Supporting Information for: Binding Sites of Bicarbonate in Phosphoenolpyruvate Carboxylase

*Nicolas Chéron**

PASTEUR, Département de chimie, École normale supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France

Corresponding Author: nicolas.cheron@ens.psl.eu

Table of contents

I. Crystallographic structures of PEPC

ID	Year	Organism	Family	Resolution	Ligands	#Chains in .pdb	Loops 174-184 and 640-649	Loop 761-768	#Missing residues
1FIY[1]	1999	*E. coli*	Bacteria	2.80 Å	Asp	1	Inactive	No	7 (1)
1QB4[2]	1999	*E. coli*	Bacteria	2.60 Å	Mn^{2+}, Asp	1	Inactive	No	6 (1)
1JQN[3]	2002	*E. coli*	Bacteria	2.35 Å	Mn^{2+}, Asp, DCDP	1	Inactive	No	6 (1)
1JQO[3]	2002	*Z. mays*	Plant	3.00 Å	Sulfate	2	Active	No	32 (3)
3ODM[4]	2011	*C. perfringens*	Archae	2.95 Å	Malonate	8			7 (1)
3ZGB[5]	2013	*F. pringlei*	Plant	2.71 Å	Sulfate	2	Inactive	No	41 (3)
3ZGE[5]	2013	*F. trinervia*	Plant	2.49 Å	Sulfate	2	Inactive	No	43 (4)
4BXC[6]	2014	*E. coli*	Bacteria	2.86 Å	G6P, Sulfate	2	Inactive	No	31 (7)
4BXH[6]	2014	*E. coli*	Bacteria	2.24 Å	Sulfate	2	Inactive	No	32 (2)
5FDN[7]	2017	*A. thaliana*	Plant	2.20 Å	Citrate, Asp	2	Inactive	No	40 (3)
5VYJ[8]	2018	*Z. mays*	Plant	3.30 Å	Glycine	4	Active	Yes, Open	17 (2)
6MGI[9]	2019	*Z. mays*	Plant	2 .99 Å	G6P, PGA, Gly	2	Active	No	21 (2)
6U2T[10]	2020	*Z. mays*	Plant	2.80 Å	Malate, Sulfate	4	Inactive	Yes, Open	40 (3)
6V3O[10]	2020	*Z. mays*	Plant	2.91 Å	Sulfate, Citrate	8	Inactive	Yes, Open	15 (1)
8OJ9[11]	2023	*A. thaliana*	Plant	3.25 Å	–	3	Inactive	Yes, Open	15 (2)
8OJE[11]	2023	*A. thaliana*	Plant	3.14 Å	Malate	8	Inactive	~Yes, Open	29 (3)
8OJF[11]	2023	*A. thaliana*	Plant	3.04 Å	Phosphate	2	Active	~Yes, Open	31 (4)

Table SI-1. Available crystallographic structures of PEPC. Only relevant ligands are listed (Asp = aspartate, G6P = glucose-6-phosphate, PGA = 2-phosphoglycolic acid, Gly = glycine).

We present in Table SI-1 details regarding the available PEPC structures. 3ODM is described in less details because its structure is quite different from the ones of bacteria and plants. For loops 174-184 and 640-649 (maize numbering), we state whether His177 and Arg647 are found in the active site (active) or not (inactive). For loop 761-768, we state whether this loop is resolved in the structure and in which conformation. The “numbers of missing residues” column considers the chain with the fewest missing blocs, and then with the fewest missing residues; in parenthesis is given the number of missing blocs. We didn’t count missing residues in the N-terminal and C-terminal parts.

II. Presentation of PEP Carboxylase

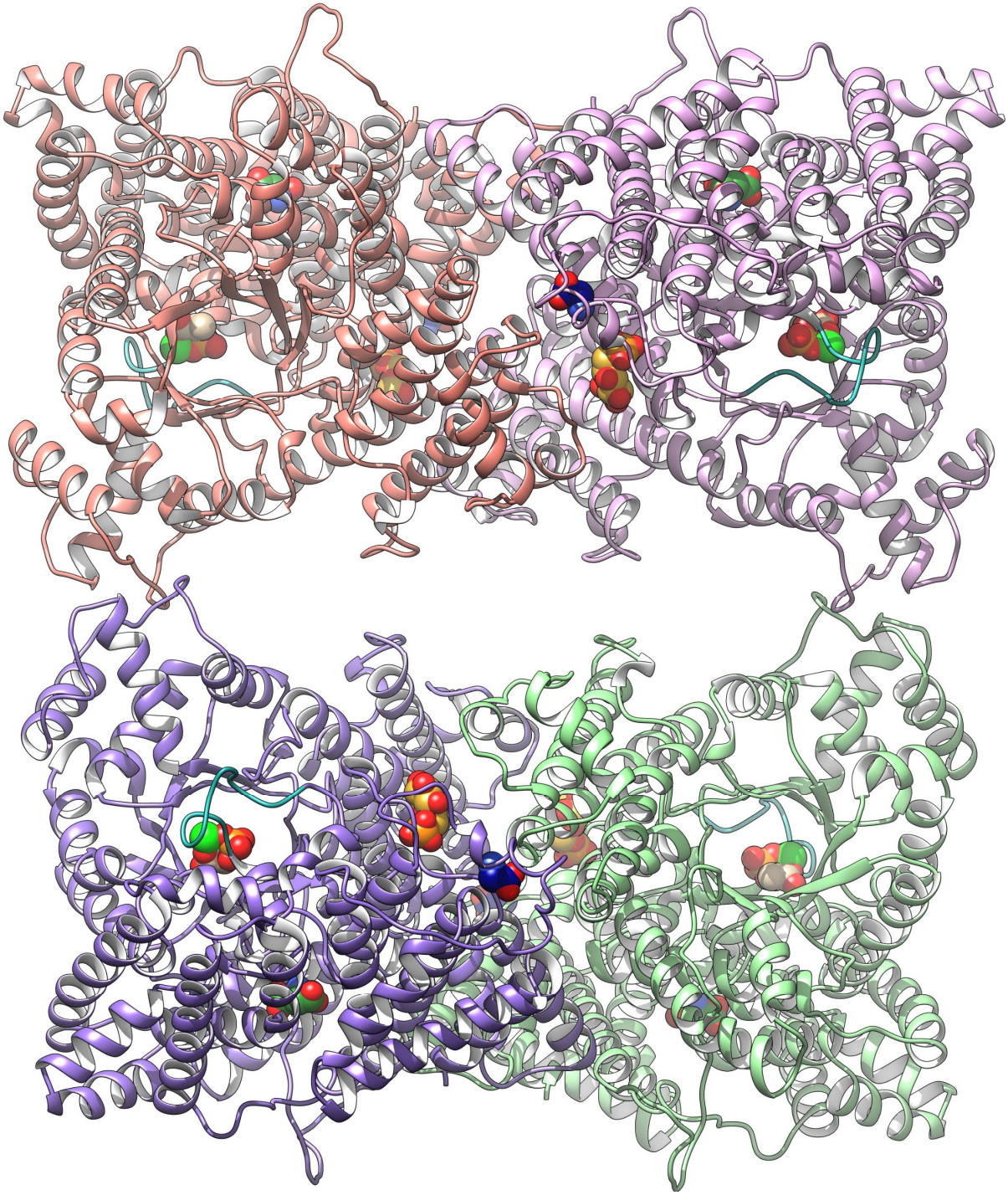


Figure SI-1. Tetramer of PEP Carboxylase. Same structure as Figure 2 of main text, which is seen from a different angle after a rotation of 90° (i.e. Figure 2 in main text is seen from the center of the tetramers and looking "above"). Ligands are in spheres with colors that matches the ones in Figure 2 of main text. The contact surface area between orange and pink subunits (top row) is 3000 Å^2, whereas the contact surface area between orange and purple subunits (left column) is 450 Å^2. The system used during this study was built from PDB 5VYJ.

III. Computational details

Molecular dynamics simulations were performed with Gromacs[12–14]. The protein was described by the Amber14SB[15] force field, the water was described by TIP3P[16] and ligands were described by GAFF2[17]. To obtain the GAFF2 parameters, the geometry of ligands were first optimized at the M06-2X/6-31+G(d,p) level of theory with an implicit water solvent (PCM) with Gaussian09 A.02[18]. The electrostatic potentials were then computed at the HF/6-31G(d) level of theory with Gaussian09 C.01, and the fitting procedures to obtain RESP charges were performed using antechamber[19] with the Merz-Kollman scheme. Finally, acpype[20] was used to obtain all GAFF2 parameters compatible with Gromacs (optimized geometries and RESP charges are provided below). To describe non-bonding interactions, we used standard parameters for the chosen force field: van der Waals interactions had a cutoff at 8 Å, and electrostatics interactions were computed with PME[21] with a separation between direct and reciprocal spaces at 8 Å. PME parameters were set to defaults values (i.e. Fourier spacing at 0.12 nm, interpolation order at 4 and tolerance at 10^{-5}). Long range dispersion corrections for energy and pressure were used. All bonds containing a hydrogen atom were constrained with the LINCS algorithm[22,23] with standard parameters (order in expansion of 4 and 1 iteration).

The energy of the system was first minimized with the steepest descent algorithm, before performing equilibration in the NPT ensemble with a time step of δt = 1 fs and using the velocity-rescale thermostat[24] and the Berendsen barostat[25]. Velocities were generated at 100 K and the system was heated to 300 K in 400 ps, and then ran for an extra 100 ps at 300 K. Production simulations were then ran with similar parameters as in the equilibration procedure, but with a time-step increased to 2 fs and the barostat changed to Parinello-Rahman[26]. For standard MD, we have used the version 2019.4 of Gromacs with the leap-frog algorithm.

Once binding pockets were identified, we performed alchemical free energy perturbations (more details are given below). Snapshots from the free MD were extracted and were solvated again. Indeed, during the initial steps, the bicarbonate could be up to 56 Å from the binding site and the first boxes that were created could thus be much larger than needed. The energies of the system were then minimized before equilibration. Alchemical free energy perturbations were then performed with Gromacs 2021 and 2021.4 patched with Plumed 2.7.1 and 2.7.3[27–29], using stochastic dynamics. All the simulations with different λ values were performed in parallel, starting from the same checkpoint that followed the equilibration.

The full computational protocol is summarized in Figure SI-2. Overall, we estimate that we have performed ~47µs of simulations, which needed ~9000 GPU.h and ~2.6 million CPU.h.

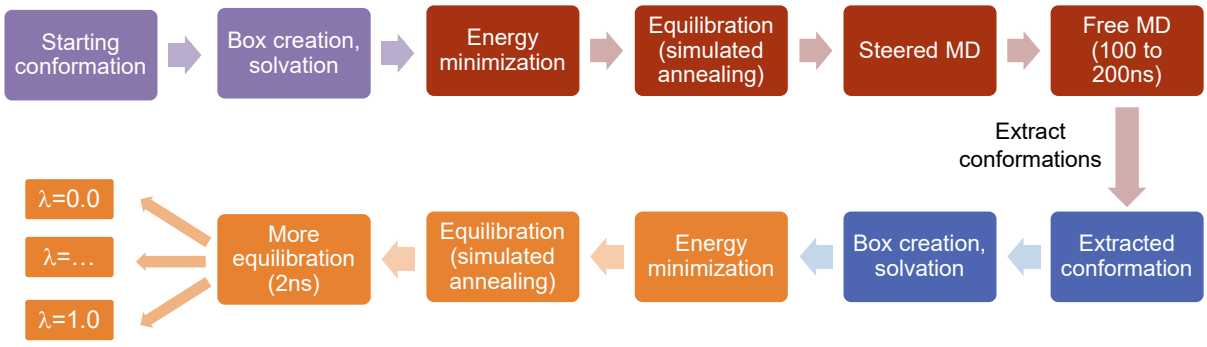


Figure SI-2. Full computational protocol.

IV. Geometry and charges of the ligands

We provide below the optimized geometry and the RESP charges of the ligands (x/y/z/q).

Bicarbonate (total charge = -1)

C	-0.030136	0.000000	-0.043110	1.005609
O	-0.050887	0.000000	1.353090	-0.698833
O	1.106026	0.000000	-0.569701	-0.829425
O	-1.156713	0.000000	-0.573780	-0.829425
H	0.871494	0.000000	1.640168	0.352074

PEP (total charge = -3)

O	20.725144	34.100514	-11.738947	-0.862954
C	21.171626	33.798357	-12.877262	0.798098
O	20.648872	34.089126	-13.981333	-0.862954
C	22.492845	33.001599	-12.910810	0.409303
O	22.840154	32.597252	-14.149807	-0.473855
P	24.230166	31.650782	-14.522289	1.273550
O	24.039468	30.342424	-13.749976	-0.953287
O	24.074326	31.526921	-16.040046	-0.953287
O	25.425218	32.494885	-14.070307	-0.953287
C	23.188118	32.779186	-11.783215	-0.802221
H	22.790979	33.156637	-10.850238	0.190447
H	24.126083	32.238316	-11.798769	0.190447

Glycine (total charge = 0)

N	60.811856	-47.690089	11.959909	-0.441453
C	60.884537	-48.380463	10.636305	-0.082733
C	61.534473	-49.769894	10.865258	0.830957
O	61.851537	-50.019171	12.057652	-0.743437
O	61.667317	-50.475013	9.850101	-0.743437
H	61.350636	-46.824355	11.994038	0.326157
H	59.881566	-48.493982	10.228239	0.100816
H	61.486695	-47.788950	9.948674	0.100816
H	59.857718	-47.480212	12.254071	0.326157
H	61.232653	-48.389020	12.613649	0.326157

Malate (total charge = -2)

C	-0.827968	-1.254155	-0.505052	0.838588
O	-1.468946	-1.937806	0.338307	-0.853664
O	-0.831335	-1.427494	-1.753166	-0.853664
C	0.008149	-0.079339	0.033820	-0.154169
C	1.399281	-0.009199	-0.582679	0.450950
C	2.171582	1.231455	-0.077875	0.675078
O	3.231260	0.988213	0.565294	-0.808356
O	1.687995	2.356615	-0.342298	-0.808356
O	2.135259	-1.186001	-0.277232	-0.747836
H	0.090806	-0.155974	1.123486	-0.015324
H	-0.522354	0.850442	-0.203371	-0.015324
H	1.295095	0.071891	-1.672275	-0.071515
H	2.900114	-0.845673	0.224419	0.363591

VI. Pushing bicarbonate inside the protein

Additional images to complement Figure 3 of main text are provided Figure SI-3 to show the starting positions of bicarbonate before steered MD. Steered MD simulations were performed on the dimer of PEPC with a single bicarbonate in each simulation.

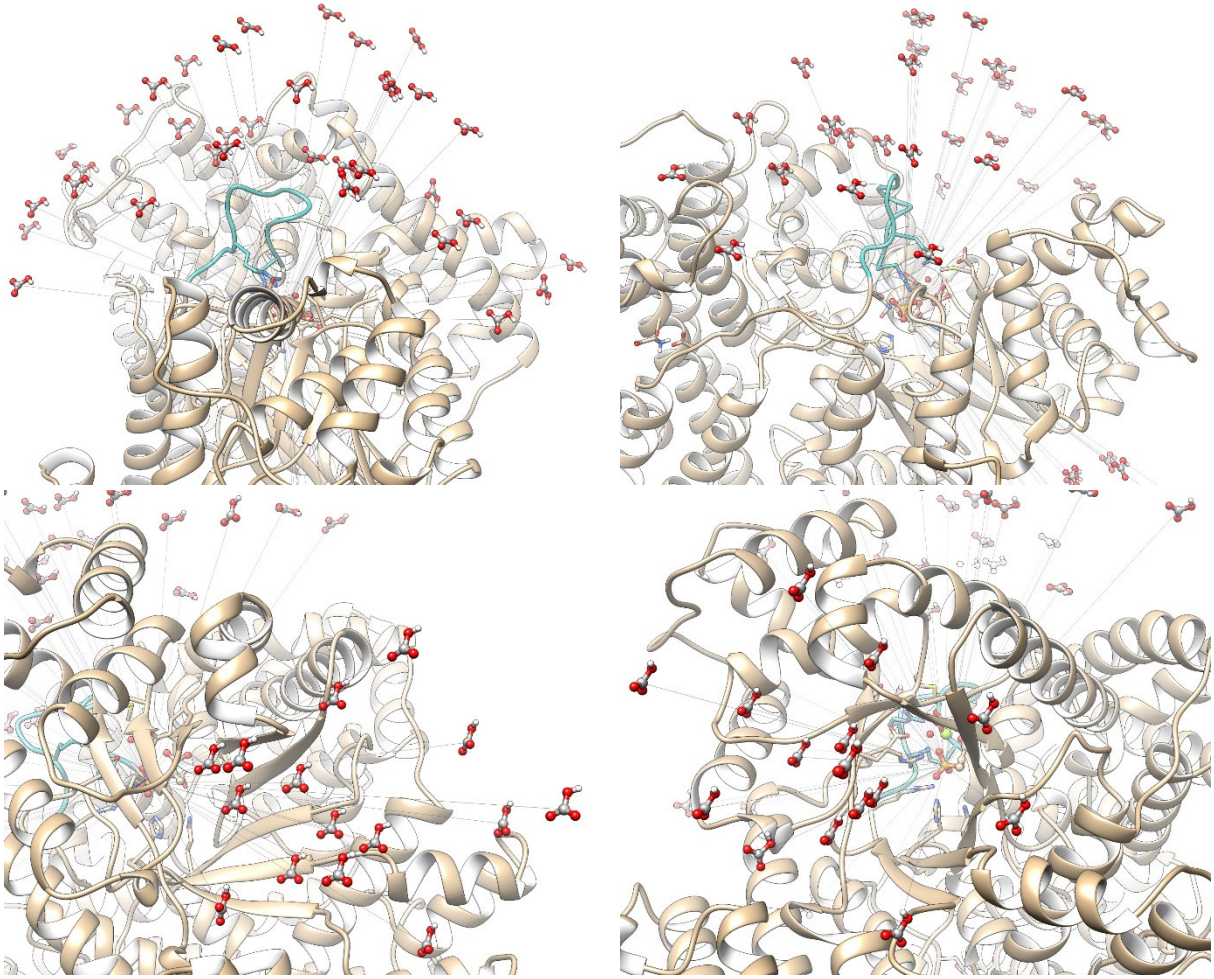


Figure SI-3. Super-imposition of starting positions of bicarbonate before steered MD simulations. Top: zoom on 35 positions with the HCO_3^-/PEP distance aligned with loop II. Bottom: zoom on 15 positions with the HCO_3^-/PEP distance along the β-barrel. The lines joining C of bicarbonate and P of PEP are shown in grey to guide the eye. Only one monomer is displayed.

Among all the steered MD simulations that were performed, two were unsuccessful and 14 led to unrealistic paths where bicarbonate moved on the other side of the protein at the beginning of the simulations (when bicarbonate could freely wander) and then penetrated through the protein when the target distance decreased. In the remaining simulations, five pathways were identified:

1. Path A (29 simulations): bicarbonate penetrates through the "top" of the protein, near loop II and "in front" of it.
2. Path B (16 simulations): bicarbonate penetrates through the β-barrel.
3. Path C (24 simulations): bicarbonate penetrates through the "top" of the protein, near loop II and "behind" it, and then reaches the active site through the bottom of the hairpin.
4. Path D (14 simulations): bicarbonate penetrates through the "top" of the protein, but slightly "on the side".
5. Path E (1 simulation): bicarbonate penetrates through the "top" of the protein, in front of loop II. Since this path was observed for a single trajectory, we didn't investigate it further.

