bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Selective Permeability of Carboxysome Shell Pores to Anionic Molecules

Paween Mahinthichaichan[a,1], Dylan M. Morris[b,1], Yi Wang[d], Grant J. Jensen[b,c,1], Emad Tajkhorshid[a,1]

[a]*Department of Biochemistry, NIH Center for Macromolecular Modeling and Bioinformatics, Center for Biophysics and Quantitative Biology, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign*
[b]*Division of Biology and Biological Engineering, California Institute of Technology*
[c]*Howard Hughes Medical Institute*
[d]*Department of Physics, Chinese University of Hong Kong*

Abstract

Carboxysomes are closed polyhedral cellular microcompartments that increase the efficiency of carbon fixation in autotrophic bacteria. Carboxysome shells consist of small proteins that form hexameric units with semi-permeable central pores containing binding sites for anions. This feature is thought to selectively allow access to RuBisCO enzymes inside the carboxysome by HCO_3^- (the dominant form of CO_2 in the aqueous solution at pH 7.4) but not O_2, which leads to a non-productive reaction. To test this hypothesis, here we use molecular dynamics simulations to characterize the energetics and permeability of CO_2, O_2, and HCO_3^- through the central pores of two different shell proteins, namely, CsoS1A of α-carboxysome and CcmK4 of β-carboxysome shells. We find that the central pores are in fact selectively permeable to anions such as HCO_3^-, as predicted by the model.

Introduction

All life depends on the ability of cells to fix atmospheric carbon into organic matter. The key enzyme in this process is ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), which catalyzes the fixation reaction of CO_2 and ribulose-1,5-bisphosphate (RuBP) molecules to produce two molecules of 3-phosphoglycerate (3PGA), a precursor molecule for sugar and amino acid biosynthesis. Besides fixing CO_2 and RuBP, the enzyme fixes O_2 and RuBP, producing one molecule of 3PGA and one molecule of 2-phosphoglycolate, a wasteful compound [1–5]. RuBisCO is notoriously inefficient with K_m of $>150\,\mu$M for CO_2 and k_{cat} of the

*Corresponding authors: G.J.J. jensen@caltech.edu; E.T. emad@life.illinois.edu
[1]Equal contribution

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reaction in the order of $10\,s^{-1}$ [6–8]. It is important to note that the concentration of O_2 in the atmosphere is ∼21% whereas that of CO_2 is only ∼0.04%, which is lower than the K_m of CO_2 for the RuBisCO enzymes [3, 5].

To increase the efficiency of RuBisCO, cyanobacteria and carbon-fixing chemoautotrophic bacteria encapsulate RuBisCO and carbonic anhydrase in specialized protein-enclosed cytoplasmic microcompartments called carboxysome [9–12]. To mitigate the occurrence of the O_2 fixation reaction, the carboxysome needs to minimize the penetration of O_2 into its lumen by mechanisms which would also impact effective entry of CO_2 due to its chemical resemblance to O_2. CO_2 is envisioned to enter the carboxysomal lumen in the form of HCO_3^- [10, 11, 13], its predominant form at physiological pH. Carbonic anhydrase then converts the HCO_3^- to CO_2 [14], providing a mechanism for concentrating CO_2 in the immediate vicinity of the RuBisCO enzymes.

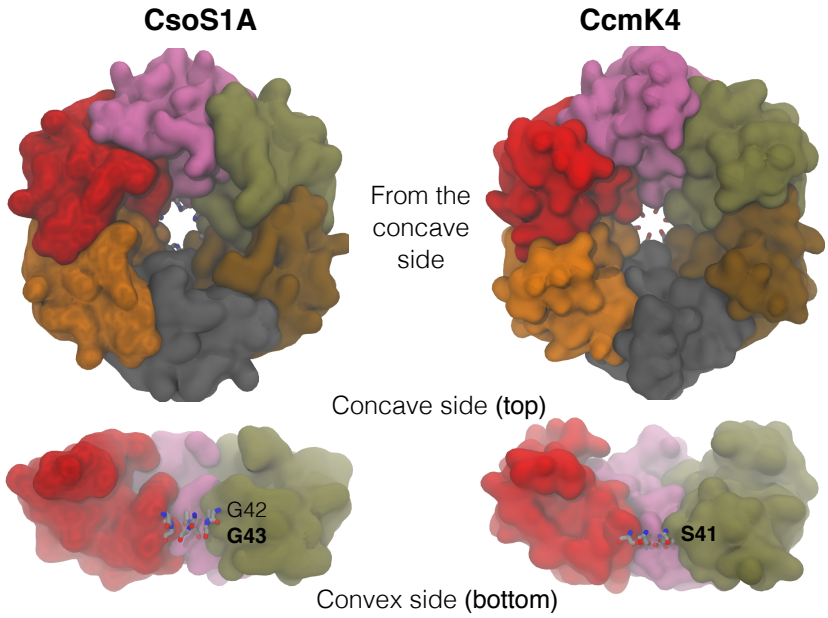

Figure 1: Arrangement of carboxysome shell proteins CsoS1A and CcmK4. Top view from the concave side (upper panel) and side view (lower panel). Individual colored molecular surfaces represent individual subunits of the homohexamers. Each hexamer forms a central pore. The narrowest section (bottleneck) of the pore is formed by six G43 residues for CsoS1A and six S41 residues for CcmK4.

