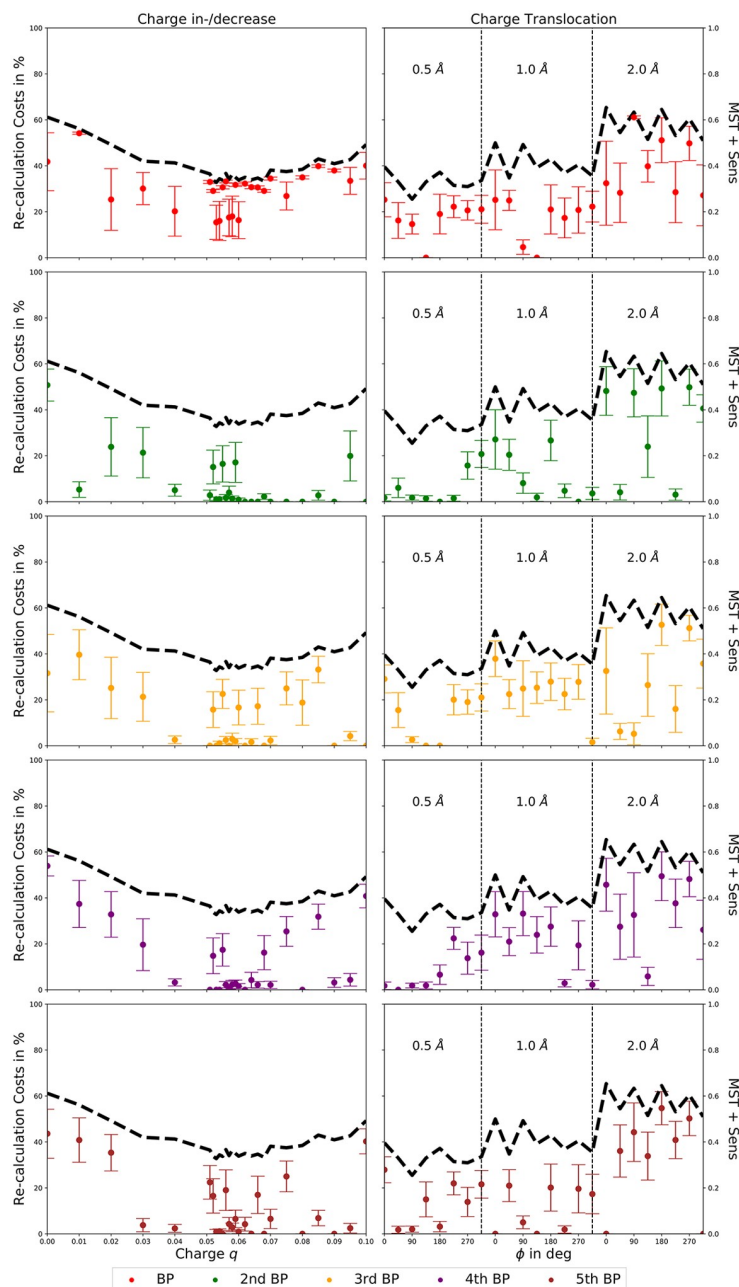


Fig 6. Average calculation cost of next best pathway predictions. Average calculation costs (in % of the average length of the ordered non-MST edge lists without coarse-graining) for the coarse-grained prediction of the perturbed MBP (red), 2nd MBP (green), 3rd MBP (orange), 4th MBP (purple), and 5th MBP (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: charge in-/decrease (left column), charge translocation (right column). The average length (as before in % of the average length of the ordered non-MST edge lists without coarse-graining) of the coarse-grained, ordered non-MST edge lists used for the calculations is depicted in black, as dashed line for a better visualization.

<https://doi.org/10.1371/journal.pone.0207718.g006>

MBPs. Thus, a complete calculation of the coarse-grained non-MST edge lists (dashed black lines), guarantees not only the most accurate determination of the MBPs, but also the most accurate determination of the second, third, fourth, and fifth best MBPs. Thereby, the TN prediction method is able to provide a proper description of chemical reactions in a cost efficient manner.