Figure SI-4. Representation of the penetration paths during steered MD simulations: green=A, blue=B, orange=C, red=D.

Images showing the density of presence of bicarbonate during representative steered MD simulations (one for each path) are displayed on Figure SI-4. Free energy profiles for the penetration of bicarbonate in PEPC were then computed. For each path, five simulations were used. Conformations were randomly extracted from each steered MD simulation every 0.5 Å of the reaction coordinate (i.e. the bicarbonate/PEP distances). After a 500 ps equilibration, 2 ns of umbrella sampling simulations in each window were performed with a force constant of k=1000 kJ/mol/nm^2 (1157 windows in total). All data from each path were then gathered and analysed with the WHAM utility from Gromacs. Resulting free energy profiles are displayed on Figure SI-5: we observed that (1) penetration of bicarbonate through path B (i.e. the β-barrel) is highly unfavourable, and that (2) penetration through paths A, C and D are comparable and more realistic. However, no energy well is observed at short distances of paths A, C and D since at the end of the steered MD simulations, bicarbonate didn't reach an actual binding site. We also underline that the same value of the reaction coordinate in two profiles doesn't mean that the bicarbonate is at the same location in the active site since different paths were followed.

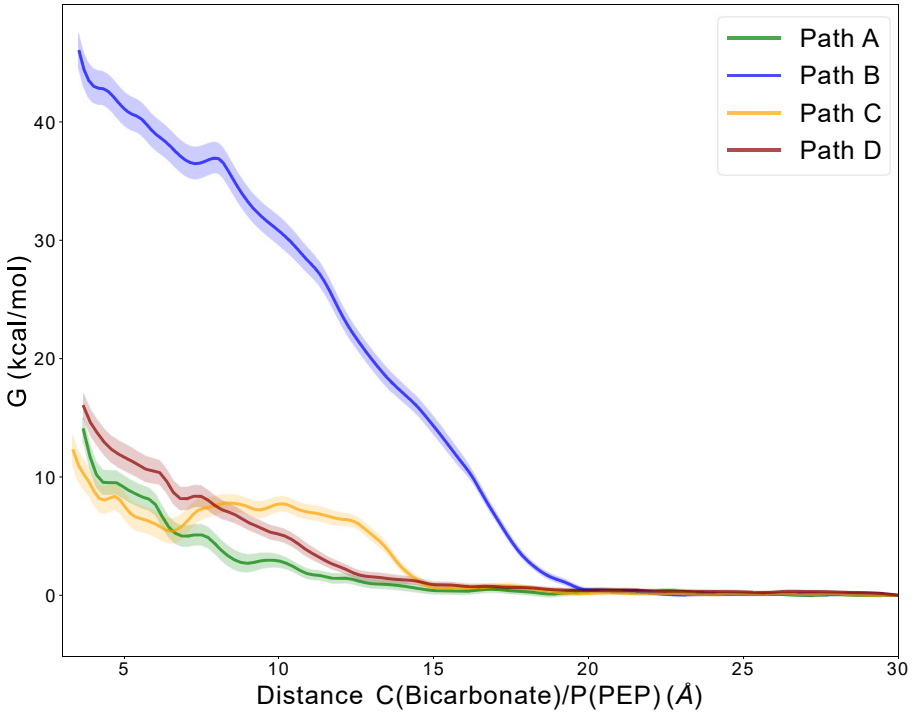
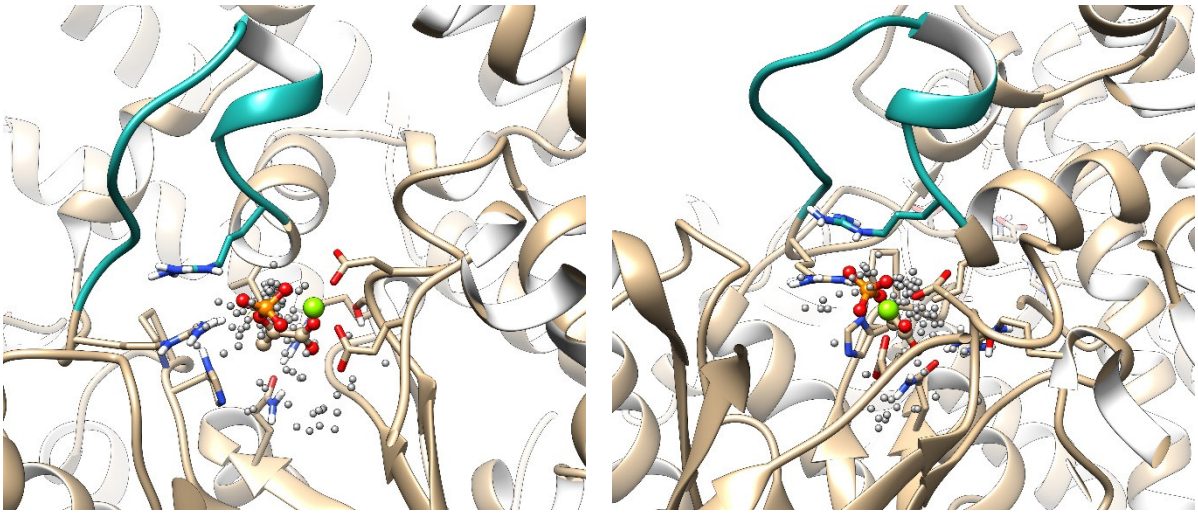


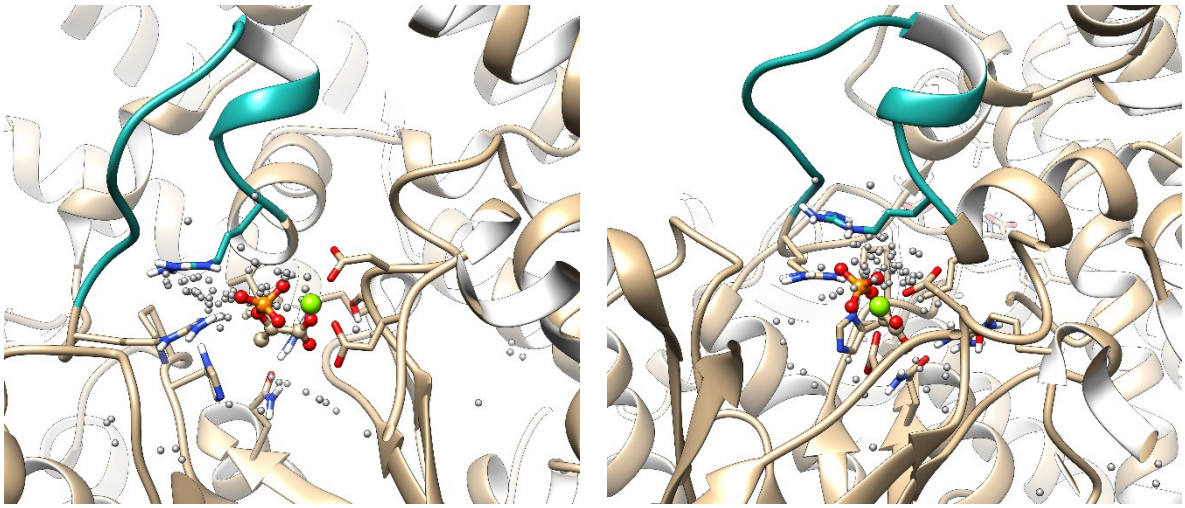
Figure SI-5. Free energy profiles along the four penetration paths (colours match the ones used in Figure SI-4). Error bars were computed with 200 cycles of bootstrapping.

VII. Sampling the active site and finding binding sites of bicarbonate

Following the steered MD simulations, free MD simulations without any constraint were performed to let bicarbonate sample the active site region. Even though some of the steered MD simulations led to non-realistic paths for the penetration, at the end of these simulations the bicarbonate was near PEP in the active site: thus, the ends of these steered MD simulations were still used as starting points for the free sampling. By default, the free MD simulations were of 100 ns long. However, eight of them were extended to 200 ns since the position of bicarbonate changed during the second half (last 50 ns). Moreover, we note that in 10 simulations the bicarbonate left the active site. The position of bicarbonate at the beginning and at the end of the free MD simulations are shown in Figure SI-6.



(a) Positions at the beginning of the free MD simulations (i.e. at the end of the steered MD) with two different orientations.



(b) Positions at the end of the free MD simulations with two different orientations.

Figure SI-6. Position of carbon from bicarbonate at the beginning and at the end of the free MD simulations.

The free MD simulations were analyzed and pockets were identified. Additional images to complement Figure 4 of main text and to show the localization of the identified pockets are provided in Figure SI-7. For each pocket, a representative free MD simulation was analyzed with GROMaps to obtain the density of presence of carbon from bicarbonate; all densities were then aligned and are displayed with a given threshold. Once the pockets were identified, the free MD simulations were analyzed to compute the distance between the bicarbonate and the pockets: first, we clustered one of the simulations to obtain a representative structure of the enzyme denoted `Reference.pdb`, then we post-processed the simulations using Plumed with the procedure described below.

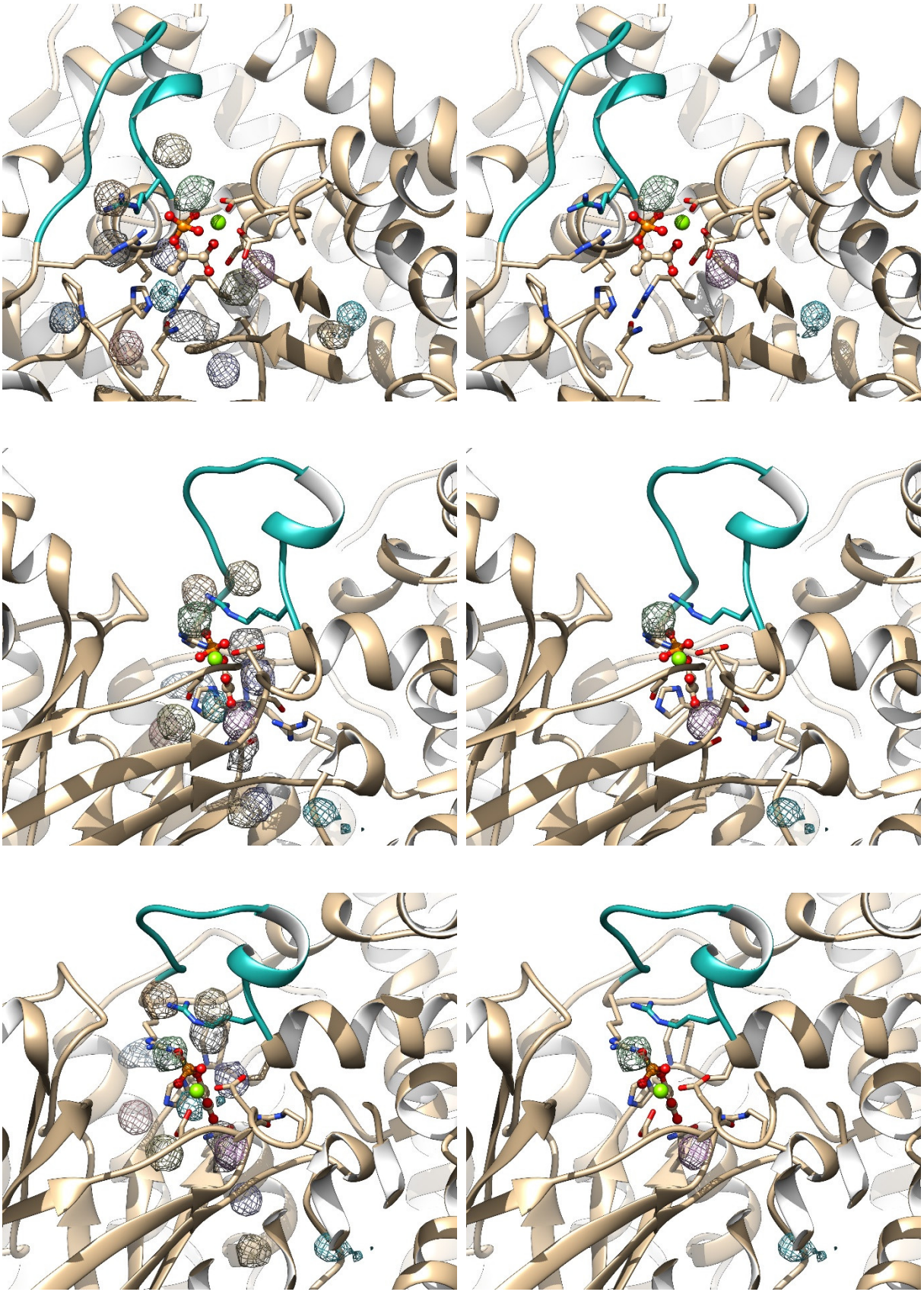


Figure SI-7. Identified pockets of bicarbonate seen under different angles after alignment. On the right, only the three meaningful pockets are displayed. For the sake of clarity, only one monomer is shown and labels of pockets are omitted.

The procedure used to identify and validate the binding sites was the following:

(a) Analysis of the free MD trajectories: the cartesian coordinates of the carbon from bicarbonate in the frame of the reference structure was obtained with the following Plumed commands:

```
FIT_TO_TEMPLATE REFERENCE=Reference.pdb TYPE=OPTIMAL
lig: CENTER ATOMS=29812          #C from bicarbonate
posLig: POSITION ATOM=lig NOPBC
PRINT ARG=posLig.x,posLig.y,posLig.z
FILE=Position_BindingSite.dat
```

(b) Distance verification: the average position of bicarbonate in the frame of `Reference.pdb` was obtained by averaging the coordinates on the equilibrated part of the simulations. The distance between bicarbonate and the averaged center of the pocket was then extracted:

```
FIT_TO_TEMPLATE REFERENCE=Reference.pdb TYPE=OPTIMAL
lig: CENTER ATOMS=29812          #C from bicarbonate
site: FIXEDATOM AT=9.41308,8.75443,3.70346
d: DISTANCE ATOMS=lig,site
PRINT ARG=d FILE=Distance_Site-Lig.dat STRIDE=1
```

The cartesian coordinates of the pockets in the frame of `Reference.pdb` are given in Table SI-2. On average, the distance between two pockets is 11.2 Å with distances ranging from 3.6 Å to 24.3 Å.

Pocket	x (Å)	y (Å)	z (Å)	Pocket	x (Å)	y (Å)	z (Å)
P1	94.1308	87.5443	37.0346	**P8**	93.8116	85.8082	31.8510
P2	92.1831	79.5831	33.8611	**P9**	97.4579	93.6789	36.0510
P3	99.5197	86.3122	38.2648	**P10**	103.619	81.8897	36.8529
P4	96.2844	84.0021	39.2230	**P11**	100.493	86.4109	41.7633
P5	100.945	79.2288	40.5813	**P12**	106.916	77.2117	45.2089
P6	102.354	91.4787	36.5804	**P13**	103.479	74.5631	49.7733
P7	96.8710	80.6208	32.9977	**P14**	103.665	85.4238	45.3082

Table SI-2.Cartesian coordinates of pockets in the frame of the reference structure.

The computed distances for 14 trajectories (one for each pocket) are displayed in Figure SI-8: we can see that at the end, bicarbonate stayed at less than 2 Å from the center of pockets, except for the simulation of pocket P3 for which the last 10 ns were omitted when conformations were extracted.

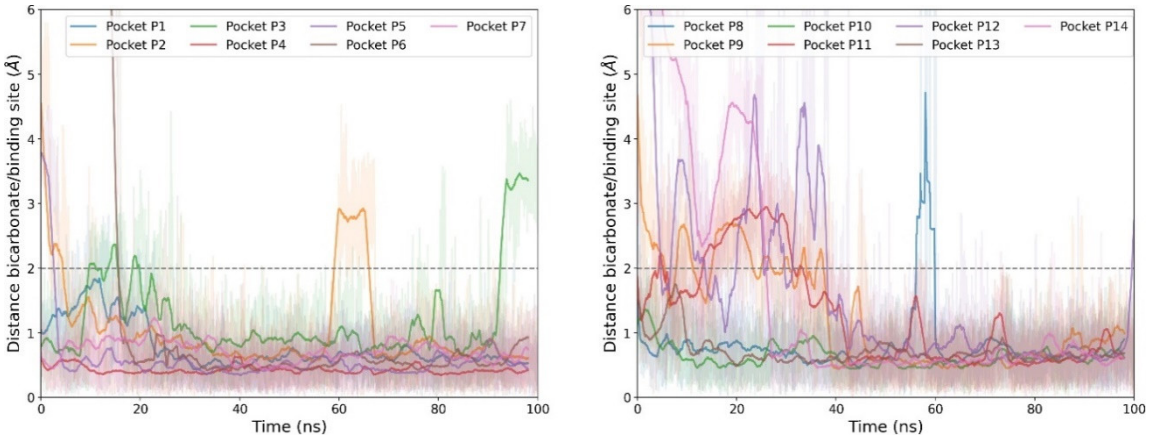


Figure SI-8. Distance between bicarbonate and the center of pockets during the free MD simulations. For each pocket, one representative simulation was chosen. For simulations of pockets P3 and P13 which were extended to 200 ns, the first 100 ns are omitted and only the last 100 ns are displayed. Pale colors: raw data; full colors: running average over 2ns. For the sake of clarity, data are separated in two panels with seven pockets per panel.

In Figure SI-9, we present the distance between bicarbonate and the center of all pockets in three representative simulations where bicarbonate ended in pockets P5, P7 and P13 respectively. No constraints were applied during these free MD simulations. We can see that pockets are well separated, i.e. for each simulation bicarbonate is at less than 2 Å from only one pocket.

