THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

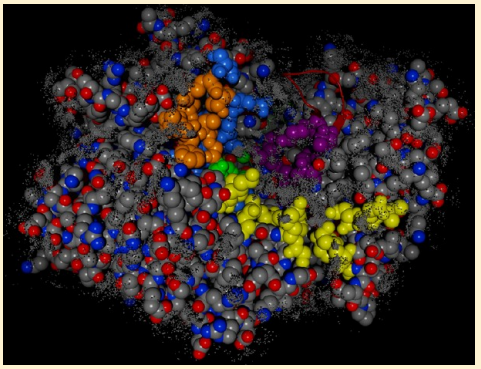
Nature of Protein–CO_2 Interactions as Elucidated via Molecular Dynamics

Michael L. Drummond,* Angela K. Wilson, and Thomas R. Cundari

Department of Chemistry, Center for Advanced Scientific Computing and Modeling (CASCaM), University of North Texas, Denton, Texas 76201, United States

Supporting Information

ABSTRACT: Rising global temperatures require innovative measures to reduce atmospheric concentrations of CO_2. The most successful carbon capture technology on Earth is the enzymatic capture of CO_2 and its sequestration in the form of glucose. Efforts to improve upon or mimic this naturally occurring process will require a rich understanding of protein–CO_2 interactions. Toward that end, extensive all-atom molecular dynamics (MD) simulations were performed on the CO_2-utilizing enzyme phosphoenolpyruvate carboxykinase (PEPCK). Preliminary simulations were performed using implicit and explicit solvent models, which yielded similar results: arginine, lysine, tyrosine, and asparagine enhance the ability of a protein to bind carbon dioxide. Extensive explicit solvent simulations were performed for both wild-type PEPCK and five single-point PEPCK mutants, revealing three prevalent channels by which CO_2 enters (or exits) the active site cleft, as well as a fourth channel (observed only once), the existence of which can be rationalized in terms of the position of a key Arg residue. The strongest CO_2 binding sites in these simulations consist of appropriately positioned hydrogen bond donors and acceptors. Interactions between CO_2 and both Mn^{2+} and Mg^{2+} present in PEPCK are minimal due to the stable protein- and solvent-based coordination environments of these cations. His 232, suggested by X-ray crystallography as being a potential important CO_2 binding site, is indeed found to be particularly "CO_2-philic" in these simulations. Finally, a recent mechanism, proposed on the basis of X-ray crystallography, for PEPCK active site lid closure is discussed in light of the MD trajectories. Overall, the results of this work will prove useful not only to scientists investigating PEPCK, but also to groups seeking to develop an environmentally benign, protein-based carbon capture, sequestration, and utilization system.



1. INTRODUCTION

It was recently proposed[1] that an atmospheric CO_2 concentration of less than 350 ppm must be maintained to avoid catastrophic, irreversible damage due to global climate change. While the precision of this specific target can be debated—in fact, this calamitous threshold has already been crossed, based on the most recent measurements of atmospheric CO_2 levels (396 ppm)[2]—the need to develop a mechanism to ameliorate rising quantities of CO_2 in the air is certainly urgent.[3] The most widely used framework on Earth for removing atmospheric carbon dioxide is through enzymatic fixation; one protein alone, ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO), fixes on the order of 10^{11} tons of CO_2 each year.[4] In addition to its role in mitigating atmospheric levels of CO_2, RuBisCO is also responsible for providing energy for all nonphotosynthetic life on Earth, as the captured CO_2 is ultimately sequestered through the Calvin cycle in the form of glucose. However, despite its unquestioned utility and impressive chemistry, the poor catalytic efficiency of RuBisCO (~1–10 CO_2 fixation events/second)[5] renders it ill-suited to play a larger role in reducing the concentration of atmospheric CO_2. Nevertheless, RuBisCO clearly demonstrates that a protein-based platform can successfully capture and utilize CO_2; indeed, Siemens Energy has recently announced a conceptually similar aqueous amino acid salt solution with a claimed CO_2 capture efficiency of over 90% on a pilot-plant scale.[6] To efficiently develop further advances in protein-based CO_2 capture technologies, an in-depth understanding of protein–CO_2 interactions is required.