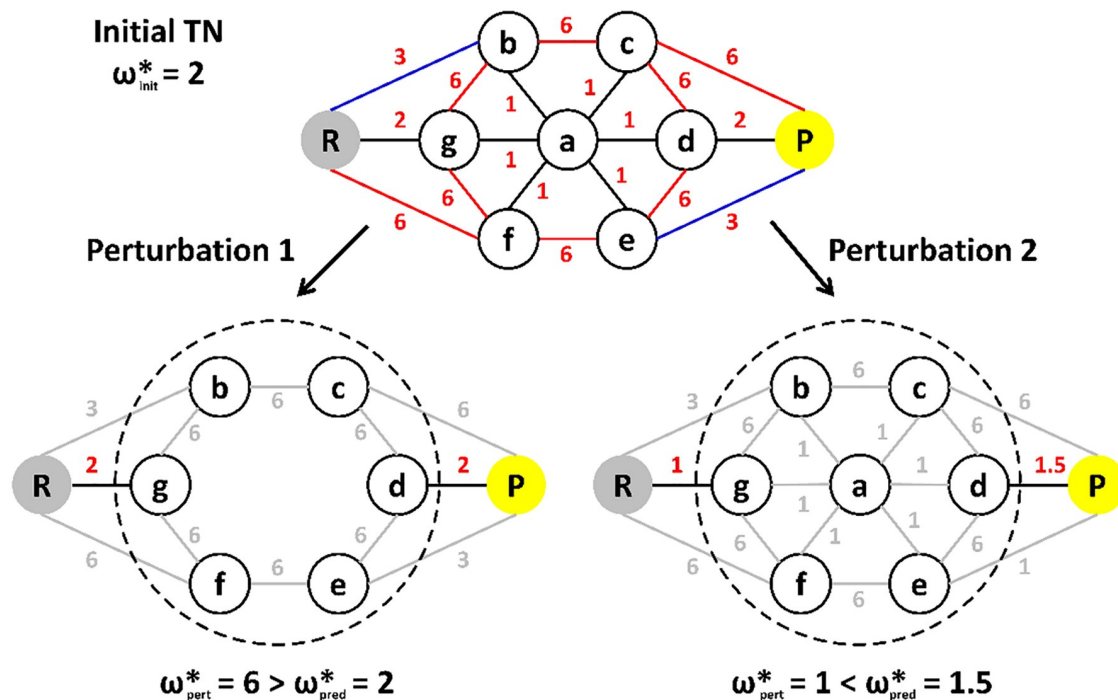


Fig 7. Coarse-graining problems. Initial TN: Nodes are depicted as circles, edges as lines. Reactant and product state are shown in gray and yellow labelled R and P, respectively, other nodes are labelled from a to g. The edge weights are in kcal/mol and shown in red. Perturbed TNs: Node labeling as 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.

<https://doi.org/10.1371/journal.pone.0207718.g007>

Coarse-graining problems

In all perturbation scenarios significant cost reductions were achieved by the initial and “on-the-fly” coarse-graining steps. These steps, however, are also potential sources of error regarding the determined maximal transition barriers of the perturbed MBPs or properties related to the MSTs. Here, we focus on the accuracy of the maximal transition barriers of the perturbed MBPs.

The initial coarse-graining is based on the assumption that the integrity of the coarse-grained nodes is preserved beyond the perturbation, 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 a coarse-grained node can be reached by crossing barriers below ω_{init}^* only. In Fig 7 “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 assumes that all nodes can be reached by crossing barriers below 2 kcal/mol. Therefore, the maximal transition barrier of the perturbed MBP would be 6 kcal/mol if all nodes are considered and 2 kcal/mol if the initial coarse-graining is applied. Hence, the initial coarse-graining step is prone to 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 ω_{init}^* , thereby neglecting all further edges connecting the two coarse-grained nodes. This setup ensures an exact refinement of ω_{pert}^* if $\omega_{pert}^* \geq \omega_{init}^*$. In the

opposite case, however, the “on-the-fly” coarse-graining could stop the refinement of ω_{pert}^* too early. A typical situation is displayed in Fig 7 by “Perturbation 2”. The initial coarse-graining provided three coarse-grained nodes. Following the calculation of the coarse-grained initial MST edges in the perturbed system all coarse-grained nodes will be combined by the “on-the-fly” coarse-graining, preventing further edge calculations. Therefore, the 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 maximal transition barrier overestimations for situations in which $\omega_{pert}^* < \omega_{init}^*$, while underestimations are ruled out.

To check the accuracy of our predictions we compared the maximal transition barriers of the MBPs of the complete TN calculations (ω_{pert}^*) with those accessible after the initial coarse-graining and those finally predicted (ω_{pred}^*). The results are summarized in Fig 8.