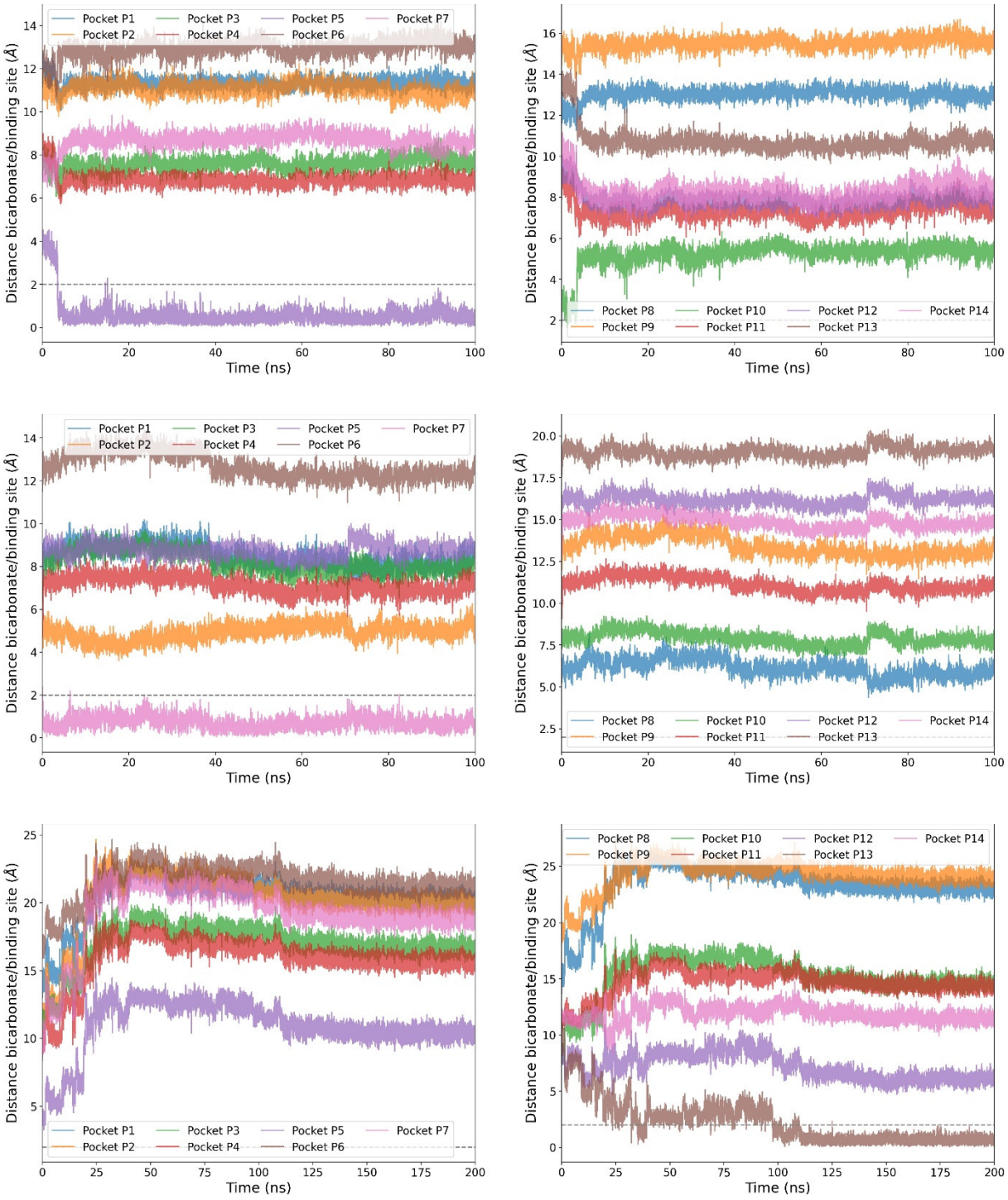


Figure SI-9. Distance between bicarbonate and the center of each pocket for three representative simulations where bicarbonate ended in pockets P5, P7 and P13 respectively. For the sake of clarity, each set of data is shown in two panels with seven pockets per panel.

VIII. Pulling bicarbonate outside the protein

To compare the different pockets, we first tested a physical approach i.e. we have computed the binding free energy along a physical reaction coordinate. Thus, we performed steered MD simulations to pull bicarbonate outside the protein. These simulations started from the last conformation of one of the free simulations where bicarbonate ended up in pocket P1. Three reaction coordinates were compared, other things remaining equal: (i) the distance between C of bicarbonate and the center of mass of the protein, (ii) the distance between C of bicarbonate and the center of mass of proline 178 (which is located at the bottom of the active site), (iii) the distance between C of bicarbonate and N from proline 178. Umbrella sampling simulations were then performed as previously: (a) snapshots from the steered MD simulations were randomly extracted every 0.5 Å along the reaction coordinate, (b) each snapshot was equilibrated 500 ps, (c) 2 ns of production with a force constant of k=1000 kJ/mol/nm^2 were performed. Free energy profiles are shown in Figure SI-10 with an illustration of the localisation of proline 178.

We observed very different results depending on the reaction coordinate, even though all simulations started from the same checkpoint file. With two reaction coordinates, pulling the bicarbonate outside PEPC is exergonic (i.e. $\Delta G_{Binding}>0$). For the third reaction coordinate (between C from bicarbonate and N from proline 178), we observed a more realistic free energy profile with a well-defined energy well. We envisioned that such a dependence on the reaction coordinate would make the convergence of the simulations a tedious task, prone to errors. Moreover, the choice of the reaction coordinate for pulling bicarbonate outside of each pocket would not be unique. Thus, we didn't investigate further the physical approach to compare the pockets and we relied on alchemical transformations.

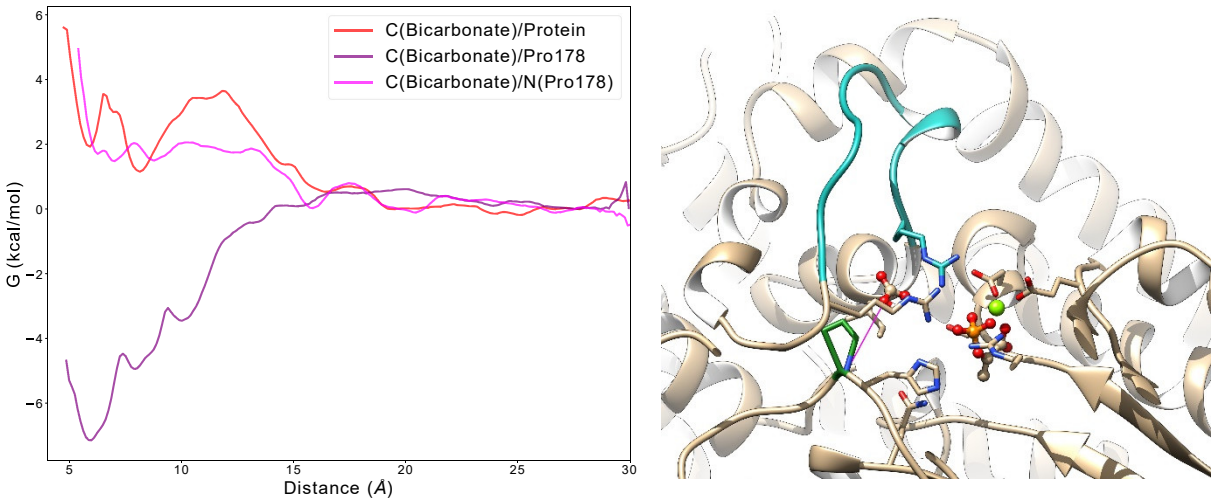


Figure SI-10. Left: Free energy profiles for pulling the bicarbonate outside PEPC. The same starting point was used for the three profiles and only the reaction coordinate changed between the simulations. Right: Active site with a highlight of proline 178 in green. The axis between C of bicarbonate and N from proline 178 is in magenta.

IX. Extracting $\Delta G_{Binding}$ from K_M

In a Michaelis-Menten scheme (see Figure SI-11), the Michaelis constant is defined as $K_M = \frac{k_{-1}+k_2}{k_1 C^0}$. Moreover, the dissociation constant is defined as $K_d = \frac{[E][S]}{[ES]C^0} = \frac{k_{-1}}{k_1 C^0}$.

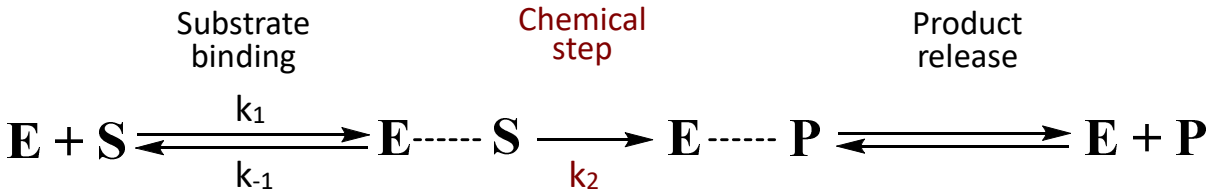


Figure SI-11. Michaelis-Menten scheme (E=enzyme, S=substrate, P=product).

Thus, if $k_2 \ll k_{-1}$ we can state that $K_d \approx K_M$ and we can then extract the binding free energy from the K_M measurements:

$$\Delta G_{Binding} = -RT \cdot ln(K_a) = -RT \cdot ln(1/K_d) \approx -RT \cdot ln(1/K_M)$$

When expressed in terms of free energies (see Figure SI-12), comparing k_{-1} and k_2 means comparing the barriers between the dissociative path and the reactive path starting from the Michaelis-Menten complex (i.e. the $\Delta\Delta G$ represented in red in Figure SI-12, starting from the red dot). If $\Delta\Delta G$ is 2.5 kcal/mol, then $k_{-1}/k_2 = 100$ and the approximation $k_2 \ll k_{-1}$ is already valid. If $\Delta\Delta G$=4.1 kcal/mol, then $k_{-1}/k_2 = 1000$.

We don't have access to the value of $\Delta\Delta G$ for PEP Carboxylase, however k_2 was measured for *E. coli* PEPC and can be converted to $\Delta G^{\ddagger}$=14.6 kcal/mol and it is really unlikely that ΔG_{-1} is below 2.5 kcal/mol from this value[1] (i.e. it is unlikely that ΔG_{-1} is higher than 12.1 kcal/mol).

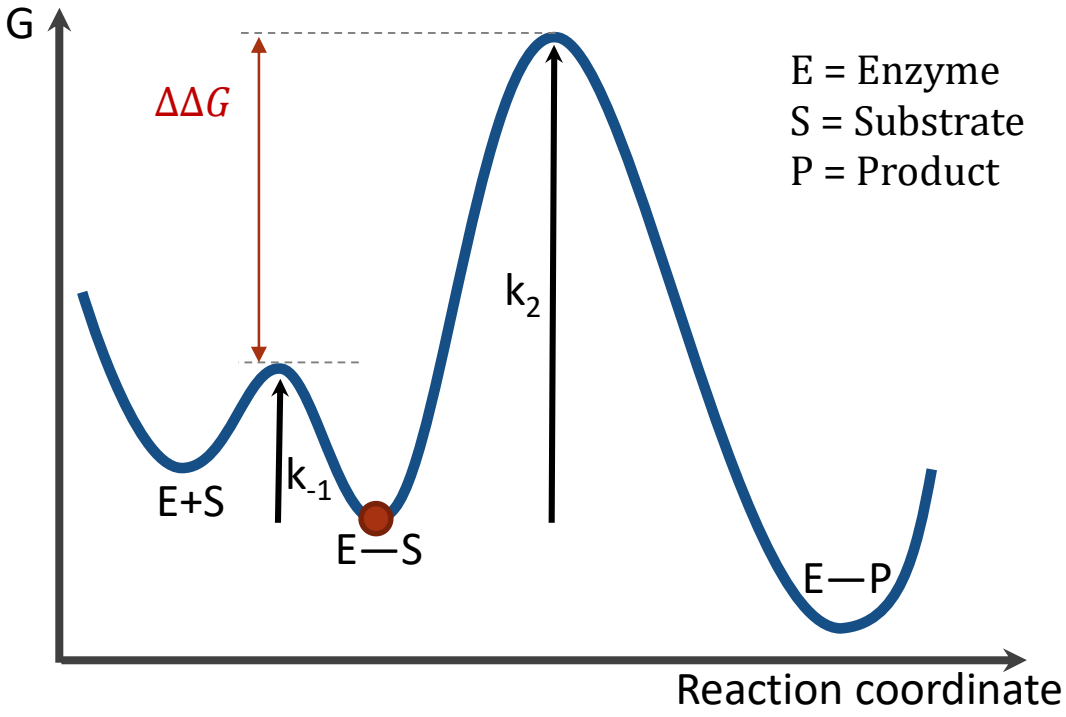


Figure SI-12. Scheme of the free energy profile along a biochemical reaction. The Michaelis-Menten complex is represented by the red dot.

X. Structural fluctuations around pockets P5 and P7

We provide below an analysis of the protein structure fluctuations around pockets P5 and P7. The vicinity of the pockets was defined as all residues within 7 Å of PEP or magnesium(II) atoms in the reference snapshot; this led to 33 residues. We analyzed the fluctuations of these residues along the 1 µs-long free MD simulations (we recall that for each pocket, two 1 µs-long simulations were performed). In Figure SI-13, we present the RMSD with respect to the first frame as well as the number of hydrogen bonds. We note that the same behavior is observed in the two simulations.

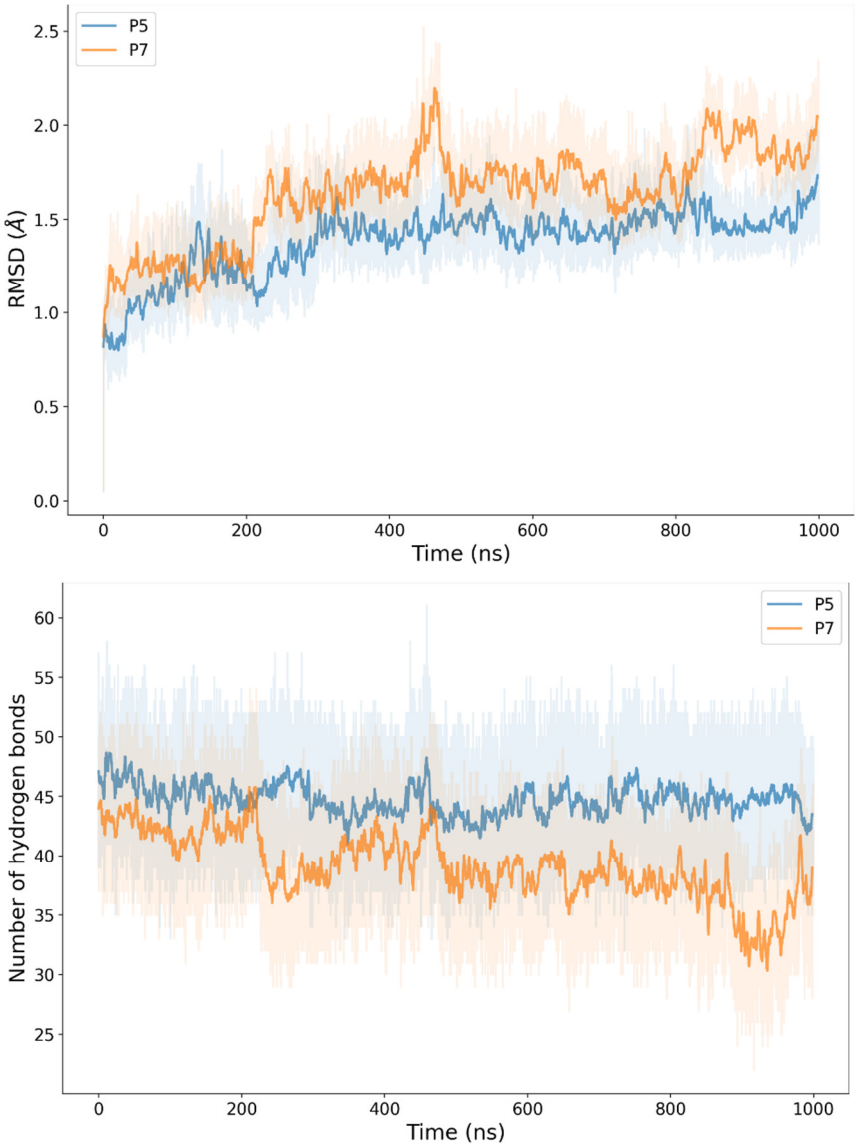


Figure SI-13. (Top) RMSD of the 33 residues around PEP and Mg^{2+} with respect to the first frame. (Bottom) Number of hydrogen bonds between the 33 residues and the remaining of the system. Pale colors: raw data; full colors: running average over 2ns. For the sake of clarity, only one simulation per pocket is displayed.

In Figure SI-14, we compare the RMSF of the residues in simulations with bicarbonate in pockets P5 and P7. For comparisons, we present in Figure SI-15 the RMSF for the two simulations in pocket P5 and in Figure SI-16 the RMSF for the two simulations in pocket P7. In these three figures, on the right panel we plot the RMSF of residues in one simulation with respect to the other, and residues around PEP and magnesium are highlighted in pink. We can clearly see that the differences observed between pockets in Figure SI-14 are comparable to the ones observe within each pocket in Figure SI-15 and in Figure SI-16. Thus, we can conclude that fluctuations in each pocket are similar and can't be distinguished.

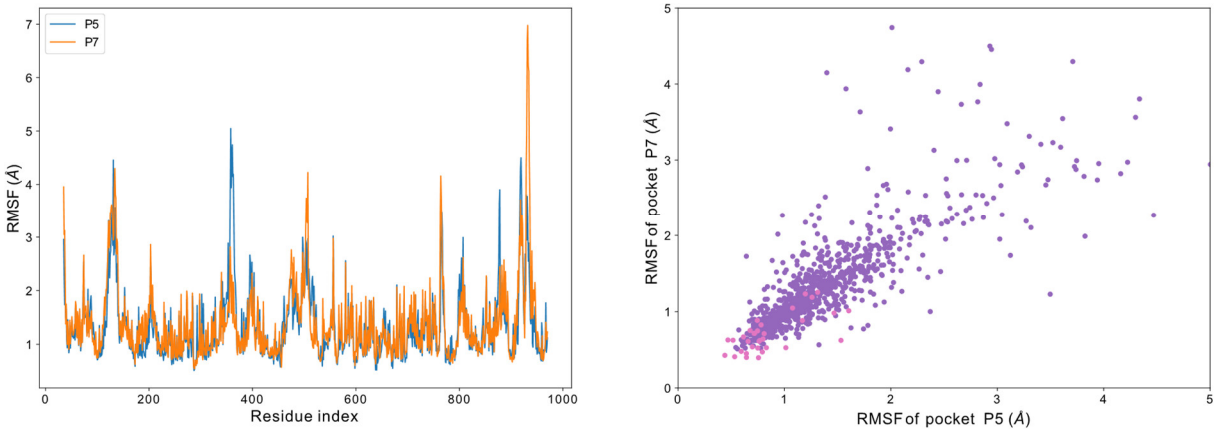


Figure SI-14. (Left) RMSF of the protein for 1 µs-long simulations with bicarbonate in pockets P5 and P7. (Right) Scatter plot of the RMSF presented on the left to compare them. In pink are highlighted the 33 residues around PEP and Mg^{2+}.

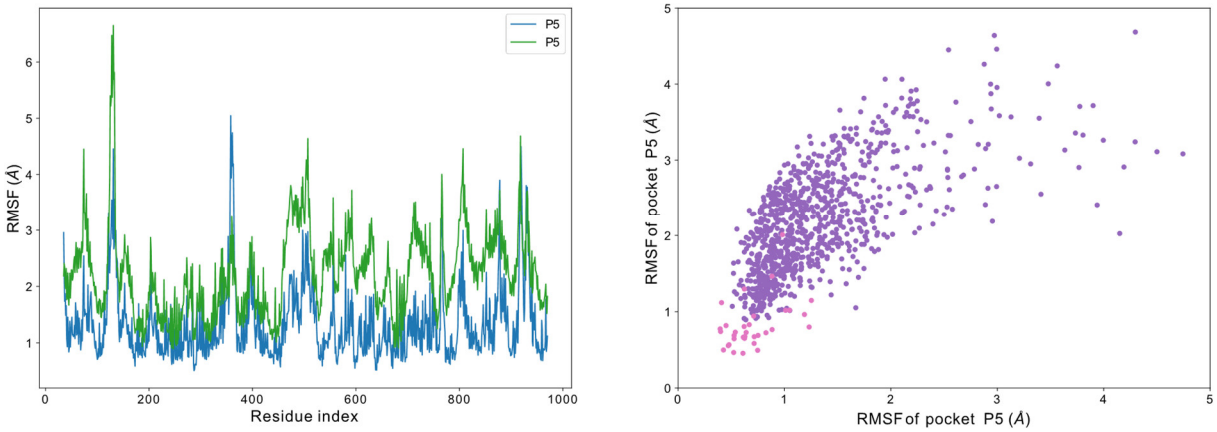


Figure SI-15. (Left) RMSF of the protein for the two 1 µs-long simulations where bicarbonate was in pocket P5. Data in blue are the same as the one in Figure SI-14. (Right) Scatter plot of the RMSF presented on the left to compare them. In pink are highlighted the 33 residues around PEP and Mg^{2+}.

