Supporting Information
Molecular basis of CLC antiporter inhibition by fluoride

Maria Gabriella Chiariello,†,‡ Viacheslav Bolnykh, ‡ Emiliano Ippoliti,†,‡ Simone Meloni,§ Jógvan Magnus Haugaard Olsen,§ Thomas Beck,‖ Ursula Rothlisberger,† Christoph Fahlke,*,# Paolo Carloni,*,†,‡,††

† Institute for Advanced Simulation (IAS-5) and Institute of Neuroscience and Medicine (INM-9), Forschungszentrum Jülich, 52425 Jülich, Germany
‡ JARA-HPC, Forschungszentrum Jülich, D-54245 Jülich, Germany
⊥ Laboratory of Computational Chemistry and Biochemistry, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
¶ Dipartimento di Scienze Chimiche e Farmaceutiche, Università degli Studi di Ferrara, Via Luigi Borsari 46, I-44121, Ferrara, Italy
§ Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway
‖ Department of Chemistry, University of Cincinnati, Cincinnati, Ohio 45221, United States
# Institute of Complex Systems (ICS), ICS-4: Zelluläre Biophysik, Forschungszentrum Jülich, Jülich, Germany.
†† Department of Physics, RWTH Aachen University, 52056 Aachen, Germany
Computational details

1. Classical molecular dynamics simulations

*Initial structural model.* Our calculations were based on the *E. coli* CLC-ec1 X-ray structure\(^1\) (PDBID: 1OTS). The ionizable residues are considered in their expected protonation state at pH 7: Glu and Asp residues were ionized, His residues were neutral. The protonation states were the same in the two subunits. The Cl\(^-\) in the central binding site\(^1\) was replaced by F\(^-\) (as in Ref. \(^2\)) in both the subunits. No other F\(^-\) ion is present in the cavity of the protein.

The protein was embedded in 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer\(^3-4\) by using the CHARMM-GUI web server\(^5\). The system was solvated with 33002 water molecules with 150 mM NaCl salt concentration. The simulation box contained a total of 165,732 atoms with a box length of \(L_x=130.21\ \text{Å}, L_y=130.21\ \text{Å}, L_z=105.14\ \text{Å}\).

*Simulations.* Proteins, lipids and water were described with CHARMM-CMAP\(^6\), CHARMM36\(^7\), TIP3P\(^8\) force fields, respectively. The fluoride parameters were described in Ref. \(^9\). Long range interactions were evaluated using the particle-mesh Ewald (PME) summations\(^10\) with a cutoff in real space of 12 Å. The Lennard-Jones interactions were truncated at 12 Å with an atom-based force switching function, which starts to be effective at 10 Å. The integration time step was set at 2 fs. The bonds involving hydrogen atoms were constrained by the LINCS algorithm.\(^11\)

After energy minimization, the system was equilibrated first in an NVT ensemble for 1 ns, followed by a 15ns-long simulation in the NPT ensemble, at 310 K and 1.0 atm pressure. The Nåse-Hoover chain thermostat\(^12-13\) was employed to control the temperature. A semi-isotropic Parrinello-Rahman barostat\(^14\) with reference pressure 1 atm and isothermal compressibility of \(4.5 \times 10^{-5}\) was used to maintain the pressure of the system. The production run was performed for 300 ns. All the MD calculations were performed by using GROMACS-2018.3 package\(^15\).

2. QM/MM simulations

*Initial models.* Ten QM/MM models (*TRAJ.I-X*) were based on 5 different force-field based MD snapshots: \textbf{MD1} (Fig. S3A, S4A), \textbf{MD2} (Fig. S3B, S4B), \textbf{MD3} (Fig. S3C, S4C), \textbf{MD4} (Fig. S3D, S4D), \textbf{MD5} (Fig. S3E, S4E). Obviously, the number of waters in the protein’s permeation pathway between E148 and E203, in general, differed on passing from one snapshot to another. An excess proton was generated in different positions of the water network inside the permeation pathway. The QM regions of *TRAJ. VI-X* differ from those of *TRAJ.I-V* for the inclusion of the E203, R205 and the water molecules interacting with E203. Overall, the ten QM regions in the ten models consisted of the following atoms:
The QM region in **TRAJ.I** (Fig. 1b, S3A, based on **MD1**) includes: Fluoride, E148\(^1\), R147, hydronium, 9 water molecules (55 atoms, embedded in a cubic box of length 40.0 a.u.). It includes the water molecules inside the protein’s pore between E148 and E203, except those forming H-bonds with E203, which were treated at the MM level along with E203. The excess proton was added on a water molecule at the inner border of the permeation pathway (pore).