Carboxysomes are classified into α and β types, based on their composition and evolutionary history [3]. In their native forms, both types resemble icosahedral capsids with a diameter of ∼1,000 Å [9, 15–18]. The outer shell of the carboxysome is formed by the assembly of thousands of copies of a few proteins [2]. α carboxysomes are found in *Prochlorococcus* and *Synechococcus* species such as *Halothiobacillus neapolitanus*, and in some other chemoau-

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

totrophic bacteria [2, 3, 11, 13, 19]. Their main shell protein is CsoS1A, of which the structure was determined from *H. neapolitanus* [20]. β carboxysomes are found in freshwater species such as *Synechococcus elongates PCC 7942* and *Synechocytis sp. PCC 6803* [3, 11, 13, 19], and their main shell proteins are CcmK1-4. These main shell proteins form hexamers and are arranged in a hexagonal lattice with aqueous-exposed surfaces on either side [20–23] (Fig. 1). Along the 6-fold symmetry axis of each hexamer is a pore, termed "the central pore", with the bottleneck radius of $\sim$2 Å [2], potentially permitting small molecular species, such as HCO_3^-, CO_2 and/or O_2 molecules to pass through.

Here, we examine the permeability of HCO_3^-, CO_2, and O_2 molecules through the central pores of CsoS1A and CcmK4 complexes in full atomic details using molecular dynamics simulations and free energy calculations. The umbrella sampling (US) technique is employed to calculate the free energy profiles for HCO_3^-, CO_2, and O_2 insertion. We find that the central pore of carboxysome shells are preferentially selective for HCO_3^-, over CO_2 and O_2 .

Materials and Methods

Simulation systems

The simulations were prepared using the crystal structures of CsoS1A from *H. neapolitanus* (PDB entry 2EWH) resolved at 1.4 Å [20] and CcmK4 from *Synechocytis sp. PCC 6803* (PDB entry 2A18) resolved at 2.28 Å [21]. CsoS1A is missing its first five amino acids. CcmK4 is missing its first three amino acids and its last thirteen amino acids. Since these amino acids are located neither near the central pore or at the interface of individual monomeric subunits, they were not modeled. The asymmetric unit of CsoS1A structure is provided by PDB as a monomer while that of CcmK4 structure is provided as a trimer. The biologically relevant hexameric complexes were constructed by VMD [24] using the transformation matrices provided in the PDB files. Each modeled hexamer was centered at the origin so that the 6-fold symmetry axis (and the central pore) coincided with the z-axis. A hexameric periodic box was constructed according to the crystallographic dimensions given in the PDB files. Internal water molecules were added to each complex with DOWSER [25]. The complex was then solvated with TIP3P waters [26] and ionized with 150 mM NaCl. The resulting hydrated hexameric CsoS1A and CcmK4 systems comprised 28,115 and 27,619 atoms, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Simulation protocols

The simulations were performed using NAMD2 [27] with a time step of 2 fs and the CHARMM36 force field [28, 29]. The periodic boundary condition (PBC) was used throughout the simulations. All covalent bonds involving hydrogen atoms were kept rigid using the SHAKE algorithm [30]. To evaluate long-range electrostatic interactions in PBC without truncation, the particle mesh Ewald method [31] with a grid density of $1/\text{Å}^3$ was used. The cutoff for van der Waals interactions was set at 12 Å. The simulations were performed under NPT ensemble. The temperature was maintained at 300 K by Langevin dynamics [32] with a damping coefficient γ of 1 /ps. The Nosé-Hoover Langevin piston method [32, 33] with a piston period of 200 fs was used to maintain the pressure at 1 atm.

Equilibration and steered molecular dynamics

The equilibrations of the hydrated complexes of CsoS1A and CcmK4 began with 5,000 steps of energy minimization using the conjugated gradient algorithm, followed by 0.5-ns protein heavy-atom restrained, 0.5-ns protein backbone-atom restrained and 1-ns protein C_α-atom restrained simulations with $k = 1\,\text{kcal/mol/Å}^2$, and finally 10-20 ns unrestrained simulations. In the CsoS1A system, one of the simulated Cl^- ions entered the central pore and bound to residues G43 (of the hexamer) located at the center of the pore during the unrestrained simulation. The binding of a Cl^- ion was also observed in the equivalent section of CcmK4, corresponding to the backbones of residues S41. To simulate HCO_3^-, CO_2 or O_2 molecule, the bound Cl^- molecule was replaced by HCO_3^-, CO_2 or O_2 molecules. The force field parameters of these substrate molecules are available in the CHARMM general force field [34]. For the simulation systems with a CO_2 or O_2 molecule, one Na^+ ion was removed for neutralization. 5,000 steps of energy minimization and 100 ps of equilibration were performed on these three systems. Steered molecular dynamics simulations [35] were performed to generate starting structures for the US simulations, described in the following section. The backbone nitrogen atoms of residues G42 of CsoS1A or residues S41 of CcmK4 were used to mark the center of the pore, and these were considered to be at $z = 0$. Using the center of the pore as the starting point, the localized substrate molecule was pulled out of the pore at a velocity of 10 Å/ns using a force constant $k = 10\,\text{kcal/mol/Å}^2$. The simulation consisted of two sets. In one set, the molecule was pulled from $z = 0$ to $z = 20$ Å (towards the concave surface). In the other set, it was pulled from $z = 0$ to $z = -15$ Å (towards the convex surface). To prevent protein translation artificially induced by the force applied on the pulling of the substrate molecule, the positions of C_α atoms of residue G6 of CsoS1A

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and those of residues E11 of CcmK4, located away from the pore, were restrained with $k = 10\,\mathrm{kcal/mol/\AA^2}$.

Umbrella sampling and free energy calculations

The free energy (ΔG) profiles of HCO_3^-, CO_2 and O_2 translocation along the central pores were calculated using the US technique [36–38]. The starting frames for the US simulations were taken from the steered molecular dynamics trajectories described above. US simulations spanned from z = -15 Å to z = 20 Å, comprising 71 0.5-Å windows, and each simulation lasted 2.5 ns. This 35-Å length is about the thickness of the carboxysome shell determined from cryo-electron tomography and atomic force microscopy [15, 16, 39]. A harmonic potential with $k = 10\,\mathrm{kcal/mol/\AA^2}$ was applied to confine the substrate molecule to the center of each window. To construct the ΔG profiles, the last 2-ns trajectories of all of the simulations of a substrate were combined and analyzed using the weighted histogram analysis method (WHAM) [40], with a 0.25-Å histogram bin. Insertion ΔG values of the substrate in individual bins (ΔG_i) were subtracted by ΔG in the bulk solution (ΔG_{bulk}), yielding relative insertion free energies ($\Delta G_{i,bulk}$). The WHAM code was implemented by Professor Alan Grossfield at the University of Rochester Medical Center ($http://membrane.urmc.rochester.edu/content/wham$).