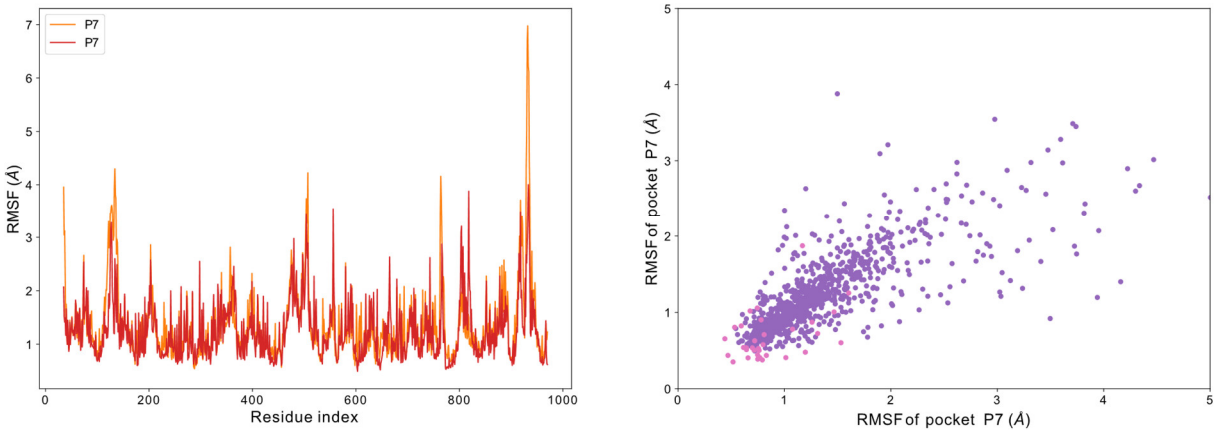


Figure SI-16. (Left) RMSF of the protein for the two 1 µs-long simulations where bicarbonate was in pocket P5. Data in orange are the same as the one in Figure SI-14. (Right) Scatter plot of the RMSF presented on the left to compare them. In pink are highlighted the 33 residues around PEP and Mg^{2+}.

X. Free Energy Perturbations

1. Main principles

Alchemical free energy calculations have been extensively reviewed (see Ref. [30–36] for ex.). We very briefly summarize the basic principles below to guide us in describing the computational details.

When one is interested in computing the binding free energy between a solvated protein and a solvated ligand, it is possible to decompose this value in several free energy contributions extracted from a thermodynamic cycle such as the one presented Figure SI-17 and described below step by step:

1. One can first compute the desolvation free energy to transfer the ligand from solvent to gas phase. This is done by switching off all intermolecular interactions between the ligand and its environment in the simulation box. This step needs simulations, and $\Delta G_{Desolvation}$ is positive.
2. Once the ligand interacts with nothing, one can then add restraints to restrict the space it can sample. $\Delta G_{Add\ restraints}$ is obtained with an analytical function and is positive.
3. Since the ligand makes no intermolecular interactions, it can be transferred at no cost between two environments as long as the restraints are equivalent. Thus, $\Delta G_{Ligand\ insertion} = 0$.
4. One can then compute the binding free energy with restraints between the ligand and the protein. This is usually done in the reverse way by starting from the complex and switching off all intermolecular interactions between the ligand and its environment. This step needs simulations, and $\Delta G_{Binding\ with\ restraints}$ is negative for good ligands in good pockets.
5. Finally, one needs to compute the gain in removing the restraints. For good ligands in a good pocket, $\Delta G_{Remove\ restraints}$ is often equal to 0 because they don't need restraints to be kept in place. Nonetheless, this value can also be computed with simulations.

All terms will be described below with technical details. Since $\Delta G_{Ligand\ insertion} = 0$, one can write:

$$\Delta G_{Binding,\ FEP} = \Delta G_{Desolvation} + \Delta G_{Add\ restraints} + \Delta G_{Binding\ with\ restraints} + \Delta G_{Remove\ restraints}$$

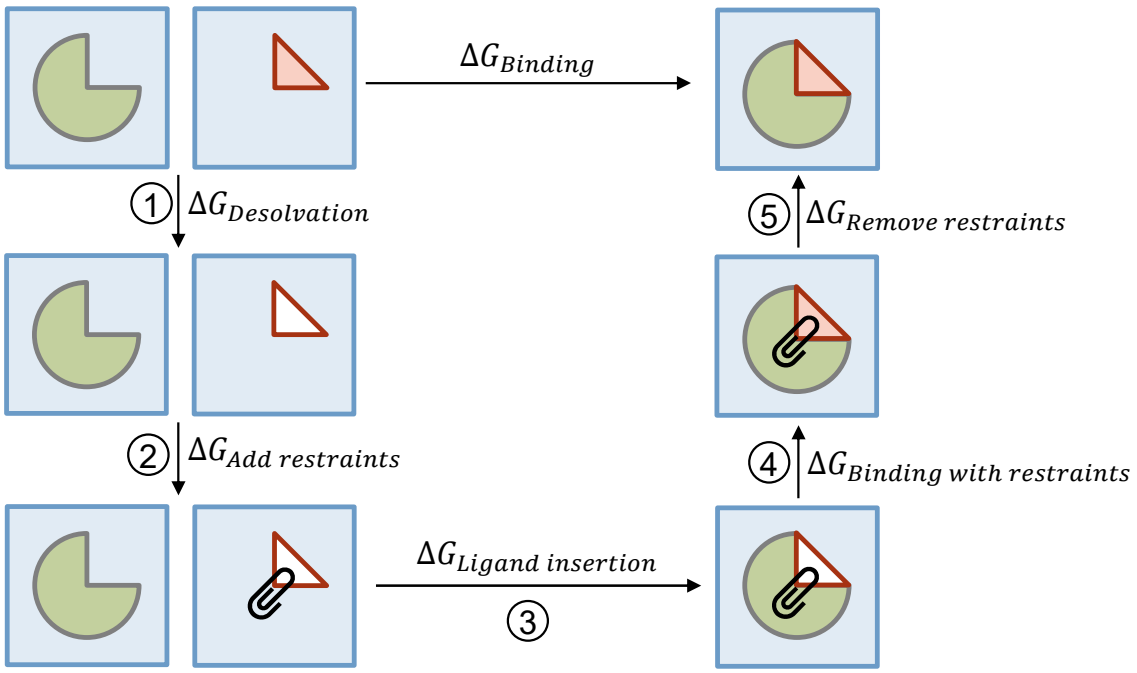


Figure SI-17. Scheme to explain the principle of computing binding free energy with alchemical transformations and FEP. The protein is in green and the ligand is in red; the blue background represents the solvent. When the ligand is not making any interaction with its environment, it is represented with no filling. The restraints are schematized with a staple.

2. Desolvation free energy and binding free energy with restraints

Steps 1 and 4 described in Figure SI-17 are equivalent in their computational implementations. For both steps, we have used soft-core potentials to decouple Lennard-Jones interactions (with standard parameters, i.e. with the classical notations α=0.5, σ=0.3 nm and a power of 1). 21 alchemical windows were used: we have first decoupled the electrostatic interactions by steps of 0.1 (0.0, 0.1, 0.2, …, 0.9, 1.0) and then decoupled the Lennard-Jones interactions also by steps of 0.1. The difference of potential between the running window and all other alchemical windows were recorded every 500 fs, as well as the derivative of the potential with respect to λ (to compare MBAR and TI). All simulations described below were ran with stochastic dynamics, with a length for each window that will be described later.

For the desolvation of bicarbonate, first we ran 500 ps of equilibration with simulated annealing of the bicarbonate in a cubic box of 4.3 nm of edges (corresponding to at least 2.0 nm between the solute and the edges of the box). We then ran 1 ns of simulations for extra equilibration. All 21 windows were then ran in parallel for 20 ns, using the same starting checkpoint. The last 15 ns were analyzed.

For the binding free energy with restraints, we extracted snapshots of the complex from the end of the free MD trajectories. For each pocket, one to six conformations were extracted every 10 ns (i.e. for example at 100 ns, 90 ns and 80 ns of the simulation). The systems were then reconstructed as previously described (i.e. box generation and solvation): the reason for doing so is that when the initial boxes were generated, they may have been quite large since bicarbonate was outside the protein (sometimes at 56 Å from PEP); creating a new box allows to save computational time, for example for one system the number of atoms decreased from ~263 000 to ~244 000. We then applied the same procedure as before: (i) energy minimization, (ii) equilibration with simulated annealing from 100 K to 300 K in 400 ps and then an extra 100 ps at 300 K, (iii) further equilibration of 2 ns under NPT. We then ran the 21 windows of alchemical transformations in parallel, all using the same starting checkpoint.

3. Adding restraints

When the ligand is fully decoupled from the protein (i.e. with no intermolecular interactions), it may wander in the simulation box if there are no restraints. However, it must explore only configurations in the binding site and must not unbind from the protein if we want to extract a meaningful binding free energy. Thus, adding restraints during step 4 is mandatory. $\Delta G_{Add\ restraints}$ is the cost to restrict the configurational space of the ligand and can be written as:

$$\Delta G_{Add\ restraints} = -RTln\left(\frac{V_L}{V^0}\right) - RTln\left(\frac{\xi_L}{8\pi^2}\right)$$

where the first term is for the restraints on translations and the second term is for the restraints on rotations. V_L is the volume occupied by the ligand under the restraints (without restraints, $V_L = V_{Box}$), ξ_L is the volume of rotational degrees of freedom and $V^0 = 1/C^0$ is the standard volume (here we have used C^0 = 1 mol/L = 1 molecule/1660 Å^3). For large ligands, the main used set of restraints are the ones derived by Boresh *et al.* where 1 distance, 2 angles and 3 dihedrals between ligand and protein are restrained[37]. The advantage of this set is that all these functions are integrable and an analytical value can thus be obtained. In the current case, the Boresh restraints cannot be used because bicarbonate is a small ligand which must be free to rotate. We have thus used flat-bottomed potential to force the ligand to stay in each identified pocket, and no rotational restraints were applied. In the following, we define r as the distance between the bicarbonate and the center of the pocket:

- If r was below r_0, no restraint was applied.
- If r was above r_0, the bicarbonate felt a harmonic potential $E = \frac{1}{2}k(r - r_0)^2$.

This function is integrable and we can write the volume occupied by the ligand under the restraints as:

$$V_L = \int_0^{\infty} e^{-\frac{E(r)}{RT}} 4\pi r^2 dr = \frac{4}{3}\pi r_0^3 + 2\left(\frac{2\pi RT}{k}\right)^{3/2} + 4r_0\frac{2\pi RT}{k} + 4\pi r_0^2\left(\frac{2\pi RT}{k}\right)^{1/2}$$

No restraints are applied on the rotations (i.e. $\xi_L = 8\pi^2$), and $\Delta G_{Add\ restraints}$ can then be computed and equals to 2.222 kcal/mol with $r_0 = 0.2\ nm$ and $k = 100\ 000\ kJ/mol/nm^2$.

To apply the restraints, we relied on Plumed. Following the procedure described above to identify the positions of pockets, we applied restraints by fitting on-the-fly during the FEP simulations the enzyme to the reference structure and by computing the distance r (this slowed down simulations by ~40%):

```
FIT_TO_TEMPLATE REFERENCE=Reference.pdb TYPE=OPTIMAL
lig: CENTER ATOMS=29812          #C from bicarbonate
site: FIXEDATOM AT=9.41308,8.75443,3.70346
d: DISTANCE ATOMS=lig,site
uwall: UPPER_WALLS ARG=d AT=0.2 KAPPA=100000 OFFSET=0.0
```

The choice of r_0 at 2 Å came from the analysis described in Part VII. We have shown in Figure SI-8 the distance between bicarbonate and the center of pockets in representative simulations for each pocket, and in Figure SI-9 the distance between bicarbonate and the center of all pockets in three simulations. Once the position of the ligand was equilibrated, we observed that it stayed at less than 2 Å from the center of the pocket. Moreover, when the pockets are visualized as in Figure 4 of main text or in Figure SI-7, we have observed that if the cutoffs for the densities are set in such a way that the volumes displayed are the same as those of spheres of radius 2.5 Å, some pockets overlapped (see Figure SI-18). This barely happens for volumes corresponding to spheres of lower radius such as 2 Å. This observation is in line with the minimal distance between the center of two pockets of 3.6 Å given in Table SI-2. Since we aimed at well separated pockets, we set r_0 at 2 Å.

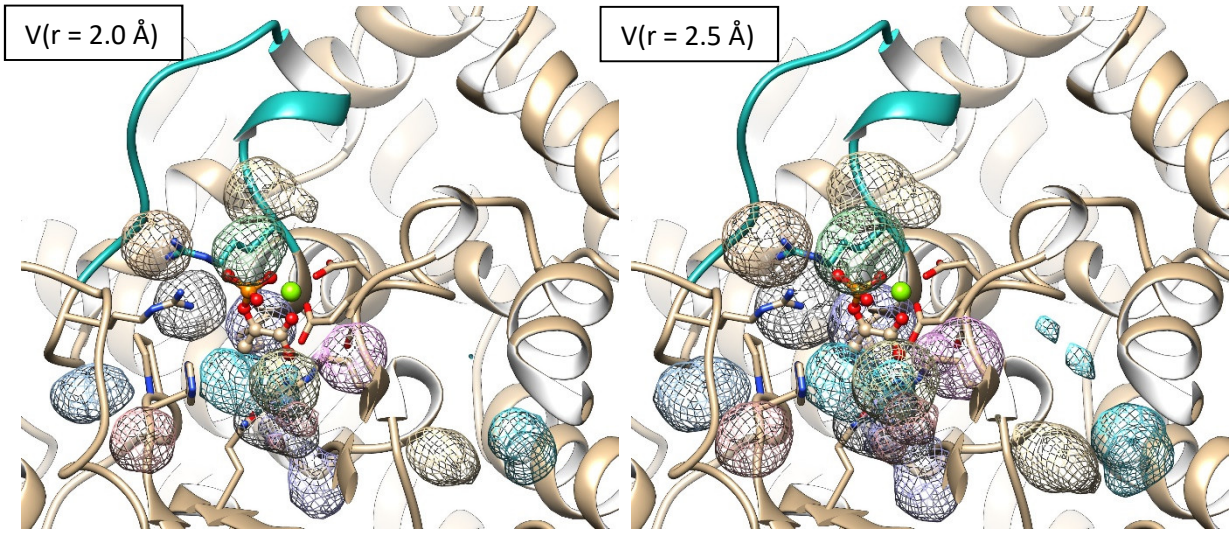


Figure SI-18. Identified pockets for bicarbonate (same orientation as Figure 4 of main text). Left: each pocket is represented with the volume of a sphere of radius 2.0 Å. Right: each pocket is represented with the volume of a sphere of radius 2.5 Å.

4. Removing restraints

In most studies using FEP for computing protein-ligand interactions, the term $\Delta G_{Remove\ restraints}$ equals to 0. Indeed, this term represents the contribution of removing the restraints between the ligand and the protein when these two partners are interacting. Since good ligands don't unbind from good pockets, there is no energy cost for restraining the ligand. However, we are not in this situation since some of the pockets we have identified are not good pockets, and the bicarbonate can unbind from them. We have thus derived a protocol to compute the term $\Delta G_{Remove\ restraints}$.

In the MBAR equations that are used to analyze the simulations, what is needed is the difference of potential energy between windows i and j. In other words, one needs to perform the simulations under state i and calculate the potential energy of each configuration if it were in state j (and do this for all possible states j). When one is removing the restraints for a fully coupled system, the only difference of potential energy between two states is the restraining potential (flat-bottomed harmonic in our case). Thus, for each binding free energy calculation in each pocket, we have performed three additional simulations of 10 ns: one with 100% of the restraining force constant, one with 50% of the restraining force constant and one with 0% of the restraining force constant. The distance between bicarbonate and the center of the pocket was recorded, and we have then computed ΔU_{ij} between each pair of states with a bash script that read the distances and then wrote the differences of flat-bottomed potential energies in each window. These new files were then directly feed to alchemlyb.

Unless for a few "not-good" pockets, the MBAR overlap matrix between windows were good to perfect (from 10 to 33% overlap). For two cases, we have compared the use of three and five windows (in the latter case, the restraining force constants were set at 100%, 75%, 50%, 25% and 0%); the differences in $\Delta G_{Remove\ restraints}$ were respectively $0.17 \cdot 10^{-3}$ and $7.3 \cdot 10^{-3}$ kcal/mol, which validates the use of only three windows. MBAR data are given in Table SI-5, and we add that the MBAR statistical uncertainties were all below 0.007 kcal/mol. We have observed that for some good pockets (i.e. with a high absolute value of $\Delta G_{Binding\ with\ restraints}$), the value for removing the restraints was exactly 0 or close to it ($4 \cdot 10^{-7}$ kcal/mol for example). However, for some "not-good" pockets, we have computed a value for removing the restraints of up to 2.6 kcal/mol. This workflow can't be used with TI equations, but this is not important since for meaningful pockets the cost for removing restraints is always 0.

5. Convergence

Analysis of the data from alchemical free energy perturbations were made with the python library alchemlyb v0.7.0 which uses the module pymbar v3.1.1 to solve the MBAR equations[38]. Convergence was first assessed with MBAR, before comparing with TI. Once a given set of simulations with all alchemical windows were fed to alchemlyb through a Jupyter notebook, we checked the convergence with a forward/backward analysis: in such a procedure, different fractions of the data are analyzed both in the forward and in the backward directions. All the analyzed fractions provide an estimate for the free energy (see Figure SI-19). This set of data is considered as converged if the two estimations at 50% of the simulation time differs by less than 1 k_BT[35]. In other words, if one analyses the first half or the second half of the simulation, the same free energy is obtained within 0.6 kcal/mol.

In our protocol (see Figure SI-2), all alchemical simulations are started in parallel from the same equilibrated structure. Thus, for the fully uncoupled simulations, a void is abruptly created and an extra equilibration time is needed. Issues with the functions to detect equilibration time in alchemlyb (see further) led us to use another protocol to know how much time must be discarded: we analyzed the simulation in the backward direction (*slicing* function) and aimed at the longest simulation time that provides an error below 1 k_BT between the first half and the second half of the subsampled simulations.

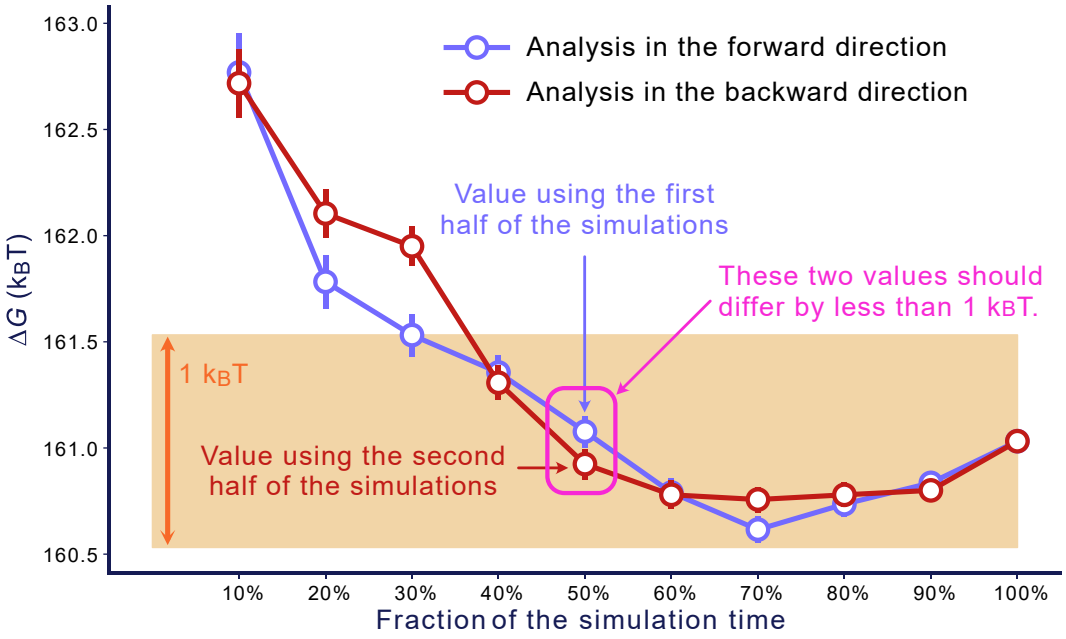


Figure SI-19. Forward and backward analysis of a given set of simulations. The orange area represents the value at 100% of the simulation time +/- 0.5 k_BT (we underline that it is not the default graphical representation of alchemlyb).

