Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2020

Supporting Information: Cooperative Protein-Solvent Tuning of Proton Transfer Energetics: the Example of Carbonic Anhydrase

Laura Zanetti-Polzi,[†] Massimiliano Aschi,[‡] and Isabella Daidone*[,‡]

†*CNR Institute of Nanoscience, Via Campi 213/A, I-41125 Modena, Italy*

‡*Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio, I-67010 L'Aquila, Italy*

E-mail: isabella.daidone@univaq.it

The Perturbed Matrix Method

The MD-PMM approach is a hybrid quantum/classical theoretical-computational approach, similar in spirit to other hybrid methods,[S1–S5] based on MD simulations and on the PMM. In the MD-PMM approach,[S6,S7] the part of the system where the quantum processes of interest occur, that is the quantum center (QC), is treated at the quantum level, and the rest of the system is modeled as an atomistic-molecular classical subsystem exerting an electrostatic effect on the QC electronic states. The main difference with other hybrid methods is that in the MD-PMM, the whole system (including the QC) phase space is sampled by classical MD simulations, allowing an extensive sampling of the QC and environment configurational space. The electrostatic perturbation of the environment is included *a posteriori*: the electronic properties of the isolated QC (unperturbed properties) are calculated quantum-chemically

in vacuum (i.e., in the gas phase) and then, for each configuration generated by all-atoms classical MD simulations of the whole system, the electrostatic effect of the instantaneous atomistic configurations of the environment is included as a perturbing term within the QC Hamiltonian operator. This allows to take into account the effect of the fluctuating perturbing environment (the solvent and the part of the solute which is not treated at the quantum level) on the quantum properties of the QC.

For each configuration of the whole system obtained from the MD simulation, the effect of the external environment on the QC eigenstates is included by building and diagonalizing the perturbed electronic Hamiltonian matrix $\widehat{H}$ constructed in the basis set of the unperturbed Hamiltonian eigenstates of the QC. Indicating with $\mathcal{V}$ and $\mathbf{E}$ the perturbing electric potential and field, respectively, exerted by the environment on the QC:

$$\widehat{H} \cong \widehat{H}^0 + \widetilde{I} q_T \mathcal{V} + \widetilde{Z}_1 \tag{1}$$

$$[\widetilde{Z}_1]_{j,j'} = -\mathbf{E} \cdot \langle \phi_j^0 | \hat{\boldsymbol{\mu}} | \phi_{j'}^0 \rangle \tag{2}$$

where $\widehat{H}^0$ is the QC unperturbed electronic Hamiltonian (i.e., as-obtained considering the isolated QC) and q_T, $\hat{\boldsymbol{\mu}}$ and ϕ_j^0 are the QC total charge, dipole operator and unperturbed electronic eigenfunctions, respectively, $\widetilde{I}$ is the identity matrix and the angled brackets indicate integration over the electronic coordinates.

At each frame of the MD simulation, the perturbed electronic Hamiltonian matrix is constructed and diagonalized, providing a continuous trajectory of perturbed eigenvalues and eigenvectors to be used for evaluating the QC instantaneous perturbed quantum observable of interest as, in the present case, the QC ground state energy in the protonated and deprotonated states. More details on the method can be found in the original articles.[S6,S7]

For the more specific task of investigating the proton transfer thermodynamics, the free energy change associated to the proton transfer (PT) reaction between the active site and

His64 is calculated. Given the PT reaction:

$$\mathrm{ZnOH^- + His64H^+ \rightleftharpoons ZnH_2O + His64} \tag{3}$$

the energy variation upon PT (the PT energy) is calculated with the MD-PMM approach for each configuration obtained from the MD simulations in both the reactants and the products states, as defined in Eq. 3. In the present case, the proton donor and acceptor are located at a relatively high distance inside the protein matrix (0.7 and 1.1 nm in the crystal structure of wt HCA II for the inward and outward conformers of His64, respectively) and sample variable relative orientations. Therefore, the active site and His64 are treated as separate QCs interacting with each other as well as with their environment. The perturbed states of the active site, either in the protonated or in the deprotonated condition, are obtained considering the corresponding QC as perturbed by the electric field provided by His64 as well as the rest of the atomic-molecular system, both treated within the semiclassical approximation. The same procedure is performed for His64, the corresponding QC of which is perturbed by the active site and the rest of the atomic-molecular system treated within the semiclassical approximation. The diagonalization at each MD frame of the perturbed electronic Hamiltonian matrices (see Eq. 1) with either the active site or His64 (in the protonated and in the deprotonated condition) allows the calculation of the perturbed energy variation upon PT at each MD configuration i.e., the time evolution of the PT energy $\Delta\mathcal{U}_e$.