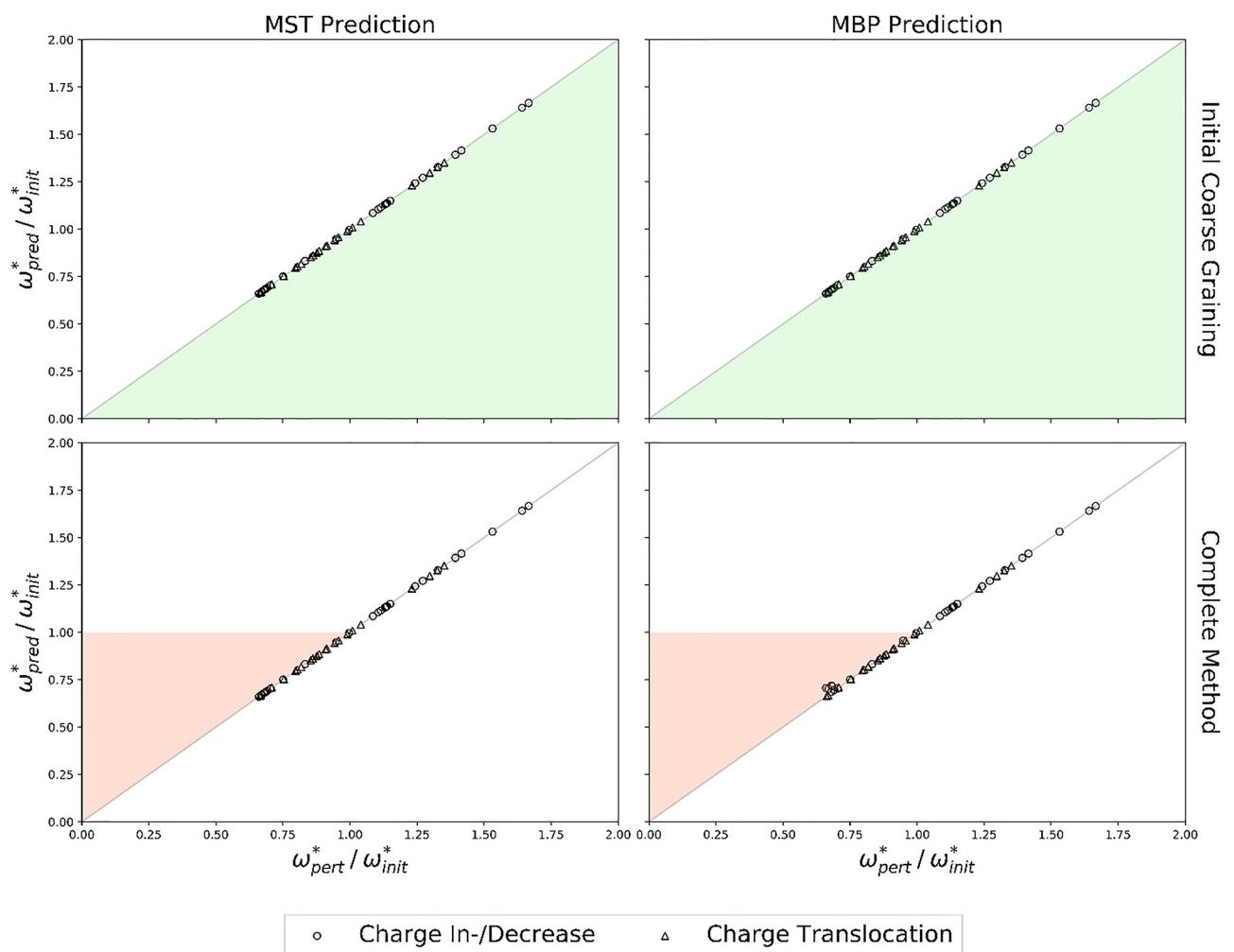


Fig 8. Accuracy of ω_{pert}^* predictions. Comparison of ω^* from complete TN calculations (ω_{pert}^*) and MST (left column) or MBP (right column) predictions (ω_{pred}^* , averaged over 1000 prediction runs per perturbation) following the initial coarse-graining (top row) or the complete method (bottom row). Charge in-/decreases (circles), charge translocations (triangles). Shaded areas indicate potential error regions, underestimations (green), overestimations (red).

<https://doi.org/10.1371/journal.pone.0207718.g008>

The maximal transition barriers of the MBPs accessible after the initial coarse-graining are in perfect agreement to the barriers determined by the complete TN calculations, regardless of the type or severity of the perturbation or the ratio $\omega_{pert}^*/\omega_{init}^*$. Hence, the potential energy surface associated with the system is most likely stabilizing the coarse-grained nodes, rendering perturbations as displayed in Fig 7 by “Perturbation 1” unlikely. The maximal transition barriers of the MBPs derived by the complete method also agree with the barriers determined by the complete TN calculations. However, for situations in which $\omega_{pert}^*/\omega_{init}^* < 1$, i.e.

$\omega_{pert}^* < \omega_{init}^*$, 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-empirical quantum method OM2 [61]. Hence, the inaccuracies of the TN prediction method are negligible compared to the intrinsic error of the semi-empirical quantum method. Nevertheless, the potential risk of maximal transition barrier over- or underestimations should not be forgotten.

Influence of the initial transition network

Finally, we tested our TN prediction method for a larger set of perturbations by using each of the complete TNs as initial TN for the prediction of the MST or MBP of all other TNs, thus increasing the amount of TN predictions from 26 and 24 to 702 and 600 for the in-/decrease of the point charge or its translocation around the initial position, respectively. Thereby, we were able to investigate increased perturbation severities, i.e. Δq up to ± 0.100 (instead of ± 0.050) and $|\mathbf{r}|$ up to 4.0 Å (instead of 2.0 Å), and the influence of the initial TN (and the similarity of the initial and perturbed TN) on the prediction of the MST or MBP of the perturbed TNs.

For the correct prediction of the perturbed MSTs and MBPs different amounts of edges need to be calculated on average. In case of the MSTs not a single correct prediction was observed, when calculating less than $\approx 20\%$ of the perturbed edge sets. On the other hand, calculations of only 5% of the perturbed edge sets provided correctly predicted perturbed MBPs in 16% (charge in-/decrease) or 12% (charge translocation) of all TN predictions. However, once calculations of more than 20% of the perturbed edge sets are performed the increase in correctly predicted perturbed MSTs is higher than the increase in correctly predicted perturbed MBPs, giving correct predictions in 82% and 90% (charge in-/decrease) or 50% and 71% (charge translocation) of all TN predictions for the perturbed MSTs and perturbed MBPs with a calculation of 50% of the perturbed edge sets. With calculations of 60% of the perturbed edge sets the number of correct predictions is above 90% for all cases, except the perturbed MST prediction for charge translocations.