All FEP simulations were first of 15 ns, and were progressively extended by 5 ns until the convergence criteria were met (we add here that we set an additional constraint of having at least 10 ns of data). The used length of each simulation and the amount of data that were used are presented in Table SI-3 (10 to 23 ns were used, i.e. 10000 to 23000 data points per window). We present in Figure SI-20 representative data where the errors between the forward and the backward analysis at 50% of a given time bloc of the simulation are represented for different time blocs.

(a) When the simulation length was 15 ns, all the errors at 50% were above 1 k_BT (except when we analyze the last ns, but this is a too small simulation time).
(b) When the simulation length was extended to 20 ns, the error is below the threshold for 7 time blocs. The first three are at the very beginning of the simulations and are not meaningful. The last four are blocs 14-20 ns, 13-20 ns, 12-20 ns and 11-20 ns. They represent 6 to 9 ns of simulation time, which is less than the 10 ns of data that we imposed to ourselves.
(c) When the simulation length was extended to 25 ns, the error is below the threshold for the very first ns, and then for 9 times blocs (19-25 ns to 11-25 ns). The last time bloc provides the longest simulation time (14 ns) and was kept to provide the binding free energy.

We acknowledge that the protocol isn't (yet) optimal in terms of resources: we extended all windows whereas it is possible that only a single window needed to be extended. Work is under progress to develop a more efficient protocol that uses less computational time, and will be reported elsewhere.

Once the appropriate simulation length was identified, we analyzed the phase space overlap between each window from the MBAR analysis of alchemlyb. A representative matrix is provided in Figure SI-21, where each number off-diagonal is the probability of overlap between window i and window j. The diagonal numbers are the self-overlaps of windows. The proposed empirical rule is that "the values of the first off-diagonals (i.e. the diagonals above and below the main diagonal) should at least be 0.03 to obtain a reliable free energy estimate"[35,39]. The minimal values, maximal values and average values that were obtained in the diagonal and in the first off-diagonal of all overlap matrixes are provided in Table SI-3. We have observed that the off-diagonal values were on average 0.23, ranging from 0.07 to 0.33. For the diagonal, the average value was 0.40, ranging from 0.19 to 0.80. Thus, all simulations presented here have a sufficient overlap between all neighboring windows.

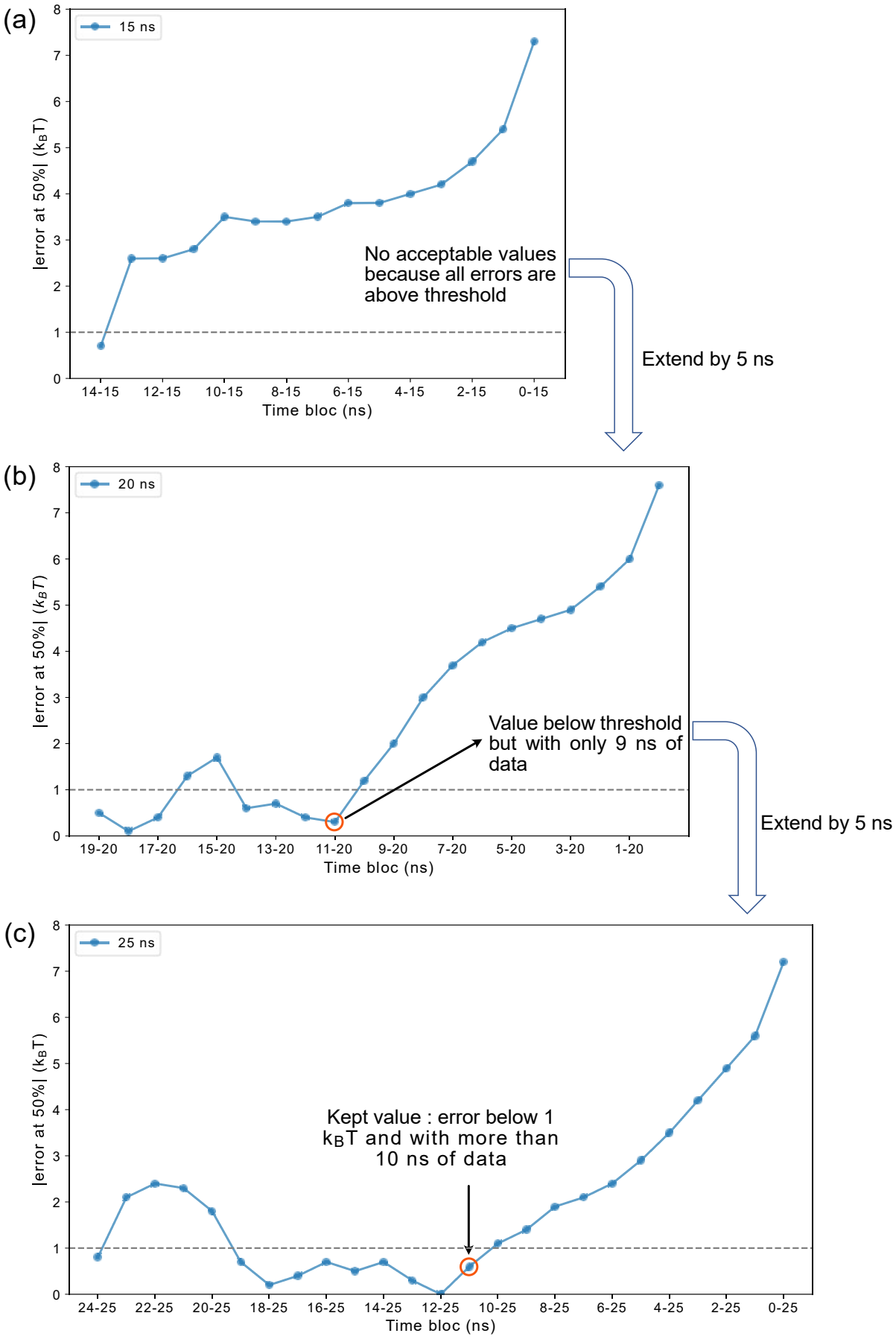


Figure SI-20. Representative data to assess the needed simulation length to reach convergence.

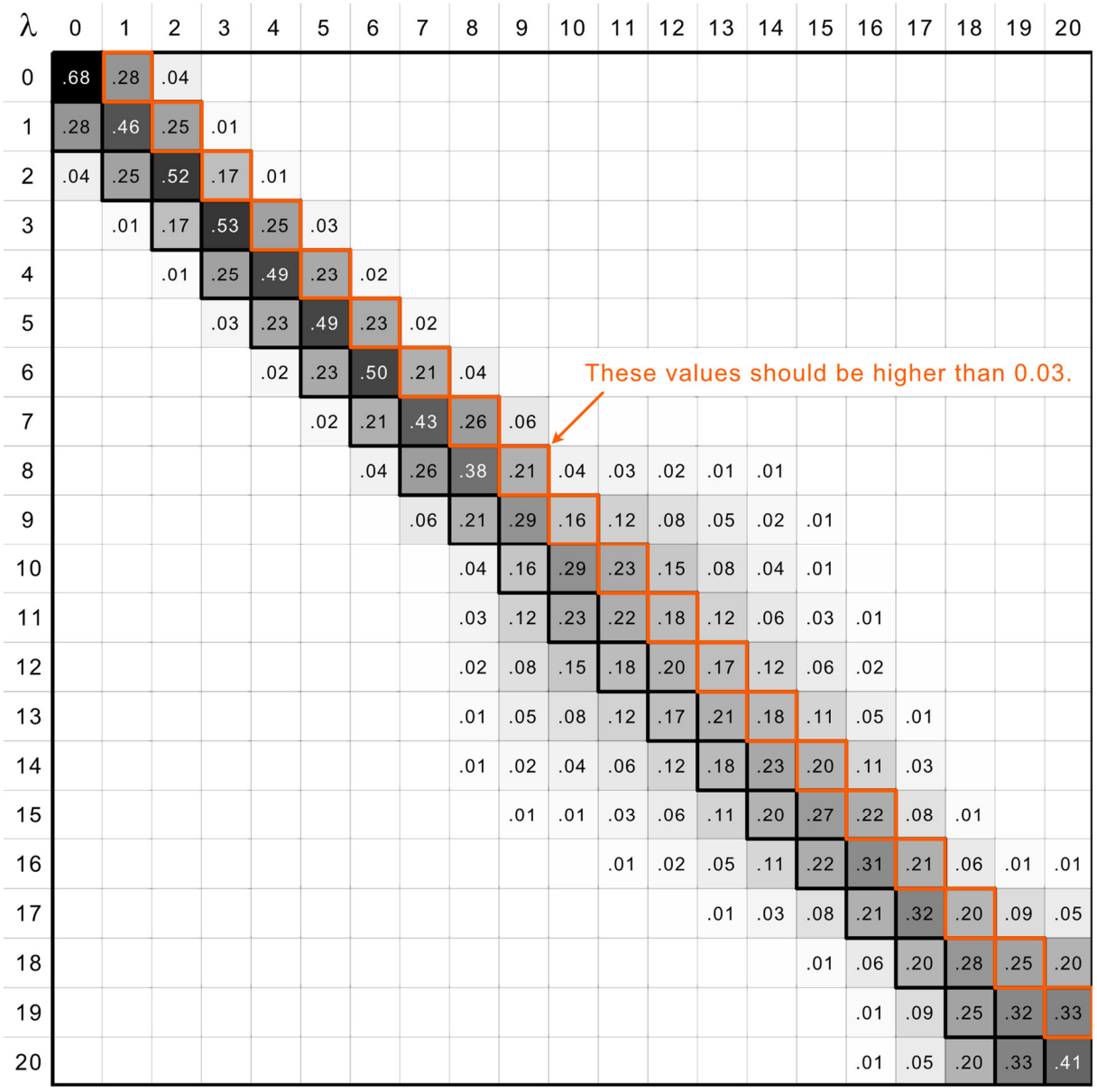


Figure SI-21. Representative overlap matrix from an MBAR analysis.

Pocket	Snapshot	Time bloc used for analysis	Length of data used	*Overlap matrix* Min/Max diagonal	Average diagonal	Min/Max first off-diagonal	Average first off-diagonal
	100 ns	1-20 ns	19 ns	0.24 / 0.69	0.38	0.18 / 0.32	0.23
P1	90 ns	1-20 ns	19 ns	0.22 / 0.70	0.38	0.15 / 0.31	0.23
	80 ns	4-20 ns	16 ns	0.20 / 0.74	0.37	0.17 / 0.33	0.22
	100 ns	3-20 ns	17 ns	0.23 / 0.74	0.40	0.17 / 0.33	0.23
P2	90 ns	6-25 ns	19 ns	0.22 / 0.71	0.39	0.17 / 0.32	0.22
	80 ns	1-20 ns	19 ns	0.23 / 0.72	0.40	0.17 / 0.33	0.23
	190 ns	11-25 ns	14 ns	0.25 / 0.73	0.41	0.17 / 0.32	0.22
P3	180 ns	13-25 ns	12 ns	0.24 / 0.75	0.41	0.17 / 0.32	0.23
	170 ns	4-25 ns	21 ns	0.23 / 0.72	0.39	0.16 / 0.33	0.23
	100 ns	22-40 ns	18 ns	0.20 / 0.63	0.36	0.15 / 0.33	0.22
P4	90 ns	5-20 ns	15 ns	0.27 / 0.70	0.39	0.18 / 0.31	0.23
	80 ns	9-25 ns	16 ns	0.29 / 0.68	0.40	0.16 / 0.32	0.23
	100 ns	13-25 ns	12 ns	0.29 / 0.72	0.43	0.07 / 0.32	0.23
P5	90 ns	10-30 ns	20 ns	0.25 / 0.74	0.43	0.13 / 0.31	0.22
	80 ns	8-30 ns	22 ns	0.26 / 0.66	0.42	0.08 / 0.33	0.23
	70 ns	17-30 ns	13 ns	0.25 / 0.67	0.40	0.08 / 0.33	0.23
	60 ns	11-30 ns	19 ns	0.26 / 0.66	0.42	0.08 / 0.33	0.23
	100 ns	10-25 ns	15 ns	0.25 / 0.68	0.39	0.17 / 0.33	0.23
P6	90 ns	3-20 ns	17 ns	0.24 / 0.67	0.38	0.18 / 0.32	0.23
	80 ns	7-30 ns	23 ns	0.28 / 0.68	0.40	0.17 / 0.31	0.24
	100 ns	12-25 ns	13 ns	0.21 / 0.75	0.41	0.13 / 0.32	0.22
P7	90 ns	3-20 ns	17 ns	0.22 / 0.75	0.41	0.15 / 0.32	0.22
	80 ns	8-20 ns	12 ns	0.24 / 0.78	0.43	0.18 / 0.33	0.22
	70 ns	8-20 ns	12 ns	0.25 / 0.73	0.42	0.16 / 0.30	0.22
	60 ns	3-20 ns	17 ns	0.21 / 0.80	0.42	0.13 / 0.33	0.22
P8	100 ns	5-20 ns	15 ns	0.23 / 0.67	0.38	0.17 / 0.33	0.23
	90 ns	1-20 ns	19 ns	0.22 / 0.73	0.38	0.18 / 0.32	0.23
	80 ns	1-20 ns	19 ns	0.22 / 0.72	0.39	0.17 / 0.32	0.22
P9	100 ns	8-20 ns	12 ns	0.26 / 0.67	0.41	0.18 / 0.32	0.23
P10	100 ns	5-20 ns	15 ns	0.20 / 0.76	0.37	0.17 / 0.32	0.22
P11	100 ns	14-30 ns	16 ns	0.26 / 0.65	0.40	0.16 / 0.33	0.23
P12	100 ns	10-25 ns	15 ns	0.23 / 0.70	0.39	0.18 / 0.32	0.23
	200 ns	3-20 ns	17 ns	0.20 / 0.68	0.37	0.15 / 0.33	0.22
P13	190 ns	10-20 ns	10 ns	0.19 / 0.66	0.37	0.15 / 0.33	0.22
	180 ns	7-20 ns	13 ns	0.19 / 0.73	0.37	0.16 / 0.33	0.22
	170 ns	1-20 ns	19 ns	0.19 / 0.74	0.38	0.15 / 0.33	0.22
	160 ns	9-25 ns	14 ns	0.19 / 0.70	0.38	0.15 / 0.33	0.22
P14	100 ns	10-25 ns	15 ns	0.22 / 0.76	0.39	0.16 / 0.33	0.23

Table SI-3. Used data for all the alchemical simulations performed.

6. Averaged FEP data

The average binding free energies from the two analysis methods for each pocket are provided in Table SI-4, with the standard error, the number of replicas that were used, and the probability to find bicarbonate in each pocket. We also report in Table SI-4 the number of free MD simulations where bicarbonate ended in each pocket. Data in Table SI-4 were used to draw Figure 5 from main text.

Pocket	Free MD	MBAR				TI			
		$\langle \Delta G \rangle$	$\sigma/\sqrt{N}$	N	%	$\langle \Delta G \rangle$	$\sigma/\sqrt{N}$	N	%
P1	31	0.6	0.6	3	0.0	0.7	0.6	3	0.0
P2	4	-1.7	0.6	3	0.0	-1.5	0.5	3	0.0
P3	3	-4.8	0.8	3	5.9	-4.7	0.8	3	6.5
P4	18	-1.2	1.6	3	0.0	-0.9	1.5	3	0.0
P5	**2**	**-5.4**	**1.1**	**5**	**15.3**	**-5.2**	**0.9**	**5**	**14.9**
P6	3	2.5	0.3	3	0.0	2.7	0.3	3	0.0
P7	**11**	**-5.7**	**0.1**	**4**	**24.1**	**-5.4**	**0.1**	**4**	**19.5**
P8	3	-2.1	0.5	3	0.1	-2.0	0.5	3	0.1
P9	4	-2.5	–	1	0.1	-2.3	–	1	0.1
P10	2	4.7	–	1	0.0	4.9	–	1	0.0
P11	4	1.2	–	1	0.0	1.0	–	1	0.0
P12	1	0.8	–	1	0.0	1.1	–	1	0.0
P13	**3**	**-6.1**	**0.9**	**5**	**54.5**	**-6.1**	**0.9**	**5**	**58.9**
P14	1	-1.7	–	1	0.0	-1.9	–	1	0.1

Table SI-4. Number of free MD simulations where bicarbonate ended in each pocket and average binding free energies in each pocket with MBAR and TI data. N is the number of replicas used to compute the average binding free energy after Dixon's Q-test. The column "%" gives the probability to find bicarbonate in each pocket.

7. Additional images of binding site with bicarbonate in pockets P5 and P7

We show in Figure SI-22 an image to complement Figure 7 from main text.

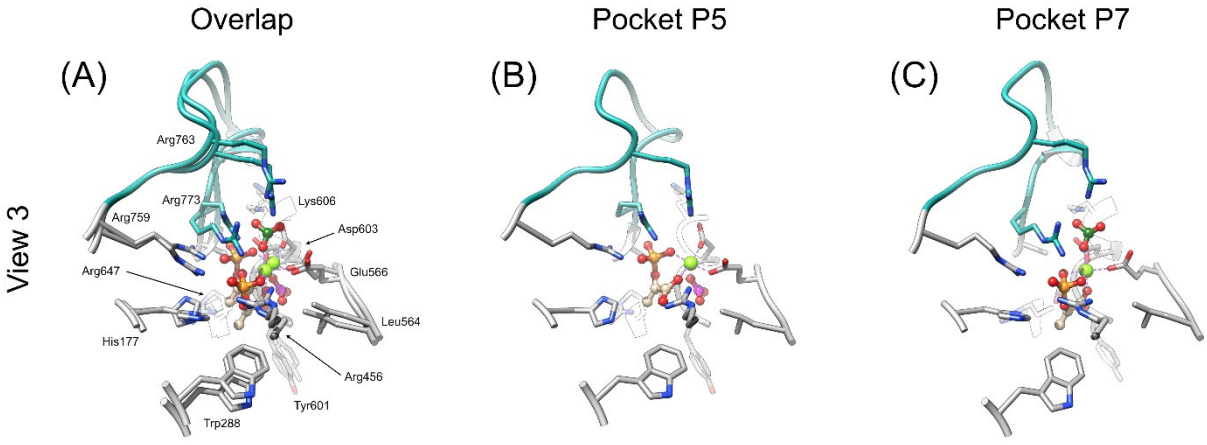


Figure SI-22. Active site of pockets P5 and P7 after alignment on the 29 displayed residues. This image is a third view of the Figure 7 from main text.