Results and Discussion

Hydration along the central pores

The hydrated complexes of CsoS1A and CcmK4 were simulated for 20 ns without RuBisCO substrates (i.e., HCO_3^-, CO_2, and O_2) in order to relax the proteins and prepare initial structures used for the US simulations.

Although the central pores were initially dehydrated, they became hydrated in less than 1 ns (Fig. 2). To determine the degree of hydration along each of the pores, a 14-Å radius cylinder covering the entire pore, including its concave and convex funnels, was defined. The z-coordinates of the oxygen atoms of water molecules localized within the cylinder during the last 10 ns of the simulations were recorded and clustered into a histogram with 0.5-Å bins, yielding a distribution profile of the water. The pore radius profiles were calculated using the HOLE program [41] to determine the degree of accessibility of the pores, and the water profiles were normalized with respect to cross-sectional areas along the sections within

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

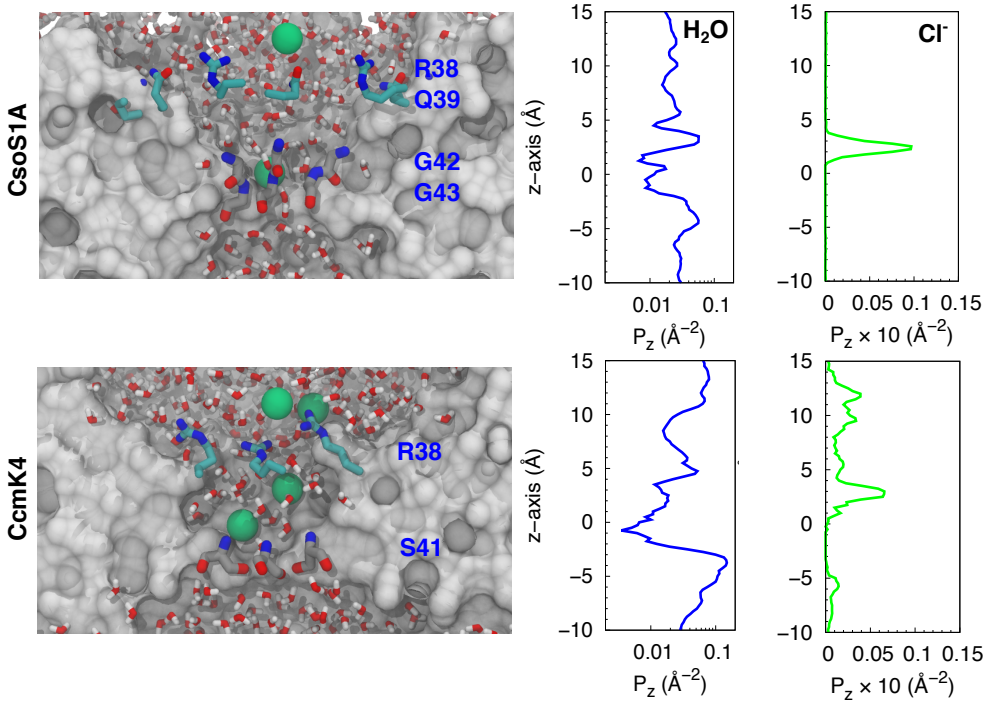

Figure 2: Hydration and Cl^- partitioning along the central pores. Left panels, taken from the equilibrium simulations, delineate the localizations of water and Cl^- (green spheres) within the central pores. Middle and right panels show the distribution profiles of water molecules and Cl^- ions normalized using cross-sectional areas along the sections within the proes. The backbone nitrogen atoms of residues G43 of CsoS1A and S41 of CcmK4 mark the center of the pore and are considered $z = 0$.

the pores. The normalized profiles shown in the middle panels of Fig. 2 represent the lowest occupancy site for water molecules at $z = 0$, which corresponds to the narrowest section, or the bottleneck, of the pore (Fig. 3, right panels). For CsoS1A, this bottleneck is formed by residues G42 and G43. For CcmK4, the equivalent section is formed by residues S41.

These simulations along with previous structural studies suggest favorable anion binding along the central pores. During the 20-ns period of each of the simulations, spontaneous binding of Cl^- ions was observed. The Cl^- ions bound the amine groups of residues G43 of CsoS1A and those of residues S41 of CcmK4. As shown in the left panels of Fig. 2, the highest Cl^- occupancy site was located at $z = 3$ Å. For CcmK4, apparent accumulation of Cl^- ions was also found near residues R38, which lie between $z = 8$ Å and $z = 14$ Å. The binding of other ions such as SO_4^{2-} has also been reported by X-ray crystallography [20, 42]. The crystal structure of CsoS1A, which was used in the simulations, contains one SO_4^{2-} molecule bound to the amine groups of G43 [20]. The same binding is also found in the crystal structures of CcmK1 and CcmK2, which are homologues of CcmK4 [42]. For CcmK1 and CcmK2, the serine residues equivalent to S41 of CcmK4 bind a SO_4^{2-} molecule. These

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

results all point to the presence of a positive electrostatic potential within the pore attracting negatively charged molecules.