The QM regions in **TRAJ.II** (Fig. 1c and S3B, based on **MD2**) and in **TRAJ.V** (Fig. S3 E, based on **MD5**) are the same as the QM region of **TRAJ.I**, except that 10 water molecules are present instead of 9 (58 atoms, embedded in a cubic box of length 40.0 a.u.). The excess proton was added on a water molecule at the inner border of the permeation pathway.

The QM region in **TRAJ.III** (Fig. 1d and S3C, based on **MD3**) includes: Fluoride, E148, R147, S107, Y445, hydronium, 6 water molecules (65 atoms, embedded in a cubic box of length 48.0 a.u.). It includes two water molecules located between F\(^-\) and E148 along with the F\(^-\) coordinating groups (4 water molecules and Y445 and S107 side chains). The other waters inside the channel were not included as they are not expected to participate in the proton transfer. The residues coordinating F\(^-\) (Y445 and S107) are also included to complete the description of the coordination sphere of the halide at QM level. The excess proton was added on a water molecule between F\(^-\) and E148.

The QM region in **TRAJ.IV** (Fig. S3D, based on **MD4**) is the same as in the QM region of **TRAJ.III**, except that 5 water molecules are present instead of 6 (62 atoms, embedded in a cubic box of length 48.0 a.u.). Indeed, it includes the two water molecules located between F\(^-\) and E148 along with F\(^-\) coordinating groups (3 water molecules and Y445 and S107 side chains). The excess proton was added on a water molecule between F\(^-\) and E148.

The QM region in **TRAJ.VI** (Fig. S4A, based on **MD1**) is the same as **TRAJ.I**, except that it also includes E203, and R205, and 11 water molecules instead of 9 (84 atoms, embedded in a box whose lengths are: 50.0 a.u., 42.0 a.u., 58.0 a.u.).

The QM regions in **TRAJ.VII** (Figure S4B, based on **MD2**) and in **TRAJ.X** (Fig. S4 E, based on **MD5**) are the same as **TRAJ.VI**, except that they feature 13 water molecules instead of 11 (90 atoms embedded in a box whose lengths are 50.0 a.u., 42.0 a.u., 58.0 a.u.)

The QM region in **TRAJ.VIII** (Fig. S4C, based on **MD3**) is the same as the QM region of **TRAJ.III**, except that it also includes E203, and R205, and 12 water molecules instead of 6 (106 atoms, embedded in a box whose lengths are 50.0 a.u., 42.0 a.u., 58.0 a.u.).

The QM region in **TRAJ.IX** (Fig. S4D, based on **MD4**) is the same as the QM region of **TRAJ.IV**, except that it also includes E203, and R205, and 10 water molecules instead of 6 (100 atoms, embedded in a box whose lengths are: 50.0 a.u., 42.0 a.u., 58.0 a.u.)\(^2\)

---

\(^1\) Unless otherwise specified, only the side chain of the residues is part of the QM regions of **TRAJ.I-TRAJ.X**.

\(^2\) Additional water molecules were not included in the calculations, as they are not expected to participate in the PT between the three negative moieties, F\(^-\), E148 and E203.
The MM regions consisted of the rest of MD1-5 (protein frame, membrane, solvent), except that a chloride ion was added in the bulk solvent to ensure the overall neutrality of the system.

**Calculations:** The QM part was described by density functional theory (DFT),\textsuperscript{16} using B3LYP exchange-correlation functional \textsuperscript{17-18} (for TRAJ I-V) and BLYP\textsuperscript{19-20} (for TRAJ VI-X). We used a plane wave basis set cutoff of 90 Ry. Monovalent carbon pseudopotentials were used to saturate the dangling bonds at the boundaries between the QM and MM regions. The core electrons were described using norm-conserving pseudopotentials of the Martins–Troullier type while the valence electrons were treated explicitly.\textsuperscript{21} Constant temperature simulations were achieved by coupling the system with the Nose-Hoover thermostat,\textsuperscript{13} with a target temperature of 300 K and a length of the Nose-Hoover chain of 4. For each system, we used a time step of 0.5 fs. To describe the electrostatic interactions between the QM and MM subsystems, the fully Hamiltonian electrostatic coupling scheme developed by Laio et al. as implemented in MiMiC, was employed.\textsuperscript{22} The QM and MM systems of TRAJ I-V and TRAJ VI-X evolved following DFT-based and force-field based molecular dynamics, respectively. Trajectories for TRAJ I-V were collected for 4 ps. As the proton transfer to either F\textsuperscript{-} or E148 turned out to occur within 0.5 ps for all of these simulations (See Figs 1 and S3), the trajectories for TRAJ VI-X were collected for only 0.7 ps. In this timeframe, a proton transfer to either F\textsuperscript{-}, or E148 or E203 occurred (See Figs S4).