Then, the Gibbs free energy change ΔG^0 associated to the PT can be calculated as:

$$\begin{aligned} \Delta G^0 &= -k_B T \ln\langle e^{-\beta\Delta\mathcal{H}}\rangle_R = k_B T \ln\langle e^{\beta\Delta\mathcal{H}}\rangle_P \\ &\cong -k_B T \ln\langle e^{-\beta\Delta\varepsilon}\rangle_R = k_B T \ln\langle e^{\beta\Delta\varepsilon}\rangle_P \end{aligned} \tag{4}$$

In the above equation $\Delta\mathcal{H}$ is the QC-environment whole energy change upon PT, with $\Delta\varepsilon$ the corresponding QC perturbed electronic ground state energy change. The angle

brackets subscripts R and P indicate that both the energy change as well as the averaging are obtained either in the reactants (R) or products (P) ensemble, as defined in Eq. 3, and the approximation $\Delta\mathcal{H} \cong \Delta\varepsilon$ is used, i.e., the environment internal energy change associated with the QC reaction is disregarded (being exactly zero when considering typical MD force fields).

From Eq. 4 it follows that $-k_BT\ln\langle e^{-\beta\Delta\varepsilon}\rangle_R$ and $k_BT\ln\langle e^{\beta\Delta\varepsilon}\rangle_P$ provide the upper and lower bounds of ΔG^0 and hence it can be written:

$$\Delta G^0 \cong \frac{k_BT}{2}\ \ln\frac{\langle e^{\beta\Delta\varepsilon}\rangle_P}{\langle e^{-\beta\Delta\varepsilon}\rangle_R} \tag{5}$$

In this last equation the perturbed electronic ground state energy change as well as the ensemble averages are evaluated via the previously described MD-PMM approach, i.e. by diagonalizing at each MD frame the perturbed electronic Hamiltonian matrices (Eq. 1).

Quantum chemical calculations

To calculate the free energy change corresponding to the proton transfer reaction between the active site and His64, both groups are selected as QCs. For the active site, the Zn atom, the active site water molecule/hydroxide ion and the side chains of the three active site histidines (His94, His96 and His119) modeled as imidazole groups are included in the system sub-part to be treated at the quantum level. For His64, the side chain group (imidazole) is selected as QC. For both QCs, quantum chemical calculations are performed in both protonation states (i.e., for the active site considering both a Zn-bound water molecule and a Zn-bound hydroxide ion and for His64 on imidazole and imidazolium) in order to obtain the unperturbed electronic eigenfunctions and properties to be used in the MD-PMM approach. Quantum calculations are performed at the Density Functional Theory (DFT) level[S8] with the 6-31+G(d) basis set[S9] in conjunction with the B3LYP functional[S10] (Time-dependent DFT[S11] is used for evaluating the properties of the excited states). For each

protonation state, a 6×6-dimensional Hamiltonian matrix is evaluated and diagonalized at each MD simulation step, according to the MD-PMM procedure (see Perturbed Matrix Method section). All quantum calculations are carried out using the Gaussian09 package.[S12]

Molecular dynamics simulations

To evaluate the proton transfer reaction free energy change, MD simulations are performed in both the reactants and the products states, i.e, with the Zn-bound hydroxide ion in the active site and double protonated (+1 charged) His64 and with the Zn-bound water molecule in the active site and neutral His64. The crystal structure of wt HCA II (PDB ID 3ks3[S13]) and of the Y7F/N67Q Mutant (PDB ID 4idr[S14]) are used as starting structure for the MD simulations. The OPLS/AA force field parameters[S15] are adopted for the protein. To properly model the active site, bonds are added between the Zn atom and its four ligands at the crystallographic bond length for both wt HCA II (3ks3) and Y7F/N67Q HCA II. All the angles coming from these bonds are set at the crystallographic value. The atomic charges for the zinc and its ligands in both protonation states are obtained at the DFT level by fitting the classical electrostatic potential outside the molecule to the corresponding QM potential using the electrostatic potential fit procedure (ESP charges)[S16] and properly scaled to ensure the compatibility with the OPLS/AA force field. In the crystal structure of both wt HCA II (3ks3) and Y7F/N67Q HCA II (4idr) the ε nitrogen of His64 is closer to the Zn atom of the active site with respect to the δ nitrogen. Therefore, the MD simulations in the products ensemble are performed with His64 protonated at N_δ.

MD simulations are performed using the GROMACS package[S17] in the NPT ensemble at 300 K and 1 bar using the velocity rescaling temperature coupling[S18] and the Parrinello-Rahman barostat.[S19] The starting configurations are solvated in a dodecahedral water box large enough to contain the solute and at least 1 nm of solvent on all sides. Periodic boundary conditions are used and the long range electrostatic interactions are treated using the particle

mesh Ewald method.[S20] The LINCS algorithm is used to constrain bond lengths.[S21] After a solute optimization and a subsequent solvent relaxation, each system is gradually heated from 50 to 300 K using short MD simulations. The trajectories are then propagated for 75 ns for each system and each ensemble. Subsequently, a structure randomly extracted from the reactants(products) ensemble is used as starting structure for a new 75 ns-long MD simulation in the products(reactants) ensemble. Therefore, a total sampling of 150 ns is obtained for each system and each ensemble. Coordinates are saved at every 1 ps.