The dashed black lines in Fig 9 indicate the number of correct MST and MBP predictions by calculating the complete coarse-grained non-MST edge lists and thus the length of the coarse-grained ordered non-MST edge lists. In principle, 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 calculations 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 our TN prediction method, for charge in-/decreases or charge translocations, respectively. Hence, in 77% or 72% of all TN predictions the predicted MSTs or MBPs are unequivocally correct (within the coarse-graining errors described before) when only calculating 50% or 60% of the perturbed edge sets, because there are no further edges left to calculate. As already described before, further edge calculation reductions are possible (cf Fig 9 blue and red lines), but inevitably bear the risk of incorrect

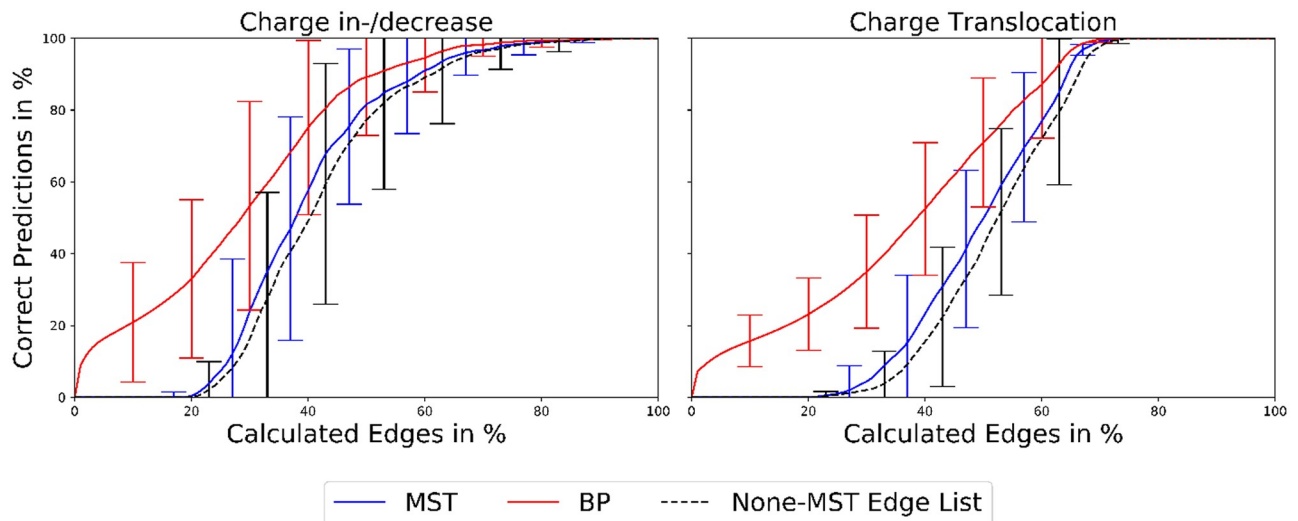


Fig 9. Correct MST and MBP predictions. Correct predictions of the perturbed MSTs (blue) or MBPs (red) for individual numbers of calculated edges averaged over the different initial TNs. The dashed black lines indicate the number of correct MST and MBP predictions by calculating the complete coarse-grained ordered non-MST edge lists averaged over the different initial TNs. Left: charge in-/decrease, right: charge translocation.

<https://doi.org/10.1371/journal.pone.0207718.g009>

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 determination of the MSTs or MBPs or the complete 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 9 standard deviations).

Once again we checked the accuracy of our TN predictions by comparing the maximal transition barriers of the MBPs from the complete TN calculations and respective TN predictions. The results are summarized in Fig 10.

As before (cf. Fig 8) not a single maximal transition barrier underestimation was observed following the initial coarse-graining of the perturbed TNs, supporting the hypothesis that the potential energy surface of the particular model studied here is preventing such deviations. The same holds for the maximal transition barriers derived by the complete TN prediction method if $\omega_{pert}^*/\omega_{init}^* > 1$. For the opposite case, however, deviations from the perturbed maximal transition barriers were observed in twelve of all 1302 initial and perturbed TN combinations. Five out of twelve 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 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 overestimations. On the other hand, eleven out of twelve overestimations occurred in TN predictions in which the perturbed TNs displayed the lowest maximal transition barrier of 3 kcal/mol (4 kcal/mol for the remaining one), once again providing the maximal range for overestimations. Hence, further investigations might be necessary for situations in which the perturbed 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 maximal transition barriers.