Selectivity of anionic substrates

Free energy (ΔG) profiles of HCO_3^-, CO_2, and O_2 insertion (Fig. 3, left panels) were calculated in order to determine the selectivity of the central pores for these substrates. In agreement with the results of our equilibrium simulations, the resulting free energy profiles demonstrate favorable binding of anionic species (i.e., HCO_3^-) to the central pores, in both CsoS1A and CcmK4. For HCO_3^- insertion in CsoS1A, a free energy well was found in the region between z = 1 Å and z = 5 Å. The lowest ΔG was located at z = 3 Å and is ~-2 kcal/mol. For CcmK4, favorable HCO_3^- binding regions (relative to bulk water) spanned from z = 1 Å to z = 15 Å, with the global ΔG minimum of ~-2.5 kcal/mol occurring between z = 2 Å and z = 5 Å. Another binding region with a weaker insertion free energy (~-2 kcal/mol) is discernible between z = 8 Å and z = 14 Å, and corresponds to the position of a second Cl^- binding site observed in one of our equilibrium simulations (Fig. 2, bottom panels).

While the binding of HCO_3^- in the central pore is generally favorable, in both CsoS1A and CcmK4 central pores, this substrate has to overcome uphill free energy changes during its translocation from one side of the shell proteins to the other. Both CsoS1A and CcmK4 exhibited high free energy values at z = -2.5 Å (Fig. 3, left panels). These barriers are significantly smaller for HCO_3^- when compared to CO_2 (see below). Furthermore, the presence of a long attractive region (at $0 < z < 5$ Å) will provide a high probability for the substrate presence right next to the high-energy region.

In contrast to HCO_3^-, the permeation of CO_2 and O_2 through these central pores was more unfavorable with insertion ΔG of 2-4 kcal/mol, extending from z = 5 Å to z = -5 Å. The highest ΔG for CO_2 and O_2 insertion, ~4 kcal/mol, was located at z = ~3 Å, corresponding to the global ΔG minima for HCO_3^- insertion (Fig. 3, left panels). The same section was found to have a high density of water molecules and Cl^- ions (Fig. 2). In CsoS1A too, this highest barrier coincides with the region where the mobility of water molecules (as well as other molecular species) was found to be minimum (Fig. 4, left panel).

Because compacted water molecules can hinder the passage of relatively nonpolar CO_2 and O_2, we analyzed the mobility of HCO_3^-, CO_2 and O_2 along the central pores by calculating lateral diffusion coefficients (D) of these molecules. The diffusion profiles showed that as any

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

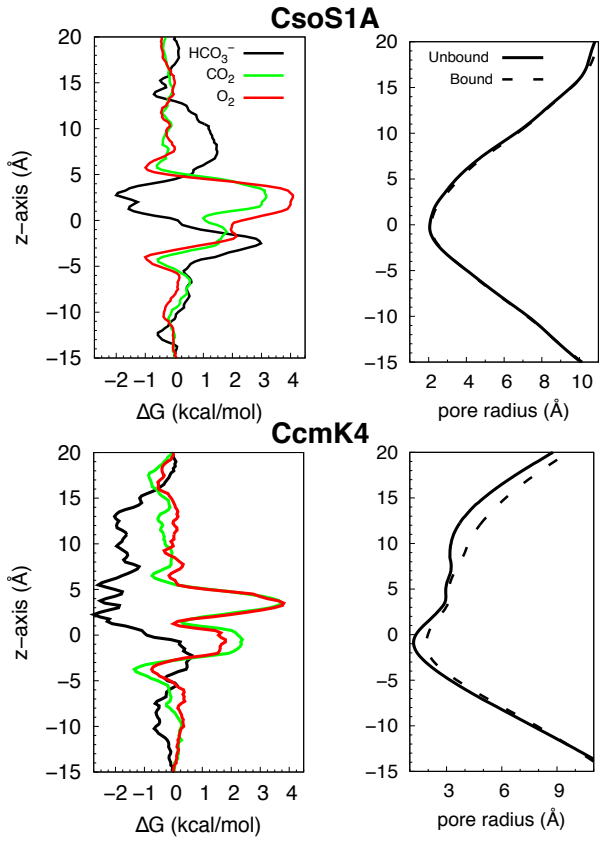

Figure 3: Partitioning of the substrates along the central pores. Left, free energy (ΔG) profiles for inserting substrates of the RuBisCO enzymes along the pores. Right, pore radius profiles of the central pores in the presence and absence of the substrate. HCO_3^- was used to represent the substrate as the other substrates similar profiles showed similar profiles.

of these molecules approach the bottleneck, their diffusivity becomes significantly diminished, by 10-40 fold (Fig. 4), consistent with a decrease in pore size (Fig. 3, right panels).

Dynamics of protein and substrates

The movement of a molecule through a protein may involve not only structural perturbations of the protein, from breathing motions of lining amino acids [43–46] to largescale conformational changes [47–51], but also the dynamics of the passing molecule (substrate or ligand) [52]. Although the pores of CsoS1A and CcmK4 are relatively wide to accommodate a free passage of small molecules, such as HCO_3^-, CO_2 and O_2, the calculated ΔG profiles indicated that such molecules cannot pass through easily.

Pore radii along the central pores were re-calculated upon the localization of a substrate molecule within the bottleneck. For CsoS1A, the bottleneck pore radius remained $\sim$2 Å (Fig.