8. Full data with MBAR and TI

All the data from the thermodynamic cycle with MBAR analysis are presented Table SI-5.

Pocket	Data	Desolvation of HCO_3^-	Artefacts for HCO_3^-	Add restraints	Ligand insertion	Binding w. restraints	Artefacts for binding	Remove restraints	ΔG
P1	19 ns	92.7	-17.6	2.2	0.0	-91.2	15.7	0.0	**1.8**
	19 ns	92.7	-17.6	2.2	0.0	-92.2	15.7	-0.1	**0.7**
	16 ns	92.7	-17.6	2.2	0.0	-93.5	15.7	-0.1	**-0.6**
P2	17 ns	92.7	-17.6	2.2	0.0	-92.5	15.6	-0.7	**-0.4**
	19 ns	92.7	-17.6	2.2	0.0	-93.5	15.6	-1.3	**-2.0**
	19 ns	92.7	-17.6	2.2	0.0	-95.0	15.6	-0.4	**-2.6**
P3	14 ns	92.7	-17.6	2.2	0.0	-99.6	15.7	0.0	**-6.7**
	12 ns	92.7	-17.6	2.2	0.0	-96.6	15.7	0.0	**-3.7**
	21 ns	92.7	-17.6	2.2	0.0	-96.9	15.7	-0.1	**-4.1**
P4	18 ns	92.7	-17.6	2.2	0.0	-95.3	15.6	-1.9	**-4.4**
	15 ns	92.7	-17.6	2.2	0.0	-90.6	15.6	0.0	**2.2**
	16 ns	92.7	-17.6	2.2	0.0	-93.7	15.6	-0.7	**-1.5**
P5	12 ns	92.7	-17.6	2.2	0.0	-100.9	15.5	0.0	**-8.1**
	20 ns	92.7	-17.6	2.2	0.0	-101.3	15.5	0.0	**-8.5**
	22 ns	92.7	-17.6	2.2	0.0	-96.2	15.5	0.0	**-3.4**
	13 ns	92.7	-17.6	2.2	0.0	-96.4	15.5	0.0	**-3.6**
	19 ns	92.7	-17.6	2.2	0.0	-96.2	15.5	0.0	**-3.4**
P6	15 ns	92.7	-17.6	2.2	0.0	-91.1	15.6	0.0	**1.7**
	17 ns	92.7	-17.6	2.2	0.0	-89.9	15.6	0.0	**2.9**
	23 ns	92.7	-17.6	2.2	0.0	-90.0	15.6	0.0	**2.8**
P7	13 ns	92.7	-17.6	2.2	0.0	-103.3	15.5	0.0	**-10.5**
	17 ns	92.7	-17.6	2.2	0.0	-98.1	15.5	0.0	**-5.3**
	12 ns	92.7	-17.6	2.2	0.0	-98.5	15.5	0.0	**-5.7**
	12 ns	92.7	-17.6	2.2	0.0	-98.9	15.5	0.0	**-6.1**
	17 ns	92.7	-17.6	2.2	0.0	-98.3	15.5	0.0	**-5.5**
P8	15 ns	92.7	-17.6	2.2	0.0	-95.3	15.5	-0.7	**-3.2**
	19 ns	92.7	-17.6	2.2	0.0	-93.9	15.5	0.0	**-1.1**
	19 ns	92.7	-17.6	2.2	0.0	-94.9	15.5	0.0	**-2.1**
P9	12 ns	92.7	-17.6	2.2	0.0	-95.4	15.7	0.0	**-2.5**
P10	15 ns	92.7	-17.6	2.2	0.0	-85.7	15.6	-2.4	**4.7**
P11	16 ns	92.7	-17.6	2.2	0.0	-89.0	15.6	-2.6	**1.2**
P12	15 ns	92.7	-17.6	2.2	0.0	-91.7	15.5	-0.3	**0.8**
P13	17 ns	92.7	-17.6	2.2	0.0	-99.7	15.6	-1.1	**-8.0**
	10 ns	92.7	-17.6	2.2	0.0	-100.5	15.6	0.0	**-7.7**
	13 ns	92.7	-17.6	2.2	0.0	-96.9	15.6	0.0	**-4.1**
	19 ns	92.7	-17.6	2.2	0.0	-96.2	15.6	0.0	**-3.4**
	14 ns	92.7	-17.6	2.2	0.0	-100.4	15.6	0.0	**-7.6**
P14	15 ns	92.7	-17.6	2.2	0.0	-94.7	15.7	0.0	**-1.7**

Table SI-5. Decomposition of the binding free energy of bicarbonate for each replica of each pocket. Values are in kcal/mol and are from the MBAR analysis. Column 3, 5, 6, 7 and 9 corresponds to the five terms of the thermodynamic cycle. Column 6 is strictly 0 by definition. Column 4 and 8 are corrections for finite-size effects (see below).

The analysis of convergence to assess the length of the simulation and the overlap between windows were made with MBAR data. Another way to check the convergence is to compare the data from MBAR with those from thermodynamic integration (TI) since from the same simulations the two methods use different data and different equations. The decomposition of the binding free energy of bicarbonate in each replica in each pocket with TI analysis is given in Table SI-6. We plot the correlation between the two sets of binding free energies in Figure SI-23. As expected, the correlation is high with a Pearson correlation coefficient of R=0.9985.

Pocket	Data	Desolvation of HCO_3^-	Artefacts for HCO_3^-	Add restraints	Ligand insertion	Binding w. restraints	Artefacts for binding	Remove restraints	ΔG
P1	19 ns	92.9	-17.6	2.2	0.0	-91.3	15.7	0.0	**1.9**
	19 ns	92.9	-17.6	2.2	0.0	-92.4	15.7	-0.1	**0.7**
	16 ns	92.9	-17.6	2.2	0.0	-93.7	15.7	-0.1	**-0.6**
P2	17 ns	92.9	-17.6	2.2	0.0	-92.6	15.6	-0.7	**-0.2**
	19 ns	92.9	-17.6	2.2	0.0	-93.7	15.6	-1.3	**-1.9**
	19 ns	92.9	-17.6	2.2	0.0	-95.2	15.6	-0.4	**-2.5**
P3	14 ns	92.9	-17.6	2.2	0.0	-100.0	15.7	0.0	**-6.8**
	12 ns	92.9	-17.6	2.2	0.0	-96.8	15.7	0.0	**-3.6**
	21 ns	92.9	-17.6	2.2	0.0	-96.9	15.7	-0.1	**-3.8**
P4	18 ns	92.9	-17.6	2.2	0.0	-95.2	15.6	-1.9	**-4.1**
	15 ns	92.9	-17.6	2.2	0.0	-90.8	15.6	0.0	**2.3**
	16 ns	92.9	-17.6	2.2	0.0	-93.5	15.6	-0.7	**-1.1**
P5	12 ns	92.9	-17.6	2.2	0.0	-100.3	15.5	0.0	**-7.3**
	20 ns	92.9	-17.6	2.2	0.0	-101.4	15.5	0.0	**-8.3**
	22 ns	92.9	-17.6	2.2	0.0	-96.5	15.5	0.0	**-3.4**
	13 ns	92.9	-17.6	2.2	0.0	-96.8	15.5	0.0	**-3.7**
	19 ns	92.9	-17.6	2.2	0.0	-96.5	15.5	0.0	**-3.4**
P6	15 ns	92.9	-17.6	2.2	0.0	-91.2	15.6	0.0	**1.9**
	17 ns	92.9	-17.6	2.2	0.0	-89.8	15.6	0.0	**3.2**
	23 ns	92.9	-17.6	2.2	0.0	-90.1	15.6	0.0	**2.9**
P7	13 ns	92.9	-17.6	2.2	0.0	-103.4	15.5	0.0	**-10.4**
	17 ns	92.9	-17.6	2.2	0.0	-98.0	15.5	0.0	**-5.0**
	12 ns	92.9	-17.6	2.2	0.0	-98.6	15.5	0.0	**-5.6**
	12 ns	92.9	-17.6	2.2	0.0	-98.7	15.5	0.0	**-5.6**
	17 ns	92.9	-17.6	2.2	0.0	-98.4	15.5	0.0	**-5.4**
P8	15 ns	92.9	-17.6	2.2	0.0	-95.5	15.5	-0.7	**-3.1**
	19 ns	92.9	-17.6	2.2	0.0	-93.9	15.5	0.0	**-0.9**
	19 ns	92.9	-17.6	2.2	0.0	-95.0	15.5	0.0	**-1.9**
P9	12 ns	92.9	-17.6	2.2	0.0	-95.5	15.7	0.0	**-2.3**
P10	15 ns	92.9	-17.6	2.2	0.0	-85.8	15.6	-2.4	**4.9**
P11	16 ns	92.9	-17.6	2.2	0.0	-89.5	15.6	-2.6	**1.0**
P12	15 ns	92.9	-17.6	2.2	0.0	-91.7	15.5	-0.3	**1.1**
P13	17 ns	92.9	-17.6	2.2	0.0	-99.7	15.6	-1.1	**-7.8**
	10 ns	92.9	-17.6	2.2	0.0	-100.8	15.6	0.0	**-7.7**
	13 ns	92.9	-17.6	2.2	0.0	-97.3	15.6	0.0	**-4.2**
	19 ns	92.9	-17.6	2.2	0.0	-96.2	15.6	0.0	**-3.2**
	14 ns	92.9	-17.6	2.2	0.0	-100.5	15.6	0.0	**-7.4**
P14	15 ns	92.9	-17.6	2.2	0.0	-95.1	15.7	0.0	**-1.9**

Table SI-6. Decomposition of the binding free energy of bicarbonate for each replica of each pocket. Values are in kcal/mol and are from the TI analysis, except column 9. Column 3, 5, 6, 7 and 9 corresponds to the five terms of the thermodynamic cycle. Column 6 is strictly 0 by definition. Column 4 and 8 are corrections for finite-size effects (see below).

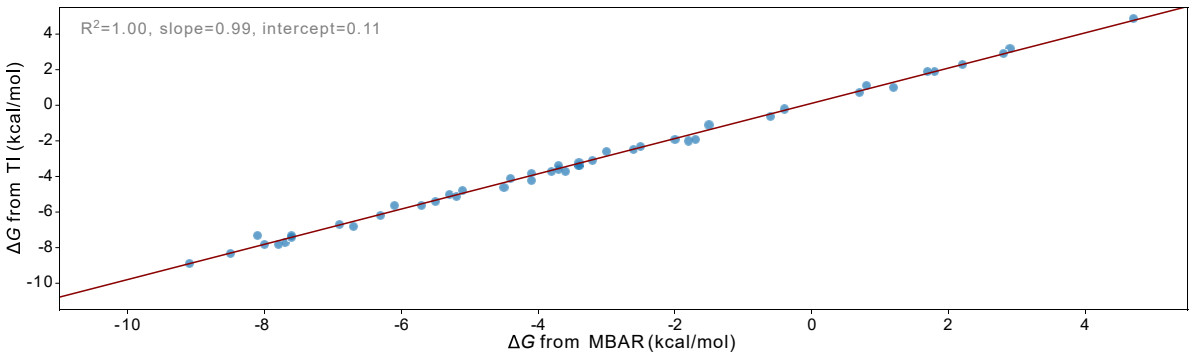


Figure SI-23. Correlation between data from TI and from MBAR.

The decomposition of binding free energies of bicarbonate in pockets P5 and P7 of the mutant K606N are provided in Table SI-7 and Table SI-8 with analysis from MBAR and TI respectively.

Pocket	Data	Desolvation of HCO_3^-	Artefacts for HCO_3^-	Add restraints	Ligand insertion	Binding w. restraints	Artefacts for binding	Remove restraints	ΔG
P5 K606N	15 ns	92.7	-17.6	2.2	0.0	-98.5	16.1	0.0	**-5.2**
	19 ns	92.7	-17.6	2.2	0.0	-98.4	16.1	-0.1	**-5.1**
	20 ns	92.7	-17.6	2.2	0.0	-96.3	16.1	0.0	**-3.0**
	19 ns	92.7	-17.6	2.2	00	-100.2	16.1	-0.7	**-7.6**
	19 ns	92.7	-17.6	2.2	0.0	-97.1	16.1	0.0	**-3.7**
	16 ns	92.7	-17.6	2.2	0.0	-95.0	16.1	-0.2	**-1.8**
P7 K606N	11 ns	92.7	-17.6	2.2	0.0	-101.0	15.9	0.0	**-7.8**
	15 ns	92.7	-17.6	2.2	0.0	-97.0	15.9	0.0	**-3.8**
	19 ns	92.7	-17.6	2.2	0.0	-102.3	15.9	0.0	**-9.1**
	17 ns	92.7	-17.6	2.2	0.0	-100.1	15.9	0.0	**-6.9**
	16 ns	92.7	-17.6	2.2	0.0	-97.7	15.9	0.0	**-4.5**
	17 ns	92.7	-17.6	2.2	0.0	-99.8	15.9	0.0	**-6.6**

Table SI-7. Decomposition of the binding free energy of bicarbonate for each replica in two pockets of the mutant K606N. Values are in kcal/mol and are from the MBAR analysis. Column 3, 5, 6, 7 and 9 corresponds to the five terms of the thermodynamic cycle. Column 6 is strictly 0 by definition. Column 4 and 8 are corrections for finite-size effects (see below).

Pocket	Data	Desolvation of HCO_3^-	Artefacts for HCO_3^-	Add restraints	Ligand insertion	Binding w. restraints	Artefacts for binding	Remove restraints	ΔG
P5 K606N	15 ns	92.9	-17.6	2.2	0.0	-98.7	16.1	0.0	**-5.1**
	19 ns	92.9	-17.6	2.2	0.0	-98.3	16.1	-0.1	**-4.8**
	20 ns	92.9	-17.6	2.2	0.0	-96.2	16.1	0.0	**-2.6**
	19 ns	92.9	-17.6	2.2	00	-100.2	16.1	-0.7	**-7.3**
	19 ns	92.9	-17.6	2.2	0.0	-97.0	16.1	0.0	**-3.4**
	16 ns	92.9	-17.6	2.2	0.0	-95.4	16.1	-0.2	**-2.0**
P7 K606N	11 ns	92.9	-17.6	2.2	0.0	-101.3	15.9	0.0	**-7.8**
	15 ns	92.9	-17.6	2.2	0.0	-97.1	15.9	0.0	**-3.7**
	19 ns	92.9	-17.6	2.2	0.0	-102.4	15.9	0.0	**-8.9**
	17 ns	92.9	-17.6	2.2	0.0	-100.1	15.9	0.0	**-6.7**
	16 ns	92.9	-17.6	2.2	0.0	-98.1	15.9	0.0	**-4.6**
	17 ns	92.9	-17.6	2.2	0.0	-100.0	15.9	0.0	**-6.6**

Table SI-8. Decomposition of the binding free energy of bicarbonate for each replica in two pockets of the mutant K606N. Values are in kcal/mol and are from the TI analysis, except column 9. Column 3, 5, 6, 7 and 9 corresponds to the five terms of the thermodynamic cycle. Column 6 is strictly 0 by definition. Column 4 and 8 are corrections for finite-size effects (see below).

9. Equilibration time and frequency of data

ΔU_{ij} and $dH/d\lambda$ were recorded every 500 fs during the simulations. Alchemlyb (through pymbar) provides functions to calculate the length of needed equilibration for each window hence the amount of time to discard (noted T_0), as well as the "statistical inefficiency" which corresponds to the period at with data should be subsampled to obtain a set of uncorrelated data (noted g) and the number of uncorrelated samples (noted N_{eff})[40]. These data for the 21 alchemical windows of one set of simulations are provided in Table SI-9, together with the ratio of disappearance of electrostatic interactions (noted q) and Lennard-Jones interactions (noted vdW).

Window	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
q	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
vdW	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
T_0 (ps)	353	0	3504	245	1176	0	4867	2888	3	8294
g	8.6	7.7	6.5	7.3	7.8	6.3	6.5	6.9	8.1	7.7
N_{eff}	5753	6523	6645	6763	6110	7899	6223	6426	6209	4312

Window	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
q	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
vdW	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
T_0 (ps)	4646	11	1	16877	43	0	0	0	0	0	0
g	8.8	6.7	6.1	6.4	6.1	1.0	1.0	1.0	1.0	1.0	1.0
N_{eff}	4636	7482	8236	2558	8178	50002	50002	50002	50002	50002	50002

Table SI-9. Equilibration time, statistical inefficiency and number of uncorrelated samples for a set of simulations.

The data displayed in Table SI-9 surprised us. It is hardly understandable why between windows 12, 13 and 14 (which differ by the amount of remaining van der Waals interactions, respectively 80%, 70% and 60%), the first one needs 1 ps to be equilibrated, the second needs almost ~17000 ps, and the third needs 43 ps. These results explain why we didn't use this automated function to find the equilibration time and we instead used the protocol described above. Moreover, the same can be said for the statistical inefficiency: if we compare windows 14 and 15 (which differs only by the fact that in the former 60% of the van der Waals interactions are kept whereas it is only 50% in the latter), it is unclear why the first one would need data to be subsampled every 6.1 steps to obtain uncorrelated data (i.e. every 3 ps since data were recorded every 500 fs), whereas the second one could use all data (i.e. every 1 step).

To decide the frequency of using data, we compare the influence of using all data or a subsample of them (every n steps) in Table SI-10 for the same set of simulations as in Table SI-9. We observed that the computed free energy is not affected by the choice at which data are subsampled (range of 0.2 kcal/mol between the highest and the lowest value). As expected, the MBAR uncertainties increases with the number of steps (since there are less data) but in a very acceptable way. Thus, we set a step of 2 (i.e. data were used every 1 ps) and didn't use the statistical inefficiency functions from alchemlyb.

Steps	**ΔG (kcal/mol)**	**Error**		**Steps**	**ΔG (kcal/mol)**	**Error**
1	99.64	0.02		**10**	99.73	0.07
2	99.65	0.03		**15**	99.63	0.09
3	99.61	0.04		**20**	99.72	0.10
4	99.64	0.04		**30**	99.67	0.12
5	99.71	0.05		**40**	99.66	0.14

Table SI-10. Free energies and MBAR uncertainties when using different amount of data from a representative simulation.

10. Decomposition of free energy in alchemical transformations

The change of free energies between two neighbor states along the alchemical transformations are presented in Figure SI-24 for the desolvation of bicarbonate as well as for the binding of bicarbonate in three meaningful pockets (with data from MBAR analysis). As expected, the largest contribution of the total change in free energy is by far the removal or appearance of electrostatic interactions.