The QM/MM simulations were carried out by means of a recently developed framework for multiscale modeling (MiMiC) (please see ref. 29 and 30 of the main text). MiMiC couples two popular and established programs for ab initio and classical molecular dynamics (MD) simulations, i.e. CPMD 4.3 \textsuperscript{23} and GROMACS 2019 \textsuperscript{15}, respectively.

### 3. QM/MM well-tempered Meta Dynamics (MTD) simulations

MTD is an exact method\textsuperscript{24} to calculate the free energy of a given process as a function of a set of reaction coordinates or collective variables (CVs). We used exactly the same setup as in the section above, except that we employed only the BLYP functional, less expensive than B3LYP but still providing reasonable energetics.\textsuperscript{25} Four QM/MM models were used (MTD.I-IV). Here, we specify which atoms are included in the quantum region\textsuperscript{3}, along with the MTD parameters, the definition of the collective variables and other information regarding the simulations.

The QM region of MTD.I (Fig. 2A, based on TRAJ. IV) includes E148\textsuperscript{4}, R147, F, S107, Y445, H\textsuperscript+ and two water molecules (50 atoms, embedded in a cubic box of length 48 a.u.). We used three CVs. These are the H-O(E148) H-O(W2) and H-F distances, as defined in Fig. S6. The MTD parameters are as follows: Gaussian height 2.0 kJ/mol, Gaussian sigma 0.2 Å, bias factor 30, gaussian deposition rate 50 fs. To reconstruct the histogram of the sampled configuration we used the algorithm described in Ref. \textsuperscript{26}, reweighting on the H-F and H-  

\textsuperscript{3}The rest of the system has been treated at the MM level as in the Section above.  
\textsuperscript{4}Unless otherwise specified, we included only the side chains of the protein residues in the quantum region of MTD.I-IV.
O(E148) variables. During the simulations, a harmonic potential restraint potential $U$ was added on the F-O(wat2), O(Wat2)-O(Wat1) and O(Wat1)-O(E148) distances (Labeling of water molecules in Fig. S6) so as to avoid the insertion of the surrounding classical waters. These restraints were activated if any of the distances were greater than 3 Å. $U$ reads:

$$U = k (x_i - a_i)^2$$

Where $k = 100$ kJ mol$^{-1}$ Å$^{-2}$, $a_i$ is the threshold (set here to 3 Å). To quantify the effect of the potential, one can consider that the energy penalty for a deviation of 0.1 Å above the threshold value is 1 kJ/mol. A limitation of this model is the inclusion of only two QM waters. However, the relative stability of the two minima (H-F and H-O(E148)) is not expected to change with the inclusion of the surrounding solvation waters as proton competitors. Indeed, the presence of two anions acting as proton acceptors is expected to trap the proton in their close proximity.

The QM region of MTD.II (Fig. 2B, based on MTD. III) includes: E148, R147, F$^-$, H$^+$ (28 atoms, embedded in a cubic box of length 38 a.u.). The CV is the F-H distance. The parameters of the MTD calculations are: Gaussian height 2.0 kJ/mol, Gaussian sigma 0.2 Å, bias factor 30, gaussian deposition rate 50 fs. No water molecule interacts directly with the excess proton. The water interacting with the two negative moieties (see inset of Fig S8D) were treated at MM level. We applied the reweighted procedure on the following parameters: H-F (Fig 2B), H-F and F-O(E148) (Fig. S8A) and H-O(E148) (Fig. S8D).

The QM region of MTD.III (Fig. 2C, based on TRAJ. III) includes: protonated E148 (side chain and backbone unit), R147, F$^-$ (35 atoms, embedded in a cubic box of length 40 a.u.). The CV is the E148’s N-Cα-Cβ-Cγ dihedral angle. The parameters for the MTD are: Gaussian height 1.2 kJ/mol, Gaussian sigma 0.35 rad, bias factor 20, gaussian deposition rate 50 fs. We applied the reweighted procedure on the N-Cα-Cβ-Cγ dihedral angle (Fig. 2C).

The QM region of MTD.IV (Fig. 2D, based on TRAJ. IV) includes: E148 (side chain and backbone unit), R147, F-H (35 atoms, embedded in a cubic box of length 40 a.u.). The CV is the E148’s N-Cα-Cβ-Cγ dihedral angle. The parameters for the MTD are: Gaussian height 1.2 kJ/mol, Gaussian sigma 0.35 rad, bias factor 20, gaussian deposition rate 50 fs. We applied the reweighted procedure on the N-Cα-Cβ-Cγ dihedral angle (Fig. 2D).

In all the MTD runs, it was not necessary to include E203 as it plays no role for fluoride inhibition as a proton acceptor group. Its influence is fully taken into account by including this residue at MM level.

We used the CPMD 4.3 with MiMiC support interfaced with the PLUMED 2.5.3 plugin.