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

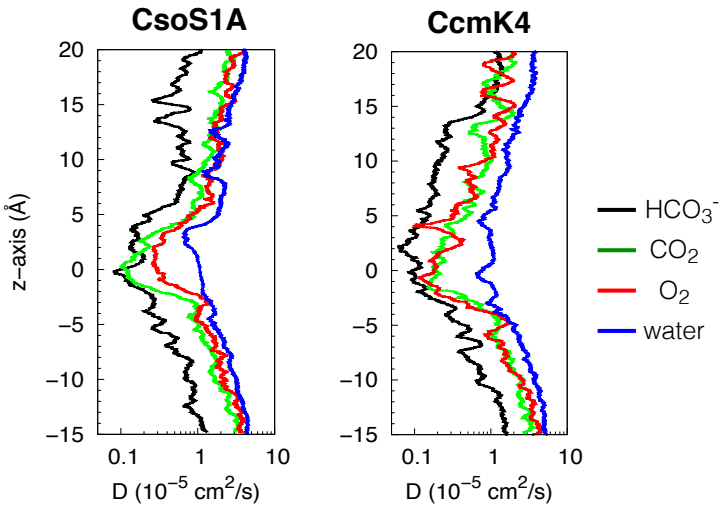

Figure 4: Lateral diffusion coefficient (D) profiles of HCO_3^-, CO_2, O_2, and water molecules along the central pores of CsoS1A and CcmK4 hexamers. D is defined as $< (\Delta x)^2 + (\Delta y)^2 > / 4\Delta t$, where Δx and Δy were the displacements of the molecule along the x-axis and the y-axis, respectively, at a z position and $\Delta t = 10$ ps. $< ... >$ denotes mean.

3, upper right panel). For CcmK4, it increased from 1.5 to 2 Å (Fig. 3, lower right panel). This change indicates a slight opening of the pathway, consistent with an increase in the radius of gyration between the six hydroxyl groups of S41 from 3 to 4 Å. In this configuration, the hydroxyl groups of S41 are oriented away from the pathway. Nevertheless, the bottleneck pore radii of both CsoS1A and CcmK4 are ~2 Å, suggesting that the permeating molecules have to orient themselves just right to transit through the bottleneck.

Conformational selectivity of the substrates was determined by calculating P_1 and P_2 order parameters, shown in Fig. 6, delineating the binding orientation of the substrate molecules upon migrating along the pore. As shown in Fig. 6, the orientation of the substrates was isotropic near the protein surface or in the bulk solution (z<-10 Å or z>15 Å). HCO_3^- became conformationally restricted when approaching or being within the bottleneck of the central pores. Upon positioning towards the concave surface, its negatively charged carboxylate faced the bottleneck when located at the ΔG minimum or z = 3 Å, as indicated by P_1 ~-1. The molecule flipped by 180° during the transition between the ΔG minimum and the bottleneck, as indicated by P_1 ~1. Between the bottleneck and z = -5 Å (convex funnel), it became perpendicularly oriented with respect to the pore, indicated by P_2 ~-0.5.

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

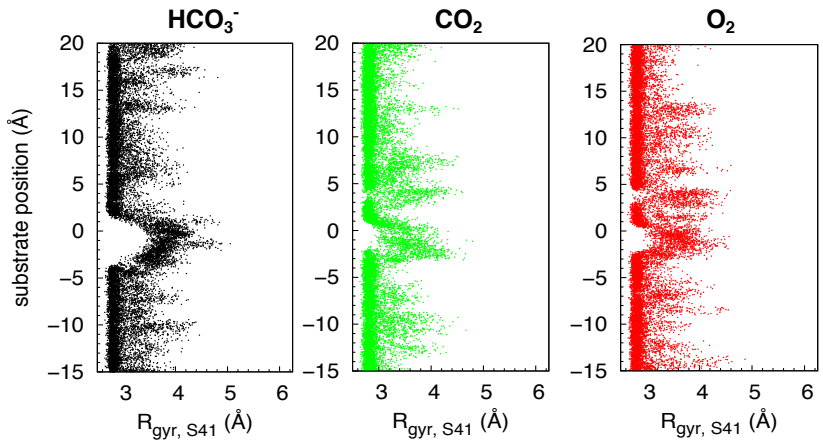

Figure 5: Conformational changes of bottlneck residues in CcmK4 upon substrate binding along the pore determined by calculating the radius of gyration (R_{gyr}) of the hydroxyl groups of residues S41. This R_{gyr} was calculated because in the crystal structure, the hydroxyl groups of these bottleneck amino acids appear to constrict the permeation pathway.

Charge distributions along the central pores

Electrostatic potential surface calculations support the preferential selectivity of the central pores of CsoS1A and CcmK4 for anions over nonpolar molecules. Time-averaged electrostatic potentials of the protein complexes were calculated using the structures of the protein hexamers generated from the last 10 ns of the equilibrium simulations (Fig. 7). The calculated electrostatic potential surfaces are consistent with ones previously calculated using the crystal structures [20, 21], supporting the idea that the pores are highly charged.

These calculations also rationalize the features observed in the calculated ΔG profiles (Fig. 3, left panels). For CcmK4, the central pore is positive, which provides attraction for HCO_3^- molecules. For CsoS1A, on the other hand, electrostatic potentials along the pore were asymmetrically distributed. The funnel from the concave surface to the bottleneck was positively charged, whereas the one from its convex surface was negatively charged (Fig. 7). This charge separation provides an explanation for the nature of the permeation ΔG barrier for HCO_3^- at the bottleneck. Because the convex surface is negatively charged, it is unfavorable for negatively charged molecules, such as HCO_3^-, to spontaneously transit through the bottleneck (z = -2.5 Å) from the concave surface to the convex surface. This is in agreement with the ΔG calculations, which showed a 2 kcal/mol higher barrier for a HCO_3^- molecule to enter the pore from the concave surface than from the convex surface.