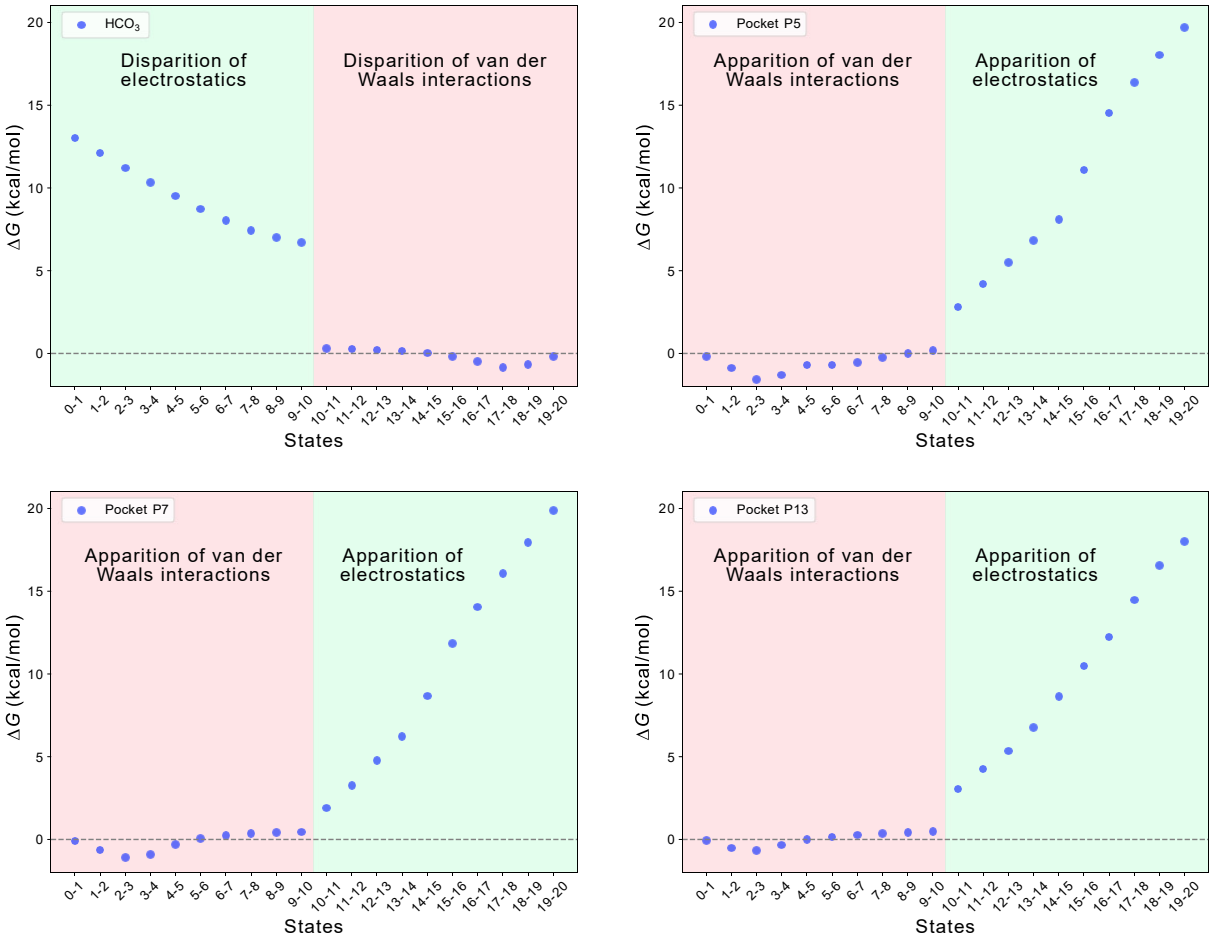


Figure SI-24. Decomposition of the full binding free energies (from MBAR) between each pair of states for (a) the desolvation of bicarbonate, (b)/(c)/(d) the binding of bicarbonate in three meaningful pockets of PEP Carboxylase.

11. Outlier in pocket P7

In pocket P7 with MBAR data, the computed binding free energies of the snapshots extracted at 100, 90, 80, 70 and 60 ns are respectively -10.5, -5.3, -5.7, -6.1 and -5.5 kcal/mol. Thus, there is a difference of 4.9 kcal/mol between the first snapshot and the four others, and the first value is found to be an outlier with Dixon's Q test (see main text). The molecular explanation lies in the position of the bicarbonate in the extracted snapshots, which is slightly shifted by 1 Å in one structure with respect to the four others (see Figure SI-25). Thus, there is some room between Asp603 and Arg773 to add an extra water molecule during the solvation procedure when building the system prior to FEP calculations since there is a void in the structure (as already pointed out, once snapshots were extracted we build again the system with solvation, addition of ions, and new equilibration). This extra water molecule explains the observed difference of binding free energies. Indeed, for the five sets of simulations, we have decomposed the binding free energies to get the change between two neighbor states along the alchemical transformations (as in Figure SI-24), and we focused on the last five transitions where most of the electrostatic interactions are created (i.e. from window 15 to window 20). With MBAR data, the change of binding free energy between these windows in the snapshot at 100 ns is 3.2 kcal/mol lower than the average of the four other snapshots, which accounts for most of the difference of 4.9 kcal/mol. With TI data, this change is of 3.8 kcal/mol out of 5.0 kcal/mol. Thus, the electrostatic interactions between one water molecule and its environment explain the low bicarbonate binding free energy found with one snapshot of pocket P7. We underline that the insertion of the extra water was not observed during the simulations, but is a consequence of the void due to position of bicarbonate in the snapshot and the re-solvation prior to FEP calculations. Moreover, having a low binding free energy for bicarbonate does not mean that this structure is more favorable than the others since the binding free energy of water in this site is probably positive due to entropic reason. Thus, this justifies to consider this data as an outlier and to remove it from the analysis.

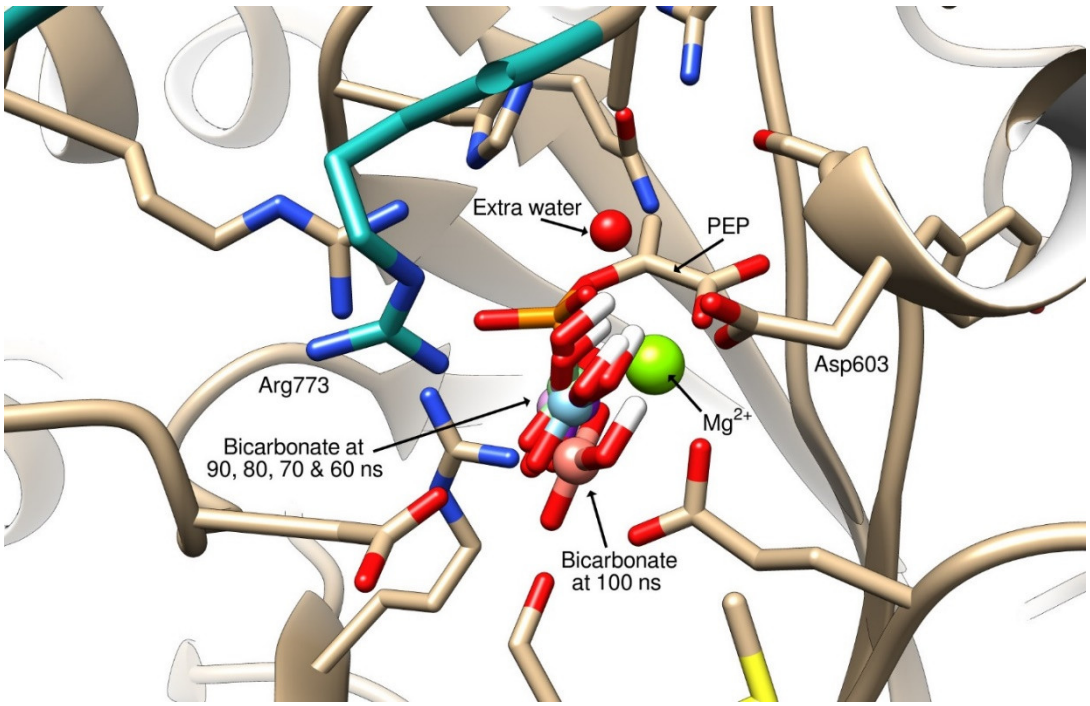


Figure SI-25. Superimposition of the position of bicarbonate in the five snapshots used for pocket P7. The water molecule that is found only in the snapshot at 100 ns is displayed as a red ball. Bicarbonate are represented as sticks and only their carbon atom is shown as a ball. Snapshot at 100 ns is in salmon, whereas the ones at 90, 80, 70 and 60 ns are in purple, blue, green and pink.

12. Corrections due to charged ligands

To reproduce experimental data (obtained in bulk), simulations are usually performed in periodic boundary conditions. However, computational artifacts (called finite-size effects) arise in the calculations of free energies under these conditions, which preclude the direct comparisons between the two. These artifacts can be due to the size of the simulation box or to the boundary conditions that are applied at its edges. "Due to the long range and large magnitude of electrostatic interactions, these errors are nearly exclusively electrostatic in nature, affect primarily the charging component of the ligand insertion process, and are most significant for systems where the ligand bears a net charge, especially when the protein is also charged"[41]. Indeed, during alchemical transformations with a charged ligand, the total charge of the system is different between the fully coupled and the fully uncoupled states. Thus, electrostatic interactions will change between the first and last windows (such as between box replicas or with the background electrostatic potential in the box) which influences the overall potential energy, hence the computed free energy.

Different approaches were proposed to deal with the case of charged ligands. First, several tricks can be used to avoid changing the total charge of the system. One could make a ligand appear far from the protein at the same time than the ligand bound to the protein disappears. To be useful, the box needs to be very large which may be prohibitive. Another option, if the studied ligand is negatively charged (e.g.), is to remove a sodium ion at the same time than the ligand, or make appear a chloride ion. In these two cases, the computed free energies are then the sum of two contributions and one must compute a solvation free energy afterwards (of the sodium or the chloride ion) to compare computed binding free energies with experiments. In these solvation free energies, a charged species is used and we are then facing the same original problem of how to correct the simulations from the finite-size effects. Thus, we didn't pursue these approaches. Two other methods were proposed to correct the computational artifacts[41,42], and it was shown that they provide comparable results[43]. We have here used the analytical scheme proposed by Rocklin *et al.*[41] since it is straightforward to use, is a post-simulation correction, and takes into account the dispersion of charges in the protein.

We will here briefly describe the terms that were used to compute the finite-size effects and correct the results from the FEP simulations. Full details are available in Ref. [41]. According to the used scheme, four terms corresponding to physical effects plus an empirical term must be computed. They are described below with the corresponding equations (adapted from Ref. [41] to take into account the more general situation of non-cubic boxes):

- Periodicity-induced net-charge interactions (equation 15 from [41]), which corrects the interaction between the reference box, the periodic replicas, and the neutralizing background:

$$\Delta\Delta G_{NET}(L) = -\frac{N_A}{8\pi\varepsilon_0}[(Q_P + Q_L)^2 - Q_P^2]\frac{\xi_{LS}}{L} \qquad (1)$$

- Periodicity-induced undersolvation (equation 16 from [41]), which corrects the fact that the volume of solutes in periodic replicas is unavailable for the solvation of the reference solute:

$$\Delta\Delta G_{USV}(L) = \frac{N_A}{8\pi\varepsilon_0}\left(1 - \frac{1}{\varepsilon_S}\right)[(Q_P + Q_L)^2 - Q_P^2]\frac{\xi_{LS}}{L} \qquad (2)$$

- Residual integrated potential effects (equation 17 from [41]), which corrects the fact that the two previous contributions were derived with the assumption that "protein and ligand were both point charges without solvent-excluded volume and located at the same position in space"[41]. Considering the excluded volume can modify the potential in the box, hence ΔG.

$$\Delta\Delta G_{RIP}(L) = [(I_P + I_L)(Q_P + Q_L) - I_P Q_P]\frac{1}{V} \quad (3)$$

I_P and I_L are the difference of integral of potentials, defined as follows (X denotes P or L, B stands for "box", and HET denotes the fact that it is a heterogeneous situation):

$$I_X = B_{HET}(X, L_{ref}) - B_{HET}(Q_X, L_{ref}) \quad (4)$$

$$With: \quad B_{HET}(X, L_{ref}) = \int_{L_{ref}} \Phi_{HET,X}(\boldsymbol{r}) d^3\boldsymbol{r} \quad (5)$$

$$B_{HET}(Q_X, L_{ref}) = -\frac{N_A \xi_{CB}}{4\pi\varepsilon_0\varepsilon_S} Q_X L_{ref}^2 \quad (6)$$

$B_{HET}(X, L_{ref})$ is obtained through a calculation of the potential $\Phi_{HET,X}$ around the solute with a continuum-electrostatics approach. In practice, since we are looking at local effects, a Poisson-Boltzmann (PB) calculations in a large cubic box (which mimics non-periodic boundary conditions) is used. Different PB solvers can be used, and we relied on APBS[44,45]. The potential is then integrated over the box. Formally, $B_{HET}(Q_X, L_{ref})$ is the integral of the potential around a single point charge Q_X. However, if the ligand is centered in the box during the Poisson-Boltzmann calculations with APBS, $B_{HET}(Q_X, L_{ref})$ can be written as above. Thus, to compute $\Delta\Delta G_{RIP}$, one only needs to perform two calculations with APBS (one for $X = P$ and on for $X = L$) and integrate the obtained potentials.

- Discrete solvent effects (equation 35 from [41]), which corrects the offset due to the fact that the references for the potential are different between the bulk situation and the periodic boundary conditions. Indeed, in bulk the reference for the electrostatic potential is 0 at infinite distances, whereas with periodic boundary conditions it is the average electrostatic potential over the box which is set at 0:

$$\Delta\Delta G_{DSC}(N_S, L) = -\frac{N_A \gamma_S Q_L}{6\varepsilon_0}\frac{N_S}{V} \quad (7)$$

- Empirical term (equation 25 from [41]), which is a correction to $\Delta\Delta G_{USV}$ "to ensure that [we] reproduce the exact analytical result in the special case of a single point charge at the center of a spherical cavity"[41]:

$$\Delta\Delta G_{EMP}(L) = -\frac{N_A}{8\pi\varepsilon_0}\frac{16\pi^2}{45}\left(1 - \frac{1}{\varepsilon_S}\right)[(Q_P + Q_L)^2 - Q_P^2]\frac{R_L^5}{V^2} \quad (8)$$

$$With: \quad R_L = \sqrt{\frac{I_{L,SLV}}{\frac{N_A}{8\pi\varepsilon_0}\frac{4\pi}{3}\left(1 - \frac{1}{\varepsilon_S}\right)Q_L}} \quad (9)$$

$I_{L,SLV}$ is defined as follows and needs a third Poisson-Boltzmann calculation with APBS:

$$I_{L,\,SLV} = I_L - \{B_{HOM}[L,\, L_{ref}] - B_{HOM}[Q_L,\, L_{ref}]\} \quad (10)$$

B_{HOM} are the analogs of B_{HET} but with "a homogeneous dielectric medium of permittivity ε_I".

$$B_{HOM}(L, L_{ref}) = RT * \int_{L_{ref}} \Phi_{HOM,L}(\boldsymbol{r}) d^3\boldsymbol{r} \quad (11)$$

$$B_{HOM}(Q_L, L_{ref}) = -\frac{N_A \xi_{CB}}{4\pi\varepsilon_0\varepsilon_I} Q_L L_{ref}^2 \quad (12)$$

In these equations, P and L denotes the protein and the ligand. Q_P is the net charge of the protein ($-38 \cdot e$), Q_L is the net charge of the ligand ($-1 \cdot e$), e is the elementary charge ($1.602176634 \cdot 10^{-19}\ C$), ε_0 is the vacuum permittivity ($8.8541878128 \cdot 10^{-12}\ J^{-1}m^{-1}C^2$), N_A is the Avogadro number, ε_S is the static relative dielectric permittivity of the solvent (97 for the TIP3P water model[46]), ε_I is the static relative dielectric permittivity (1), R is the gas constant ($8.31446261815324\ JK^{-1}mol^{-1}$), T is the absolute temperature, ξ_{LS} is the lattice-sum integration constant (≈ -2.8372, see section X-7), $\xi_{CB} = \pi/2 - 3 \cdot ln(2+\sqrt{3})$ is the cubic Coulomb integration constant[47], L is the size of the box, V is the volume of the box ($V = L^3$ for cubic boxes), L_{ref} is the length of a reference box for PB calculations and was set at $25\ nm$ (Ref. [41] proposed $15\ nm$ for simulation boxes of $7.42\ nm$, but since the highest dimension of our boxes is $15.1\ nm$ we increased L_{ref}), N_S is the number of solvent molecules ($\sim$72000), γ_S is "the quadrupole-moment trace of the solvent model relative to its single Waals interaction site" ($\gamma_S = \sum_{i=1}^{n} q_i r_i^2 = 0.00764\ e \cdot nm^2$ for the TIP3P water model). During PB calculations with APBS, the probe radius was set at 1.4 Å. Since our simulations used counter-ions, Q_P was artificially set to 0 "to account for an effective surface neutralization of the protein by the counter-ions" in the equations above. To compute the corrections for the desolvation of bicarbonate, I_P and Q_P were set to 0 in the above equations.

At the end, the correction $\Delta\Delta G_{Total}$ is the sum of 5 terms and the corrected computed free energy is:

$$\Delta G_{Simulations} = \Delta G_{From\ FEP} + \Delta\Delta G_{USV} + \Delta\Delta G_{NET} + \Delta\Delta G_{RIP} + \Delta\Delta G_{DSC} + \Delta\Delta G_{EMP}$$

Once it is set up, the procedure proposed by Rocklin *et al.*[41] is straightforward to use since the FEP protocol doesn't need to be changed and the corrections are applied afterwards. However, setting it up can be tricky. Rocklin *et al.* proposed in the Supporting Information of their article a script and input files to perform the calculations of the residual integrated potential term; unfortunately, there are typos in the script and it doesn't calculate the other terms. A python module exists to perform this task (https://github.com/xiki-tempula/rocklinc), but we aimed at an even simpler workflow. Thus, we provide all the files needed to perform the full calculations. We propose a bash script that prepares the files (calling pdb2pqr[48] and APBS[44,45]) and then calls a python script (build from the one from Rocklin *et al.*) that computes the correction terms all at once.

For each pocket, 50 snapshots were extracted (every 1ns) and the protocol was applied independently on each snapshot (see Table SI-11 for the results). Three out of the five correction terms depend on the box volume and are thus different for each snapshot, but the standard deviations are always small.

(kcal/mol)	$\Delta\Delta G_{Total}$	$\Delta\Delta G_{NET}$	$\Delta\Delta G_{USV}$	$\Delta\Delta G_{RIP}$	$\Delta\Delta G_{DSC}$	$\Delta\Delta G_{EMP}$
P1	15.7 ± 0.07	3.6	-3.5	0.0 ± 0.07	15.6 ± 0.01	-4.9e-5 ± 5.7e-6
P2	15.6 ± 0.05	3.5	-3.5	-0.1 ± 0.05	15.7 ± 0.02	-9.6e-6 ± 1.2e-5
P3	15.7 ± 0.07	3.5	-3.4	-0.1 ± 0.06	15.7 ± 0.02	-1.4e-4 ± 2.3e-5
P4	15.6 ± 0.04	3.6	-3.5	-0.1 ± 0.04	15.6 ± 0.02	-1.9e-4 ± 2.1e-5
P5	15.5 ± 0.06	3.6	-3.5	-0.1 ± 0.05	15.6 ± 0.02	-1.9e-4 ± 2.8e-5
P6	15.6 ± 0.05	3.6	-3.5	-0.1 ± 0.05	15.6 ± 0.02	-1.7e-4 ± 1.9e-5
P7	15.5 ± 0.06	3.6	-3.5	-0.1 ± 0.05	15.6 ± 0.02	-4.6e-5 ± 1.4e-5
P8	15.5 ± 0.05	3.6	-3.5	-0.1 ± 0.05	15.6 ± 0.02	-4.0e-5 ± 1.1e-5
P9	15.7 ± 0.07	3.6	-3.5	0.0 ± 0.07	15.6 ± 0.02	-8.4e-5 ± 1.7e-5
P10	15.6 ± 0.06	3.6	-3.5	-0.1 ± 0.05	15.6 ± 0.02	-9.9e-5 ± 2.4e-5
P11	15.6 ± 0.05	3.6	-3.5	-0.1 ± 0.05	15.6 ± 0.02	-2.5e-4 ± 2.9e-5
P12	15.5 ± 0.06	3.6	-3.5	-0.2 ± 0.06	15.6 ± 0.02	-2.4e-4 ± 2.2e-5
P13	15.6 ± 0.05	3.6	-3.5	-0.1 ± 0.05	15.6 ± 0.02	-2.4e-4 ± 1.7e-5
P14	15.7 ± 0.05	3.6	-3.5	0.0 ± 0.05	15.6 ± 0.02	-2.5e-4 ± 3.0e-5
HCO_3^-	17.6 ± 0.11	10.9	-10.8	0.0 ± 0.00	17.5 ± 0.11	-6.1e-5 ± 1.2e-5

Table SI-11. Full data for corrections of finite size effects, averaged over 50 snapshots when needed.