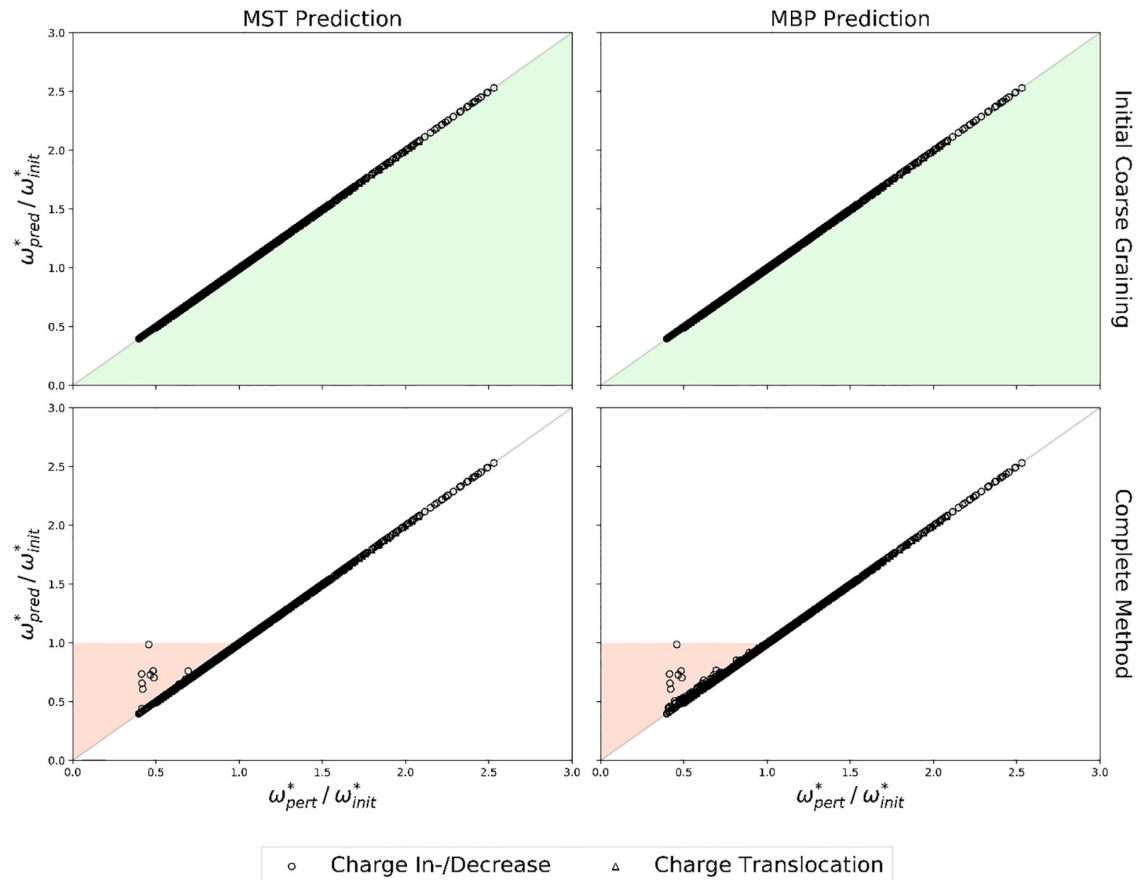


Fig 10. Accuracy of ω^* predictions. Comparison of ω^* from complete TN calculations (ω_{pert}^*) and MST (left column) or MBP (right column) predictions (ω_{pred}^*), averaged over 1000 prediction runs per perturbation) following the initial coarse-graining (top row) or the complete method (bottom row). Charge in-/decreases (circles), charge translocations (triangles). Shaded areas indicate potential error regions, underestimations (green), overestimations (red).

<https://doi.org/10.1371/journal.pone.0207718.g010>

Conclusion

The TN prediction method proposed in this paper (summarized in Fig 3) characterizes perturbed TNs by determining their MSTs or MBPs on a coarse-grained level using the MST of an existing, complete TN as initial guess. The costs for a TN prediction are flexible, self-regulated by the TN prediction method and constantly adjusted to the requirements of the perturbed TNs. Thereby significant cost reductions of up to 80% (compared to complete TN calculations) were achieved in a small model system resembling a water filled proton transfer channel for various perturbations of a point charge in the vicinity of the channel. The accuracy of the TN prediction method was tested for the determined maximal transition barriers of the perturbed MBPs, showing for the most part only inaccuracies which were well below the intrinsic error of the semi-empiric calculation method. In few cases more pronounced deviations were observed, rendering further investigations of the perturbed TNs a necessity if the perturbed maximal transition barrier is lower than the initial one.

In principle the TN prediction method proposed in this paper is an extension of Boruvka’s algorithm [62] used for the determination of MSTs. Here, the MST is determined iteratively by coarse-graining the TN until it contains a single node only. Therefore, in every iteration cycle the edges with minimum weight incident to each node are determined and added to the

MST, while the TN is coarse-grained along the edges with minimum weight and self loops and multiple edges between pairs of nodes are eliminated (except for the edge with minimum weight). The TN prediction method employs the same concept to determine perturbed MSTs and MBPs when performing the “on-the-fly” coarse-graining. However, it is not possible to use edges with minimum weight incident to each node only, since the weights of all edges are not known a priori. Furthermore, it is not possible to eliminate all multiple edges between pairs of coarse-grained nodes due to peculiarities of the MEP calculation. To compensate for all these drawbacks, an excessive pre-processing, i.e. sensitivity analysis and initial coarse-graining, is performed in a different (but overall similar) TN. Thereby, it is possible to accurately predict the MST and MBP of a perturbed TN, when only calculating a fraction of the perturbed edge set.

Supporting information

S1 Fig. Perturbed TNs with decreased point charge. $q = 0.000$ to $q = 0.040$.
(PDF)

S2 Fig. Perturbed TNs with increased point charge. $q = 0.051$ to $q = 0.055$.
(PDF)

S3 Fig. Perturbed TNs with increased point charge. $q = 0.056$ to $q = 0.060$.
(PDF)

S4 Fig. Perturbed TNs with increased point charge. $q = 0.060$ to $q = 0.070$.
(PDF)

S5 Fig. Perturbed TNs with increased point charge. $q = 0.075$ to $q = 0.100$.
(PDF)

S6 Fig. Perturbed TNs with translocated point charge. $|\mathbf{r}| = 0.5$ with $\phi = 0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ$, and 225° .
(PDF)

S7 Fig. Perturbed TNs with translocated point charge. $|\mathbf{r}| = 0.5$ with $\phi = 270^\circ$, and 315° and $|\mathbf{r}| = 1.0$ with $0^\circ, 45^\circ, 90^\circ$, and 135° .
(PDF)

S8 Fig. Perturbed TNs with translocated point charge. $|\mathbf{r}| = 1.0$ with $180^\circ, 225^\circ, 270^\circ$, and 315° and $|\mathbf{r}| = 2.0$ with 0° and 45° .
(PDF)

S9 Fig. Perturbed TNs with translocated point charge. $|\mathbf{r}| = 2.0$ with $90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$, and 315° .
(PDF)

S1 Table. Node label comparison.
(PDF)

S2 Table. TN parameters.
(PDF)

Acknowledgments

We thank the HPC support of the physics department of the Freie Universität Berlin and Frank Noe for making the sources of his Transition Networks implementation available to us.