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

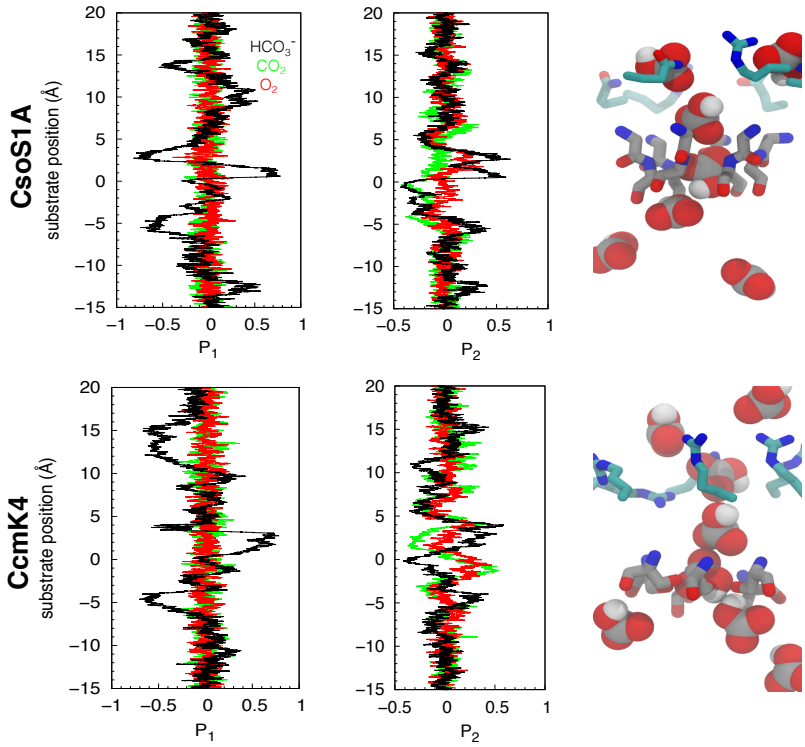

Figure 6: Binding orientation of substrates along the central pores (P_1 or P_2 order parameters versus substrate's positions along the z axis). $P_1 = <\cos\theta>$ and $P_2 = <(3\cos^2\theta\text{-}1)/2>$. P_1 is $\sim\pm1$ when a substrate molecule is oriented parallel to the pore axis and ~0 when it is oriented either isotropically or perpendicularly with respect to the pore axis. P_2 differentiates between isotropic average orientation ($P_2 = 0$) and orthogonality ($P_2 = $-0.5), in which the molecule is in a perpendicular orientation. θ corresponds to the tilt angle with respect to the pore. For HCO_3^-, θ is the angle between the C-OH bond vector and the pore axis. For linear CO_2 and O_2 molecules, θ is the angle between the main axis of the substrate and the pore axis. The right panels show several snapshots of HCO_3^- taken from the US simulations.

Conclusion

The results of the simulations and free energy calculations presented in this study show that the central pores of carboxysome shell protein complexes favor HCO_3^- over either CO_2 or O_2. Once within the carboxysomal lumen, HCO_3^- is converted to CO_2 by carbonic anhydrase. The poor permeability of the carboxysome shell to CO_2 and O_2 can improve the productivity of RuBisCO in two ways; not only can it minimize outward leakage of CO_2 from carboxysomal lumen upon production by carbonic anhydrase, but it will also prevent unwanted entry of O_2 into the lumen. These results substantiate, at a molecular level, how carboxysomes maintain the high local CO_2 concentrations around the RuBisCO enzymes necessary for adequate performance of these otherwise inefficient enzymes. Consistent with the idea that the carboxysome shell acts as a barrier against the permeation of CO_2, and that of O_2, the

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

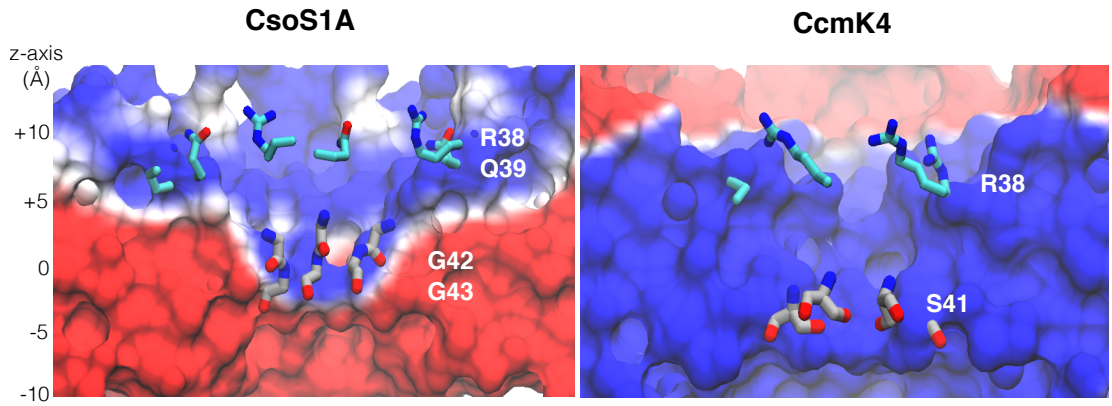

Figure 7: Electrostatic potential maps of the shell proteins. The maps were calculated using PMEpot plugin of VMD [24, 53], which uses the particle mesh Ewald method [54]. The ensemble of protein conformations collected in every 10 ps was used in each calculation. Electrostatic potentials shown in this figure are with threshold ranging from -20 (red) to +20 (blue) kT/e. "positive" z is towards the concave surface, while "negative" z is towards the convex surface.

experimental study by Dou et al [55] on *H. neapolitanus* suggests that the CO_2 supply into the carboxysome is provided mainly thorugh the entry of HCO_3^- molecules. The activity of freely soluble RuBisCO enzymes was compared to the activity of the enzymes contained inside the intact carboxysome. The intact carboxysome devoid of carbonic anhydrase showed a 3-fold increase in K_m of CO_2 with no change in V_{max}, while the ruptured carboxysome and carboxysome-free RuBisCO enzymes showed similar K_m and V_{max} to the wild type. As a consequence, they found that *H. neapolitanus* mutants lacking carbonic anhydrase required elevated CO_2 to grow.

The observed substrate selectivity appears to originate from electrostatic properties of the central pores. Positive electrostatic potentials along the central pores establish strong binding affinities for negatively charged molecules, such as HCO_3^- and Cl^- ions. This notion is supported by the calculated favorable insertion free energies of HCO_3^- from the umbrella sampling simulations and the captures of spontaneous Cl^- binding during the equilibrium simulations. Favorable binding of HCO_3^- facilitates its passage through the pore.