13. Lattice-sum integration constant for non-cubic boxes

In the equations presented above that are used to correct artifacts from periodic-boundary conditions, a term called the *lattice-sum integration constant* arises (ξ_{LS}). This term is linked to the total coulombic energy of a single particle of charge +1 placed at the origin of a box in the limit of infinitesimal distances. Indeed, in periodic boundary conditions there is a background electrostatic field whose purpose is to set the reference with $\langle \Phi \rangle_{Box} = 0$. Thus, for an ion alone in a box, the ion will interact with the background field and the total electrostatic energy (sum of direct and reciprocal terms) is:

$$E_{Coulombic} = \frac{1}{2} \frac{Na \cdot e^2}{4\pi\varepsilon_0} \frac{\xi_{LS}}{L}$$

For cubic boxes, the value of ξ_{LS} has been precisely estimated by Redlack and Grindlay[49] and is $\xi_{LS} = -2.837297479484$. However, for non-cubic boxes, no values and no protocols are found in the literature. We have thus looked for the computational parameters that allow to recover the ξ_{LS} value of Redlack and Grindlay for cubic boxes, before applying the optimal protocol to other box shapes. Since the value of ξ_{LS} given above is obtained at the limit of infinitesimal distances, for each given set of parameters we have computed the total coulombic energy for different box sizes (length L from 1.6 to 5.0 nm, every 0.1 nm). After a linear fit, we have looked for the intercept at L=0 and compared the value with the tabulated one. In the tables below, we report the error multiplied by 10^6.

In Gromacs, several parameters have an influence on the coulombic energy and we have changed them successively. First, we have looked at the interpolation order for PME (pme-order keyword). 4 is the default, and the value of 8 gives the lowest error. However, since the difference between 8 and 10 is small and we know that high values should be better, we have kept a value of 10 for the following.

pme-order	**ewald-rtol**	**rlist**	**fourierspacing**	**$10^6 \cdot$ error**
4	10^{-5}	0.8	0.12	53.7
6	10^{-5}	0.8	0.12	53.9
8	10^{-5}	0.8	0.12	50.4
10	10^{-5}	0.8	0.12	51.9

We have then looked at the influence of the "relative strength of the Ewald-shifted direct potential" at the cut-off distance (quote from Gromacs manual). The default value of ewald-rtol is 10^{-5}. Increasing ewald-rtol gives a more accurate reciprocal sum and a less accurate direct sum, which is the expected behavior in the current case. This is true from 10^{-5} to 10^{-2} (which provides the lowest error), but going above 10^{-2} largely increases the error.

pme-order	**ewald-rtol**	**rlist**	**fourierspacing**	**$10^6 \cdot$ error**
10	**10^{-5}**	0.8	0.12	51.9
10	**10^{-4}**	0.8	0.12	37.7
10	**10^{-3}**	0.8	0.12	11.9
10	**10^{-2}**	0.8	0.12	9.2
10	**10^{-1}**	0.8	0.12	-1903.0
10	**10^{0}**	0.8	0.12	$-6*10^{40}$

In parallel of ewald-rtol, we have looked at the cut-off distance. For computational efficiency, the same distance is used by Gromacs for (1) the cut-off for the neighbor list (rlist keyword), (2) the distance for the short-range/long-range electrostatics separation (rcoulomb keywork), (3) the cut-off for Lennard-Jones interactions (rvdw keywork). The default value is 0.8 nm and the value of 1.3 nm provides the lowest error (even though ξ_{LS} is obtained at infinitesimal distances, a value larger than default is

better). Since the box size must be at least twice the value of rlist, this explains why we have computed the desolvation free energy of bicarbonate in a relatively large box with a size of 4.3 x 4.3 x 4.3 nm^3.

pme-order	ewald-rtol	rlist	fourierspacing	$10^6 \cdot$ error
10	10^{-5}	**0.8**	0.12	51.9
10	10^{-5}	**1.0**	0.12	29.4
10	10^{-5}	**1.1**	0.12	30.2
10	10^{-5}	**1.2**	0.12	9.0
10	10^{-5}	**1.3**	0.12	3.8
10	10^{-5}	**1.4**	0.12	29.6

Finally, we have looked at the number of grid points that are used in the reciprocal space (fourierspacing keyword, with a default of 0.12 nm): for a box of size L, L/fourierspacing gives the number of points in the grid (so a low value should be more accurate). For this last step, we have used the optimal values for the three other keywords. All values provide comparable errors, but the lowest one is obtained with 0.10 nm.

pme-order	ewald-rtol	rlist	fourierspacing	$10^6 \cdot$ error
10	10^{-2}	1.3	**0.15**	-1.1
10	10^{-2}	1.3	**0.14**	-0.4
10	10^{-2}	1.3	**0.13**	-0.7
10	10^{-2}	1.3	**0.12**	-0.5
10	10^{-2}	1.3	**0.11**	-0.3
10	10^{-2}	1.3	**0.10**	-0.2
10	10^{-2}	1.3	**0.09**	0.9
10	10^{-2}	1.3	**0.08**	1.8
10	10^{-2}	1.3	**0.07**	0.5
10	10^{-2}	1.3	**0.06**	1.0

Using the optimal set (pme-order=10 / ewald-rtol=10^{-2} / rlist=1.3 / fourierspacing=0.10), the intercept provides a value for ξ_{LS} of -2.837297682300 for cubic boxes with an error of $-0.2 \cdot 10^{-6}$ with respect to the value of Redlack and Grindlay[49].

When equations (1) and (2) were actually used to correct a simulation, we computed the total electrostatic energy for a sodium ion placed at the origin of an empty box of the same shape as the one used in the simulation (using the average size of the box). The calculation was done with the parameters obtained above. We then obtained directly ξ_{LS}/L with the equation:

$$\frac{\xi_{LS}}{L} = \frac{8\pi\varepsilon_0}{Na \cdot e^2} \cdot E_{Coulombic}$$

In the procedure described above to obtain the optimal computational parameters, ξ_{LS} was obtained at the limit of infinitesimal distances. However, with the obtained PME parameters, the slope of ξ_{LS} as a function of L is almost 0 ($6.5 \cdot 10^{-7}\ nm^{-1}$, and we add that the slope is $-250 \cdot 10^{-7}\ nm^{-1}$ with the default parameters). This means that one can use the value of ξ_{LS}/L at any distance and doesn't need the intercept. For example, for a cubic box with $L = 4.3\ nm$, the computed value of ξ_{LS}/L provides an error of $-0.8 \cdot 10^{-6}$. To stay consistent for the correction of artifacts between the desolvation of the ligand and the binding of the ligand to the protein, we have used this protocol also for cubic boxes even if in such cases both ξ_{LS} and L are known.

XII. Data and software availability

All software used for the current article are open-source and can be obtained through their webpages. All scripts, input files, analysis files are available at the address: https://zenodo.org/doi/10.5281/zenodo.10204297.

XIII. Bibliography

(1) Kai, Y.; Matsumura, H.; Inoue, T.; Terada, K.; Nagara, Y.; Yoshinaga, T.; Kihara, A.; Tsumura, K.; Izui, K. Three-Dimensional Structure of Phosphoenolpyruvate Carboxylase: A Proposed Mechanism for Allosteric Inhibition. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (3), 823–828. https://doi.org/10.1073/pnas.96.3.823.

(2) Matsumura, H.; Terada, M.; Shirakata, S.; Inoue, T.; Yoshinaga, T.; Izui, K.; Kai, Y. Plausible Phosphoenolpyruvate Binding Site Revealed by 2.6 Å Structure of Mn2+-Bound Phosphoenolpyruvate Carboxylase from Escherichia Coli. *FEBS Letters* **1999**, *458* (2), 93–96. https://doi.org/10.1016/S0014-5793(99)01103-5.

(3) Matsumura, H.; Xie, Y.; Shirakata, S.; Inoue, T.; Yoshinaga, T.; Ueno, Y.; Izui, K.; Kai, Y. Crystal Structures of C4 Form Maize and Quaternary Complex of E. Coli Phosphoenolpyruvate Carboxylases. *Structure* **2002**, *10* (12), 1721–1730. https://doi.org/10.1016/S0969-2126(02)00913-9.

(4) Dharmarajan, L.; Kraszewski, J. L.; Mukhopadhyay, B.; Dunten, P. W. Structure of an Archaeal-Type Phosphoenolpyruvate Carboxylase Sensitive to Inhibition by Aspartate. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79* (6), 1820–1829. https://doi.org/10.1002/prot.23006.

(5) Paulus, J. K.; Schlieper, D.; Groth, G. Greater Efficiency of Photosynthetic Carbon Fixation Due to Single Amino-Acid Substitution. *Nat Commun* **2013**, *4*, 1518–1523. https://doi.org/10.1038/ncomms2504.

(6) Schlieper, D.; Förster, K.; Paulus, J. K.; Groth, G. Resolving the Activation Site of Positive Regulators in Plant Phosphoenolpyruvate Carboxylase. *Molecular Plant* **2014**, *7* (2), 437–440. https://doi.org/10.1093/mp/sst130.

(7) Connell, M. B.; Lee, M. J. Y.; Li, J.; Plaxton, W. C.; Jia, Z. Structural and Biochemical Characterization of Citrate Binding to AtPPC3, a Plant-Type Phosphoenolpyruvate Carboxylase from Arabidopsis Thaliana. *Journal of Structural Biology* **2018**, *204* (3), 507–512. https://doi.org/10.1016/j.jsb.2018.11.003.

(8) González-Segura, L.; Mújica-Jiménez, C.; Juárez-Díaz, J. A.; Güémez-Toro, R.; Martinez-Castilla, L. P.; Muñoz-Clares, R. A. Identification of the Allosteric Site for Neutral Amino Acids in the Maize C4 Isozyme of Phosphoenolpyruvate Carboxylase: The Critical Role of Ser-100. *J. Biol. Chem.* **2018**, *293* (26), 9945–9957. https://doi.org/10.1074/jbc.RA118.002884.

(9) Muñoz-Clares, R. A.; González-Segura, L.; Juárez-Díaz, J. A.; Mújica-Jiménez, C. Structural and Biochemical Evidence of the Glucose 6-Phosphate-Allosteric Site of Maize C4-Phosphoenolpyruvate Carboxylase: Its Importance in the Overall Enzyme Kinetics. *Biochemical Journal* **2020**, *477* (11), 2095–2114. https://doi.org/10.1042/BCJ20200304.

(10) Carrizosa-Carbajal, E. I.; Gonzalez-Segura, L.; Munoz-Clares, R. A. Crystal Structures of the T-State of Maize C4-Phosphoenolpyruvate Carboxylase Provides Insight into the Structural Basis of the Allosteric Transition of Plant Phosphoenolpyruvate Carboxylase Enzymes. *To be Published* **2024**.

(11) Meyer, T. J.; Sheng, J.; Haesaerts, S.; Frausto, K.; O'Leary, S.; Loris, R.; Larsen, P. B. Amino Acid Changes That Deregulate Phosphoenolpyruvate Carboxylase in Plants. *To be Published* **2024**.

(12) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GRGMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4* (3), 435–447. https://doi.org/10.1021/ct700301q.

(13) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29* (7), 845–854. https://doi.org/10.1093/bioinformatics/btt055.
(14) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.
(15) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.
(16) Jorgensen, W. L. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *Journal of the American Chemical Society* **1981**, *103* (2), 335–340. https://doi.org/10.1021/ja00392a016.
(17) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.
(18) 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.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, 2009.
(19) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *Journal of Molecular Graphics and Modelling* **2006**, *25* (2), 247–260. https://doi.org/10.1016/j.jmgm.2005.12.005.
(20) Sousa Da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes* **2012**, *5*, 1–8. https://doi.org/10.1186/1756-0500-5-367.
(21) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.
(22) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *Journal of Computational Chemistry* **1997**, *18*, 1463–1472.
(23) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4* (1), 116–122. https://doi.org/10.1021/ct700200b.
(24) Bussi, G.; Zykova-Timan, T.; Parrinello, M. Isothermal-Isobaric Molecular Dynamics Using Stochastic Velocity Rescaling. *The Journal of Chemical Physics* **2009**, *130* (7), 074101. https://doi.org/10.1063/1.3073889.
(25) Berendsen, H. J. C.; Postma, J. C.; Van Gunsteren, J. P. M.; Di Nola, W. F.; Haak, A. ; Molecular Dynamics with Coupling to an External Bath. *Studies in Molecular Dynamics. I. General Method The Journal of Chemical Physics* **1984**, *81* (8), 234505. https://doi.org/10.1063/1.448118.
(26) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52* (12), 7182–7190. https://doi.org/10.1063/1.328693.
(27) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A Portable Plugin for Free-

Energy Calculations with Molecular Dynamics. *Computer Physics Communications* **2009**, *180* (10), 1961–1972. https://doi.org/10.1016/j.cpc.2009.05.011.
(28) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Computer Physics Communications* **2014**, *185* (2), 604–613. https://doi.org/10.1016/j.cpc.2013.09.018.
(29) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; Capelli, R.; Carloni, P.; Ceriotti, M.; Cesari, A.; Chen, H.; Chen, W.; Colizzi, F.; De, S.; De La Pierre, M.; Donadio, D.; Drobot, V.; Ensing, B.; Ferguson, A. L.; Filizola, M.; Fraser, J. S.; Fu, H.; Gasparotto, P.; Gervasio, F. L.; Giberti, F.; Gil-Ley, A.; Giorgino, T.; Heller, G. T.; Hocky, G. M.; Iannuzzi, M.; Invernizzi, M.; Jelfs, K. E.; Jussupow, A.; Kirilin, E.; Laio, A.; Limongelli, V.; Lindorff-Larsen, K.; Löhr, T.; Marinelli, F.; Martin-Samos, L.; Masetti, M.; Meyer, R.; Michaelides, A.; Molteni, C.; Morishita, T.; Nava, M.; Paissoni, C.; Papaleo, E.; Parrinello, M.; Pfaendtner, J.; Piaggi, P.; Piccini, G.; Pietropaolo, A.; Pietrucci, F.; Pipolo, S.; Provasi, D.; Quigley, D.; Raiteri, P.; Raniolo, S.; Rydzewski, J.; Salvalaglio, M.; Sosso, G. C.; Spiwok, V.; Šponer, J.; Swenson, D. W. H.; Tiwary, P.; Valsson, O.; Vendruscolo, M.; Voth, G. A.; White, A.; The PLUMED consortium. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat Methods* **2019**, *16* (8), 670–673. https://doi.org/10.1038/s41592-019-0506-8.
(30) Jorgensen, W. L.; Thomas, L. L. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J. Chem. Theory Comput.* **2008**, *4* (6), 869–876. https://doi.org/10.1021/ct800011m.
(31) Mobley, D. L.; Klimovich, P. V. Perspective: Alchemical Free Energy Calculations for Drug Discovery. *J. Chem. Phys.* **2012**, *137* (23), 230901. https://doi.org/10.1063/1.4769292.
(32) Chipot, C. Frontiers in Free-Energy Calculations of Biological Systems. *WIREs Computational Molecular Science* **2014**, *4* (1), 71–89. https://doi.org/10.1002/wcms.1157.
(33) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57* (12), 2911–2937. https://doi.org/10.1021/acs.jcim.7b00564.
(34) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46* (1), 531–558. https://doi.org/10.1146/annurev-biophys-070816-033654.
(35) Mey, A. S. J. S.; Allen, B. K.; Macdonald, H. E. B.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best Practices for Alchemical Free Energy Calculations. *LiveCoMS* **2020**, *2* (1), 18378. https://doi.org/10.33011/livecoms.2.1.18378.
(36) Song, L. F.; Merz, K. M. Jr. Evolution of Alchemical Free Energy Methods in Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60* (11), 5308–5318. https://doi.org/10.1021/acs.jcim.0c00547.
(37) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **2003**, *107* (35), 9535–9551. https://doi.org/10.1021/jp0217839.
(38) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *The Journal of Chemical Physics* **2008**, *129* (12), 124105. https://doi.org/10.1063/1.2978177.
(39) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the Analysis of Free Energy Calculations. *J Comput Aided Mol Des* **2015**, *29* (5), 397–411. https://doi.org/10.1007/s10822-015-9840-9.
(40) Chodera, J. D. A Simple Method for Automated Equilibration Detection in Molecular Simulations. *J. Chem. Theory Comput.* **2016**, *12* (4), 1799–1805. https://doi.org/10.1021/acs.jctc.5b00784.
(41) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. Calculating the Binding Free Energies of Charged Species Based on Explicit-Solvent Simulations Employing Lattice-Sum Methods: An Accurate Correction Scheme for Electrostatic Finite-Size Effects. *J. Chem. Phys.* **2013**, *139* (18), 184103. https://doi.org/10.1063/1.4826261.

(42) Lin, Y.-L.; Aleksandrov, A.; Simonson, T.; Roux, B. An Overview of Electrostatic Free Energy Computations for Solutions and Proteins. *J. Chem. Theory Comput.* **2014**, *10* (7), 2690–2709. https://doi.org/10.1021/ct500195p.
(43) Chen, W.; Deng, Y.; Russell, E.; Wu, Y.; Abel, R.; Wang, L. Accurate Calculation of Relative Binding Free Energies between Ligands with Different Net Charges. *J. Chem. Theory Comput.* **2018**, *14* (12), 6346–6358. https://doi.org/10.1021/acs.jctc.8b00825.
(44) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proceedings of the National Academy of Sciences* **2001**, *98* (18), 10037–10041. https://doi.org/10.1073/pnas.181342398.
(45) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS Biomolecular Solvation Software Suite. *Protein Science* **2018**, *27* (1), 112–128. https://doi.org/10.1002/pro.3280.
(46) Höchtl, P.; Boresch, S.; Bitomsky, W.; Steinhauser, O. Rationalization of the Dielectric Properties of Common Three-Site Water Models in Terms of Their Force Field Parameters. *The Journal of Chemical Physics* **1998**, *109* (12), 4927–4937. https://doi.org/10.1063/1.477104.
(47) Nijboer, B. R. A.; Ruijgrok, Th. W. On the Energy per Particle in Three- and Two-Dimensional Wigner Lattices. *J Stat Phys* **1988**, *53* (1–2), 361–382. https://doi.org/10.1007/BF01011562.
(48) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Research* **2007**, *35* (Web Server), W522–W525. https://doi.org/10.1093/nar/gkm276.
(49) Redlack, A.; Grindlay, J. The Electrostatic Potential in a Finite Ionic Crystal. *Can. J. Phys.* **1972**, *50* (22), 2815–2825. https://doi.org/10.1139/p72-375.