4. Mulliken Population analysis on Fluoride anion

Some DFT functionals can have issues in describing the anions resulting in spurious charge transfer effects. The calculation can be in principle improved by the employment of long-range corrected functionals. Here, we performed a Mulliken population analysis to compare the
charge on the fluoride with the BLYP, B3LYP and its long-range corrected version (CAM-B3LYP)\textsuperscript{31}. We consider the fluoride embedded in two different environments i) in its binding site, coordinated by S107 and Y445 (Fig. 1A) ii) in direct contact with E148 (inset in Fig. 2B and S8D). The resulting values are reported in SI (Table S1). In both cases, the B3LYP and CAM-B3LYP functionals gave very similar results. This suggests that in this case the inclusion of the long-range correction in the functional does not appear to significantly affect the charge distribution of the fluoride.

5. **Disrupting the F\textsuperscript{−}/ protonated E148 complex of Fig. 2B**

From the free energy landscape plotted in the Fig. S8A, one can evince that a free energy of ~5 kcal/mol would be required to bring the complex from a free energy minimum (corresponding to a F-O(E148) distance of 2.6 Å) to about 3.4 Å, where the complex is basically disrupted. Because the BLYP exchange correlation functional tends to underestimate proton transfer barriers\textsuperscript{32}, the free energy may be even larger than this value.
Supporting figures

Figure S1. Classical MD simulations of *E. coli* CLC-ec1. A) Number of water molecules inside the permeation pathway plotted as function of the simulation time. B) Radial distribution function (RDF) of fluoride – hydrogen belonging to water molecules.
Figure S2 Classical MD simulations of *E. coli* CLC-ec1. A) Tyr445-F⁻ distance; B) Ser107-F⁻ distance; C) E148 $\chi_1$ dihedral angle; D) F-O1(E148) (black, average value 7.5 Å)/ F-O2(E148) (red, average value 8.43 Å) distances, plotted as a function of the simulation time. The average values of the F-O1(E148) and F-O2(E148) distances (7.53 Å and 8.43 Å) allow for the insertion of a couple of waters to stabilize the negative charges of the two anions.
Figure S3. Proton transfer processes during the unbiased QM/MM simulations: TRAJ.I-V simulations described at p. 3 SI. The left panel of A-E shows the starting configurations, taken from our classical MD simulation. They consist of hydronium, fluoride (yellow sphere) and E148, connected via water wires. The right panels show the distance between either H-F or H-O(E148) as a function of simulated time. Its inset shows a snapshot after the PT.
Figure S4. Proton transfer processes during the QM/MM simulations: TRAJ.VI-X simulations described at p. 3 SI. The left panel of A-E shows the starting configurations, taken from our classical MD simulation. They consist of hydronium, fluoride (yellow sphere), E148 and E203 connected via water wires. The right panels show the distance between either H-F, H-O(E148) or H-O(E203) as a function of simulated time. Its inset shows a snapshot after the PT.
Figure S5. Free energy surface (kcal/mol) associated with configuration in Fig. 2A of the main text, within energy range 0-2 kcal/mol.
Figure S6. QM/MM simulation of MTD.I. Time evolution of the CVs: H-F, H-O(W2) and H-O(E148).
Figure S7. QM/MM simulation of MTD.I. Convergence of the free energy profiles as function of the three CVs.
Figure S8. QM/MM simulation of MTD.II. A) Free energy of the direct PT between F⁻ and E148 reweighted on the H-F and F-O(E148) distances. The F-H-E148 triad is a free energy minimum corresponding to a F-O(E148) distance of 2.6 Å. B) Time evolution of the CV, i.e. H-F distance. C) Convergence of free energy along the CV. D) Free energy profile reweighted on the H-O(E148) distance.
Figure S9. QM/MM simulations of MTDIII (A) and MTDIV (B). Convergence of the free energy profiles along the $\chi_1$ dihedral for A) protonated E148 and B) deprotonated E148.

Supporting Table

Table S1. Mulliken charges (a.u.) of the F$^-$ ion in its binding site, with S107 and Y445 coordinating residues without (I) (Fig. 1A) or in the presence of (II) protonated E148 (Fig. 2B and S8D). The calculation has been performed at different levels of theory, with 6-31+G(d,p) basis set, using the Gaussian09 package$^{33}$.

<table>
<thead>
<tr>
<th></th>
<th>F$^-$ Mulliken charges</th>
<th>F$^-$ Mulliken charges</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>CAM-B3LYP</td>
<td>-0.815</td>
<td>-0.788</td>
</tr>
<tr>
<td>B3LYP</td>
<td>-0.803</td>
<td>-0.783</td>
</tr>
<tr>
<td>BLYP</td>
<td>-0.750</td>
<td>-0.763</td>
</tr>
</tbody>
</table>
References