Acknowledgements

We thank Dr. Catherine Oikonomou for help revising the manuscript. This work was supported in part by the National Institutes of Health (NIH P41-GM104601, U01-GM111251 and U54-GM087519 to E.T.), the Office of Naval Research (ONR N00014-16-1-2535 to E.T.),

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and the Howard Hughes Medical Institute (to G.J.J.). P.M. gratefully acknowledges a previous support as a trainee of the Molecular Biophysics Training Program by the NIH (T32-GM008276) during his graduate study.

References

[1] H. Bauwe, M. Hagemann, A. R. Fernie, Photorespiration: players, partners and origin, Trends in Plant Science 15 (2010) 330–336.

[2] J. N. Kinney, S. D. Axen, C. A. Kerfeld, Comparative analysis of carboxysome shell proteins, Photosyn. Res. 109 (2011) 21–32.

[3] B. D. Rae, B. M. Long, M. R. Badger, G. D. Price, Functions, compositions, and evolution of the two types of carboxysomes: Polyhedral microcompartments that facilitate CO_2 fixation in cyanobacteria and some proteobacteria, Microbiol. Mol. Biol. Rev. 77 (2013) 357–379.

[4] M. van Lun, J. S. Hub, D. van der Spoel, I. Andersson, CO_2 and O_2 distribution in Rubisco suggests the small subunit functions as a CO_2 reservoir, J. Am. Chem. Soc. 136 (2014) 3165–3171.

[5] A. J. Bloom, Photorespiration and nitrate assimilation: a major intersection between plant carbon and nitrogen, Photosyn. Res. 123 (2015) 117–128.

[6] M. R. Badger, T. J. Andrews, S. M. Whitney, M. Ludwig, D. C. Yellowlees, W. Leggat, G. D. Price, The diversity and coevolution of Rubisco plastids, pyrenoids, and chloroplast-based CO_2-concentrating mechanisms in algae, Can. J. Biochem. 76 (1998) 1052–1071.

[7] G. G. B. Tcherkez, G. D. Farquhar, T. J. Andrews, Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized, Proc. Natl. Acad. Sci. USA 19 (2006) 7246–7251.

[8] L. Whitehead, B. M. Long, G. D. Price, M. R. Badgers, Comparing the in Vivo Function of α-Carboxysomes and β-Carboxysomes in Two Model Cyanobacteria, Plant Physiol. 165 (2014) 398–411.

[9] J. M. Shively, F. Ball, D. H. Brown, R. E. Saunders, Functional Organells in Prokaryotes: Polyhedral Inclusions (Carboxysomes) of *Thiobacillus neapolitanus*, Science 182 (1973) 584–586.

[10] A. Kaplan, L. Reinhold, CO_2 concentrating mechanism in photosynthetic microorganisms, Annu. Rev. Plant Physiol. Plant Mol. Biol. 50 (1999) 537–570.

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[11] M. R. Badger, G. D. Price, CO_2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution, J. Exp. Bot. 54 (2003) 609–622.

[12] J. C. Cameron, S. C. Wilson, S. L. Bernstein, C. A. Kerfeld, Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway, Cell 155 (2013) 1131–1140.

[13] G. D. Price, Inorganic carbon transporters of the cyanobacteria CO_2 concentrating mechanism, Photosyn. Res. 109 (2011) 47–57.

[14] G. C. Cannon, S. Heinhorst, C. A. Kerfeld, Carboxysomal carbonic anhydrases: Structure and role in microbial CO_2 fixation, Biochim. Biophys. Acta – Proteins and Proteomics 1804 (2010) 382–392.

[15] M. F. Schmid, A. M. Paredes, H. A. Khant, F. Soyer, H. C. Aldrich, W. Chiu, J. M. Shively, Structure of *Halothiobacillus neapolitanus* carboxysomes by cryo-electron tomography, J. Mol. Biol. 364 (2006) 526–535.

[16] C. V. Iancu, H. J. Ding, D. M. Morris, D. P. Dias, A. D. Gonzales, A. Martino, G. J. Jensen, The structure of isolated *Synechococcus* strain WH8102 carboxysomes as revealed by electron cryotomography, J. Mol. Biol. 372 (2007) 764–773.

[17] S. Tanaka, C. A. Kerfeld, M. R. Sawaya, F. Cai, S. Heinhorst, G. C. Cannon, T. O. Yeates, Atomic-Level Models of the Bacterial Carboxysome Shell, Science 319 (2008) 1083–1086.

[18] W. Bonacci, P. K. Teng, B. Afonso, H. Niederrholtmeyer, P. Grob, P. A. Silver, D. F. Savage, Modularity of a carbon-fixing protein organelle, Proc. Natl. Acad. Sci. USA 109 (2012) 478–483.

[19] T. O. Yeates, C. Kerfeld, S. Heinhorst, G. C. Cannon, J. M. Shively, Protein-based organelles in bacteria: carboxysomes and related microcompartments, Nat. Rev. Microbiol. 6 (2008) 681–691.

[20] Y. Tsai, M. R. Sawaya, G. C. Cannon, F. Cai, E. B. Williams, S. Heinhorst, C. A. Kerfeld, T. O. Yeates, Structural analysis of CsoS1A and the protein shell of the *Halothiobacillus neapolitanus* carboxysome, PLoS Biol. 5 (2007) e144.

[21] C. A. Kerfeld, M. R. Sawaya, S. Tanaka, C. V. Nguyen, M. Phillips, M. Beeby, T. O. Yeates, Protein structures forming the shell of primitive bacterial organelles, Science 309 (2005) 936–938.

[22] B. D. Rae, B. M. Long, M. R. Badger, G. D. Price, Structural Determinants of the Outer Shell of β-Carboxysomes in *Synechococcus elongatus* PCC 7942: Roles for CcmK2, K3-K4, CcmO and CcmL, PLoS One 7 (2012) e43871.