Author Contributions

Conceptualization: Marco Reidelbach.

Data curation: Marco Reidelbach.

Formal analysis: Marco Reidelbach.

Funding acquisition: Petra Imhof.

Investigation: Marco Reidelbach.

Methodology: Marcus Weber.

Supervision: Marcus Weber, Petra Imhof.

Validation: Marco Reidelbach.

Visualization: Marco Reidelbach.

Writing – original draft: Marco Reidelbach.

Writing – review & editing: Marcus Weber, Petra Imhof.

References

1. Lane N. Proton gradients at the origin of life. *Bioessays*. 2017. <https://doi.org/10.1002/bies.201600217>. PMID: 28503790
2. Deamer DW, Nichols JW. Proton flux mechanisms in model and biological membranes. *The Journal of Membrane Biology*. 1989; 107:91:103.
3. Deamer DW, Nichols JW. Net proton-hydroxyl permeability of large unilamellar liposomes measured by an acid-base titration technique. *PNAS*. 1980; 77:2038–2042. <https://doi.org/10.1073/pnas.77.4.2038> PMID: 6246519
4. Decoursey TE, Cherny VV. Deuterium isotope effects on permeation and gating of proton channels in rat alveolar epithelium. *The Journal of General Physiology*. 1997; 109:415–434. <https://doi.org/10.1085/jgp.109.4.415> PMID: 9101402
5. Finkelstein A, Andersen OS. The gramicidin A channel: a review of its permeability characteristics with special reference to the single-file aspect of transport. *The Journal of Membrane Biology*. 1981; 59:155–171. <https://doi.org/10.1007/BF01875422> PMID: 6165825
6. Cukierman S. Proton mobilities in water and in different stereoisomers of covalently linked gramicidin A channels. *Biophysical Journal*. 2000; 78:1825–1834. [https://doi.org/10.1016/S0006-3495\(00\)76732-4](https://doi.org/10.1016/S0006-3495(00)76732-4) PMID: 10733963
7. Levitt DG, Elias SR, Hautman JM. Number of water molecules coupled to the transport of sodium, potassium and hydrogen ions via gramicidin, nonactin or valinomycin. *Biochim Biophys Acta*. 1978; 512:436–451. [https://doi.org/10.1016/0005-2736\(78\)90266-3](https://doi.org/10.1016/0005-2736(78)90266-3) PMID: 81687
8. Grotthuss CJTV. Sur la decomposition de l'eau et des corps qu'elle tient en dissolution a l'aide de l'electricite galvanique. *Annali Di Chimica*. 1806; 58:54–73.
9. Agre P, Preston GM, Smith BL, Jung JS, Raina S, Moon C, et al. Aquaporin CHIP: the archetypal molecular water channel. *American Journal of Physiology-Renal Physiology*. 1993; 265:463–476.
10. Zeidel ML, Ambudkar SV, Smith BL, Agre P. Reconstitution of functional water channels in liposomes containing purified red cell CHIP28 protein. *Biochemistry*. 1992; 31:7436–7440. <https://doi.org/10.1021/bi00148a002> PMID: 1510932
11. Kozono D, Yasui M, King LS, Agre P. Aquaporin water channels: atomic structure molecular dynamics meet clinical medicine. *The Journal of Clinical Investigation*. 2002; 109:1395–1399. <https://doi.org/10.1172/JCI15851> PMID: 12045251
12. Holsinger LJ, Nichani D, Pinto LH, Lamb RA. Influenza A virus M2 ion channel protein: a structure-function analysis. *Journal of Virology*. 1994; 68:1551–1563. PMID: 7508997
13. Sansom MSP, Kerr ID, Smith GR, Son HS. The influenza A virus M2 channel: a molecular modeling and simulation study. *Virology*. 1997; 233:163–177. <https://doi.org/10.1006/viro.1997.8578> PMID: 9201226