Analysis of the protein environment around His64

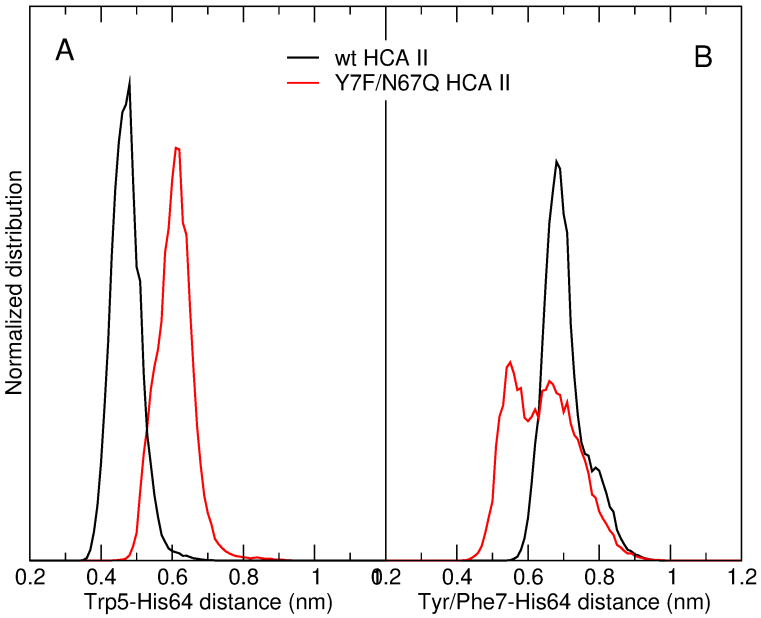

Figure S1: Normalized distribution of the distance between the center of mass of the side chains of Trp5 and His64 (A) and of Tyr/Phe7 and His64 (B) along the MD simulation in the reactants ensemble of wt HCA II (black) and Y7F/N67Q HCA II (red).

References

(S1) Gao, J.; Truhlar, D. G. Quantum mechanical methods for enzyme kinetics. *Ann. Rev. Phys. Chem.* **2002**, *53*, 467–505.

(S2) Vreven, T.; Morokuma, K. Chapter 3 Hybrid Methods: ONIOM(QM:MM) and QM/MM. *Ann. Rep. Comp. Chem.* **2006**, *2*, 35–51.

(S3) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theor. Chem. Acc.* **2007**, *117*, 185–199.

(S4) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

(S5) Liu, M.; Wang, Y.; Chen, Y.; Field, M. J.; Gao, J. QM/MM through the 1990s: the first twenty years of method development and applications. *Isr. J. Chem.* **2014**, *54*, 1250–1263.

(S6) Aschi, M.; Spezia, R.; Di Nola, A.; Amadei, A. A first priciples method to model perturbed electronic wavefunctions: the effect of an external electric field. *Chem. Phys. Lett.* **2001**, *344*, 374–380.

(S7) Amadei, A.; D'Alessandro, M.; D'Abramo, M.; Aschi, M. Theoretical characterization of electronic states in interacting chemical systems. *J. Chem. Phys.* **2009**, *130*, 08410–08415.

(S8) Parr, R. G.; Yang, W. Density-functional theory of the electronic structure of molecules. *Ann. Rev. Phys. Chem.* **1995**, *46*, 701–728.

(S9) Krishnan, R.; Binkley, J.; Seeger, R.; Pople, J. Self–consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72*, 650–654.

(S10) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(S11) Adamo, C.; Jacquemin, D. The calculations of excited-state properties with Time-Dependent Density Functional Theory. *Chem. Soc. Rev.* **2013**, *42*, 845–856.

(S12) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. Gaussian 09, revision A. 1. *Gaussian Inc., Wallingford, CT* **2009**,

(S13) Avvaru, B. S.; Kim, C. U.; Sippel, K. H.; Gruner, S. M.; Agbandje-McKenna, M.; Silverman, D. N.; McKenna, R. A short, strong hydrogen bond in the active site of human carbonic anhydrase II. *Biochemistry* **2010**, *49*, 249–251.

(S14) Mikulski, R.; West, D.; Sippel, K. H.; Avvaru, B. S.; Aggarwal, M.; Tu, C.; McKenna, R.; Silverman, D. N. Water networks in fast proton transfer during catalysis by human carbonic anhydrase II. *Biochemistry* **2013**, *52*, 125–131.

(S15) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(S16) Besler, B. H.; Merz Jr, K. M.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comp. Chem.* **1990**, *11*, 431–439.

(S17) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: fast, flexible, and free. *J. Comp. Chem.* **2005**, *26*, 1701–1718.

(S18) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(S19) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(S20) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(S21) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.* **1997**, *18*, 1463–1472.