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[23] F. Cai, M. Sutter, S. L. Bernstein, J. N. Kinney, C. A. Kerfeld, Engineering bacterial microcompartment shells: Chimeric shell proteins and chimeric carboxysome shells, ACS Synt. Biol. 4 (2014) 444–453.

[24] W. Humphrey, A. Dalke, K. Schulten, VMD – Visual Molecular Dynamics, J. Mol. Graphics 14 (1) (1996) 33–38.

[25] L. Zhang, J. Hermans, Hydrophilicity of Cavities in Proteins, Proteins: Struct., Func., Gen. 24 (1996) 433–438.

[26] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.

[27] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, Scalable Molecular Dynamics with NAMD, J. Comp. Chem. 26 (2005) 1781–1802.

[28] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. W. E. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins., J. Phys. Chem. B 102 (1998) 3586–3616.

[29] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. MacKerell, Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles, J. Chem. Theory Comput. 8 (9) (2012) 3257–3273.

[30] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes, J. Comp. Phys. 23 (1977) 327–341.

[31] T. Darden, D. York, L. G. Pedersen, Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems, J. Chem. Phys. 98 (12) (1993) 10089–10092.

[32] G. J. Martyna, D. J. Tobias, M. L. Klein, Constant Pressure Molecular Dynamics Algorithms, J. Chem. Phys. 101 (5) (1994) 4177–4189.

[33] S. E. Feller, Y. Zhang, R. W. Pastor, B. R. Brooks, Constant pressure molecular dynamics simulation: The Langevin piston method, J. Chem. Phys. 103 (11) (1995) 4613–4621.

[34] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell, Jr., CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields, J. Comp. Chem. 31 (4) (2010) 671–690.

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[35] S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, K. Schulten, Steered Molecular Dynamics, in: P. Deuflhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, R. D. Skeel (Eds.), Computational Molecular Dynamics: Challenges, Methods, Ideas, vol. 4 of *Lecture Notes in Computational Science and Engineering*, Springer-Verlag, Berlin, 39–65, 1998.

[36] G. M. Torrie, J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella Sampling, J. Comp. Phys. 23 (1977) 187–199.

[37] B. Roux, The calculation of the potential of mean force using computer simulations, Comput. Phys. Commun. 91 (1995) 275–282.

[38] D. Frenkel, B. Smit, Understanding Molecular Simulation From Algorithms to Applications, Academic Press, California, 2002.

[39] M. Faulkner, J. Rodriguez-Ramos, G. F. Dykes, S. V. Owen, S. Casella, D. M. Simpson, R. J. Beynon, L.-N. Liu, Direct characterization of the native structure and mechanics of cyanobacterial carboxysomes, Nanoscale 9 (2017) 10662–10673.

[40] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, J. M. Rosenberg, The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method, J. Comp. Chem. 13 (1992) 1011–1021.

[41] O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, M. S. P. Sansom, HOLE: A Program for the Analysis of the Pore Dimensions of Ion Channel Structural Models, J. Mol. Graphics 14 (1996) 354–360.

[42] S. Tanaka, M. R. Sawaya, M. Phillips, T. O. Yeates, Insights from multiple structures of the shell protein from the β-carboxysome, Prot. Sci. 18 (2009) 108–120.

[43] H. Frauenfelder, S. G. Sligar, P. G. Wolynes, The Energy Landscapes and Motions of Proteins, Science 254 (1991) 1598–1603.

[44] A. Ostermann, R. Waschipky, F. G. Parak, G. U. Nienhaus, Ligand binding and conformational motions in myoglobin, Nature 404 (2000) 205–208.

[45] J. Saam, I. Ivanov, M. Walther, H. Holzhutter, H. Kuhn, Molecular dioxygen enters the active site of 12/15 lipoxygenase via dynamic oxygen access channels, Proc. Natl. Acad. Sci. USA 104 (2007) 13319–13324.

[46] P. Mahinthichaichan, R. B. Gennis, E. Tajkhorshid, Bacterial denitrifying nitric oxide reductases and aerobic respiratory terminal oxidases use similar delivery pathways for their molecular substrates, Biochim. Biophys. Acta – Bioener. .

[47] H.-X. Huang, J. A. McCammon, The gates of ion channels and enzymes, Trends Biochem. Sci. 35 (2010) 179–185.

bioRxiv preprint doi: https://doi.org/10.1101/367714; this version posted July 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[48] A. Gora, J. Brezovsky, J. Damborsky, Gates of enzymes, Chem. Rev. 113 (2013) 5871–5923.

[49] K. Henzler-Wildman, D. Kern, Dynamic personalities of proteins, Nature 450 (2007) 964–972.

[50] P. J. Focke, X. Wang, H. P. Larsson, Neutrotransmitter Transporters: Structure Meets Function, Structure 21 (2013) 694–705.

[51] M. Moradi, G. Enkavi, E. Tajkhorshid, Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate transporter, Nat. Commun. 6 (2015) 8393.

[52] J. Henin, E. Tajkhorshid, K. Schulten, C. Chipot, Diffusion of glycerol through *Escherichia coli* aquaglyceroporin GlpF, Biophys. J. 94 (2008) 832–839.

[53] A. Aksimentiev, K. Schulten, Imaging Alpha-Hemolysin with Molecular Dynamics: Ionic Conductance, Osmotic Permeability and the Electrostatic Potential Map, Biophys. J. 88 (2005) 3745–3761.

[54] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, A smooth particle mesh Ewald method., J. Chem. Phys. 103 (1995) 8577–8593.

[55] Z. Dou, S. Heinhorst, E. B. Williams, C. D. Murin, J. M. Shively, G. C. Cannon, CO_2 fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusion barrier for CO_2, J. Biol. Chem. 283 (2008) 10377–10384.