14. Henry RM, Yu CH, Rodinger T, Pomés R. Functional Hydration and Conformational Gating of Proton Uptake in Cytochrome c Oxidase. *Journal of Molecular Biology*. 2009; 387:1165–1185. <https://doi.org/10.1016/j.jmb.2009.02.042> PMID: 19248790
15. Bagherpoor Helabad M, Ghane T, Reidelbach M, Woelke AL, Knapp W, Imhof P. Protonation-State-Dependent Communication in Cytochrome c Oxidase. *Biophysical Journal*. 2017; 113:817–828. <https://doi.org/10.1016/j.bpj.2017.07.005> PMID: 28834718
16. Sharma V, Jambrina PG, Kaukonen M, Rosta E, Rich PR. Insights into functions of the H channel of cytochrome c oxidase from atomistic molecular dynamics simulations. *PNAS*. 2017. <https://doi.org/10.1073/pnas.1708628114>.
17. Ghane T, Gorriz RF, Wrzalek S, Volkenandt S, Dalatieh F, Reidelbach M, et al. Hydrogen-Bonded Network and Water Dynamics in the D-channel of Cytochrome c Oxidase. *Journal of Membrane Biology*. 2018; 251:299–314. <https://doi.org/10.1007/s00232-018-0019-x> PMID: 29435610
18. Luca AD, Gamiz-Hernandez AP, Kaila VRI. Symmetry-related proton transfer pathways in respiratory complex I. *PNAS*. 2017. <https://doi.org/10.1073/pnas.1706278114>.
19. Peng Y, Swanson JMJ, Kang S, Zhou R, Voth GA. Hydrated Excess Protons Can Create Their Own Water Wires. *The Journal of Physical Chemistry B*. 2015; 119:9212–9218. <https://doi.org/10.1021/jp5095118> PMID: 25369445
20. Liang R, Swanson JMJ, Peng Y, Wikström M, Voth GA. Multiscale simulations reveal key features of the proton-pumping mechanism in cytochrome c oxidase. *PNAS*. 2016; 113:7420–7425. <https://doi.org/10.1073/pnas.1601982113> PMID: 27339133
21. Liang R, Swanson JMJ, Wikström M, Voth GA. Understanding the essential proton-pumping kinetic gates and decoupling mutations in cytochrome c oxidase. *PNAS*. 2017; 114:5924–5929. <https://doi.org/10.1073/pnas.1703654114> PMID: 28536198
22. Noe F, Fischer S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Current Opinion in Structural Biology*. 2008; 18:154–162. <https://doi.org/10.1016/j.sbi.2008.01.008> PMID: 18378442
23. Jonsson H, Mills G, Jacobsen KW. Nudged Elastic Band Method for finding minimum energy paths of transitions. In: Berne BJ, Ciccotti G, Coker DF, editors. *Classical and Quantum Dynamics in Condensed Phase Simulations*. World Scientific Publishing Co.; 1998, pp. 385–404.
24. Henkelmann G, Jonsson H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *The Journal of Chemical Physics*. 2000; 113:9978–9985. <https://doi.org/10.1063/1.1323224>
25. Henkelmann G, Uberuaga B, Jonsson H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics*. 2000; 113:9901–9904.
26. Trygubenko S, Wales D. A doubly nudged elastic band method for finding transition states. *The Journal of Chemical Physics*. 2004; 120:2082–2094. <https://doi.org/10.1063/1.1636455> PMID: 15268346
27. Fischer S, Karplus M. Conjugate Peak Refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom. *Chemical Physics Letters*. 1992; 194:252–261. [https://doi.org/10.1016/0009-2614\(92\)85543-J](https://doi.org/10.1016/0009-2614(92)85543-J)
28. Ionova IV, Carter EA. Ridge method for finding saddle points on potential energy surfaces. *The Journal of Chemical Physics*. 1993; 98:6377–6388. <https://doi.org/10.1063/1.465100>
29. Dewar MJS, Healy EF, Stewart JJP. Location of transition states in reaction mechanisms. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics*. 1984; 80:227–233. <https://doi.org/10.1039/f29848000227>
30. Henkelmann G, Jonsson H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *The Journal of Chemical Physics*. 1999; 111:7010–7022. <https://doi.org/10.1063/1.480097>
31. Henkelmann G, Johannesson G, Jonsson H. Methods for Finding Saddle Points and Minimum Energy Paths. In: Schwartz SD, editor. *Theoretical Methods in Condensed Phase Chemistry*. Springer Netherlands; 2002, pp. 269–300.
32. Arora K, Brooks CL. Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *PNAS*. 2007; 104:18496–18501. <https://doi.org/10.1073/pnas.0706443104> PMID: 18000050
33. Ito Y, Ikeguchi M. Mechanism of the ab Conformational Change in F1-ATPase after ATP Hydrolysis: Free-Energy Simulations. *Biophysical Journal*. 2014; 108:85–97. <https://doi.org/10.1016/j.bpj.2014.11.1853>
34. Ma J, Karplus M. Molecular switch in signal transduction: Reaction paths of the conformational changes in ras p21. *PNAS*. 1997; 94:11905–11910. <https://doi.org/10.1073/pnas.94.22.11905> PMID: 9342335

35. Fischer S, Windshuegel B, Horak D, Holmes KC, Smith JC. Structural mechanism of the recovery stroke in the Myosin molecular motor. *PNAS*. 2005; 102:6873–6878. <https://doi.org/10.1073/pnas.0408784102> PMID: 15863618
36. Achtyl JL, Unocic RR, Xu L, Cai Y, Raju M, Zhang W, et al. Aqueous proton transfer across single-layer graphene. *Nature Communications*. 2015; 6:6539. <https://doi.org/10.1038/ncomms7539> PMID: 25781149
37. Hu S, Lozada-Hidalgo M, Wang FC, Mishchenko A, Schedin F, Nair RR, et al. Proton transport through one-atom-thick crystals. *Nature*. 2014; 516:227–230. <https://doi.org/10.1038/nature14015> PMID: 25470058
38. Friedman R, Fischer S, Nachliel E, Scheiner S, Gutman M. Minimum energy pathways for proton transfer between adjacent sites exposed to water. *Journal of Physical Chemistry B*. 2007; 111:6059–6070. <https://doi.org/10.1021/jp070781r>
39. Weinan E, Weiqing R, Vanden-Eijnden E. String method for the study of rare events. *Physical Review B*. 2002. <https://doi.org/10.1103/PhysRevB.66.052301>.
40. Rosta E, Nowotny M, Yang W, Hummer G. Catalytic Mechanism of RNA Backbone Cleavage by Ribonuclease H from QM/MM Simulations. *Journal of the American Chemical Society*. 2011. <https://doi.org/10.1021/ja200173a>. PMID: 21539371
41. Wales DJ. Structure, Dynamics, and Thermodynamics of Clusters: Tales from Topographic Potential Surfaces. *Science*. 1996; 271:925–929. <https://doi.org/10.1126/science.271.5251.925>
42. Calvo F, Bogdan TV de Souza VK, Wales DJ. Equilibrium density of states and thermodynamic properties of a model glass former. *The Journal of Chemical Physics*. 2007; 127:044508. <https://doi.org/10.1063/1.2749725> PMID: 17672708
43. Becker OM, Karplus M. The topology of multidimensional potential energy surfaces: theory and application to peptide structure and kinetics. *The Journal of Chemical Physics*. 1997; 106:1495–1517. <https://doi.org/10.1063/1.473299>
44. Levi Y, Becker OM. Effect of conformational constraints on the topography of complex potential energy surfaces. *Physical Review Letters*. 1998; 81:1121–1126.
45. Noe F, Krachtus D, Smith JC, Fischer S. Transition Networks for the Comprehensive Characterization of Complex Conformational Change in Proteins. *Journal of Chemical Theory and Computation*. 2006; 2:840–857. <https://doi.org/10.1021/ct050162r> PMID: 26626691
46. Imhof P. Transition networks for sampling of enzymatic reactions pathways. In: Voith GA, editor. *Methods in Enzymology. Computational approaches for studying Enzyme Mechanism*. Elsevier; 2016, pp. 249–271.
47. Reidelbach M, Betz F, Mäusle RM, Imhof P. Proton transfer pathways in an aspartate-water cluster sampled by a network of discrete states. *Chemical Physics Letters*. 2016; 659:169–175. <https://doi.org/10.1016/j.cplett.2016.07.021>
48. Eyring H. The Activated Complex in Chemical Reactions. *Journal of Chemical Physics*. 1935. <https://doi.org/10.1063/1.1749604>.
49. Bowman GR, Pande VS, Noe F. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*. Springer; 2008.
50. Hanggi P, Talkner P, Borkovec M. Reaction-rate theory: fifty years after Kramers. *Reviews on Modern Physics*. 2008; 62:251–341. <https://doi.org/10.1103/RevModPhys.62.251>
51. Stillinger FA. Exponential multiplicity of inherent structures. *Physical Review E*. 1999; 59:48–51. <https://doi.org/10.1103/PhysRevE.59.48>
52. Shen C-C, Tsai W-H. A Graph Matching Approach to Optimal Task Assignment in Distributed Computing Systems Using a Minimax Criterion. *Transactions on Computers*. 1985; 34:197–203. <https://doi.org/10.1109/TC.1985.1676563>
53. Thiel W, Weber W. Orthogonalization corrections for semiempirical methods. *Theoretical Chemistry Accounts*. 2000; 103:495. <https://doi.org/10.1007/s002149900083>
54. Foloppe N, MacKerrell AD 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*. 2000; 21:86–104. [https://doi.org/10.1002/\(SICI\)1096-987X\(20000130\)21:2%3C86::AID-JCC2%3E3.0.CO;2-G](https://doi.org/10.1002/(SICI)1096-987X(20000130)21:2%3C86::AID-JCC2%3E3.0.CO;2-G)
55. Dijkstra EA. A note on two problems in connection with graphs. *Numerische Mathematik*. 1959; 1:269. <https://doi.org/10.1007/BF01386390>
56. Marcus D. Spanning Tree Algorithms. In: Marcus D, editor. *Graph Theory: A Problem Oriented Approach*. The Mathematical Association of America; 2015, pp. 41–56.

57. Kruskal BJ. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*. 1956; 7:48–50. <https://doi.org/10.1090/S0002-9939-1956-0078686-7>
58. Martini L, Kells A, Covino R, Hummer G, Buchete N-V, Rosta E. Variational Identification of Markovian Transition States. *Physical Review X*. 2017. <https://doi.org/10.1103/PhysRevX.7.031060>.
59. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *Journal of Computational Chemistry*. 1983; 4:187–217. <https://doi.org/10.1002/jcc.540040211>
60. Thiel W. MNDO Version 6.1. Max-Planck-Institut für Kohlenforschung, Mülheim a.d. Ruhr, Germany, 2004.
61. Wu X, Thiel W, Pezeshki S, Lin H. Specific Reaction Path Hamiltonian for Proton Transfer in Water: Reparameterized Semiempirical Models. *Journal of Chemical Theory and Computation*. 2013; 9:2672–2686. <https://doi.org/10.1021/ct400224n> PMID: 26583861
62. Nesetril J, Milkova E and Nesetrilova H. Otakar Boruvka on minimum spanning tree problem. Translation of both the 1926 papers, comments, history. *Discrete Mathematics*. 2001; 23:3–36. [https://doi.org/10.1016/S0012-365X\(00\)00224-7](https://doi.org/10.1016/S0012-365X(00)00224-7)