We have been developing this understanding[7−10] through a multiscale, in silico approach, ranging from high-accuracy calculations of just a few atoms to a bioinformatics study of dozens of proteins and hundreds of protein–CO_2 interactions. In the process, we have discovered that CO_2 exhibits a strong preference to bind to proteins at basic residues (Arg, Lys, His) and near β-sheets rather than α-helices;[7] this latter phenomena likely arises from the presence of hydrogen bond donors along the β-sheet edges.[9] Additionally, it appears that proteins unrelated in terms of both structure and function utilize similar three-dimensional patterns of hydrogen bond donors and

Received: May 16, 2012
Revised: July 26, 2012
Published: August 10, 2012

acceptors to bind CO_2; these patterns, which are akin to pharmacophoric patterns used in drug discovery, can also be used to detect likely CO_2 binding sites on uncharacterized proteins.[8] In the current work, we seek to add a time-resolved element to the description of protein−CO_2 interactions by expanding upon our previous work[10] utilizing extensive molecular dynamics (MD) simulations.

The protein architecture explored with MD herein is not RuBisCO due to both its aforementioned poor catalytic activity and its large size: the active form is a hexadecamer of protein subunits weighing a total of over 500000 g mol^{-1}. Instead, the interactions of CO_2 with phosphoenolpyruvate carboxykinase (PEPCK) will be discussed. PEPCK, also referred to in the literature as PCK, is an enzyme found in nearly every species that catalyzes the reversible addition (Figure 1) of a molecule of

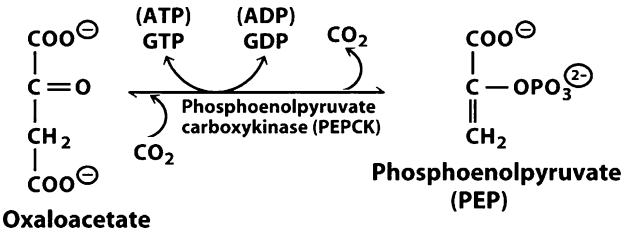


Figure 1. Reversible reaction catalyzed by PEPCK.

CO_2 to phosphoenolpyruvate (PEP), cleaving the phosphoryl group of PEP and generating oxaloacetate (OAA).[11] The direction where PEP is enzymatically formed from OAA represents a key early step in the formation of glucose from noncarbohydrate molecules (gluconeogenesis). PEPCK enzymes can be divided into two categories, depending on the nucleotide utilized: ATP-dependent and GTP-dependent. While these two categories differ greatly at the primary sequence level, within each category there is a great deal of structural similarity;[12] there is also some correspondence across the two categories among the active site residues.[13] Optimal catalysis in PEPCK requires two metal ions.[14]

PEPCK represents an ideal framework for studying protein−CO_2 interactions for a number of reasons. First, its presence in all species can be exploited to analyze common (as well as disparate) binding features found across the biosphere (although that task is beyond the scope of the current work, which focuses on the ATP-dependent enzyme from *Escherichia coli*). Second, in addition to its established biological role in gluconeogenesis, PEPCK has been implicated as a drug target due to its role in type 2 diabetes.[15] Third, it has been suggested[16] that modifying PEPCK to more efficiently form OAA from PEP could lead to a bioengineered source of succinic acid, a derivative of OAA with a wealth of current and future industrial uses in diverse fields such as pharmaceuticals, plastics, and detergents.[17,18] Finally, there is a crystal structure that shows the PEPCK−CO_2 interaction at atomic resolution,[19] one of only a handful of examples of proteins characterized with a molecule of bound carbon dioxide.[8] For these reasons, PEPCK thus serves as an ideal pilot study platform for the computational investigation of protein−CO_2 interactions.

In addition to a crystal structure, there is a wealth of additional experimental data against which the present MD simulations can be compared, both to ensure the validity of the computational protocol and to clarify, verify, or extend experimental results. Experimentally studied PEPCK enzymes fall into both the ATP- and GTP-dependent categories and have been studied for species such as yeast,[16,20−25] rat,[26−29] and various bacteria.[11−14,16,19,30−35] These diverse experimental studies have yielded valuable insight into the interactions of PEPCK with metal ions, substrates (both PEP and OAA), inhibitors, and nucleotides and have also elucidated the chemistry of phosphoryl transfer. However, the interaction of PEPCK with CO_2 has not been examined in nearly as much detail in these experiments. In one example,[12] it was found that CO_2 binding is reduced by three-and-a-half times with a single site mutation in the PEPCK of *Mycobacterium smegmatis*. Although the amino acid mutated in the GTP-dependent enzyme of *M. smegmatis* does not have an analogue in the ATP-dependent *E. coli* PEPCK studied in the current work, this study clearly establishes that changing even a single amino acid can have a large impact on carbon dioxide binding, which has obvious ramifications for bioengineered CO_2 capture platforms. More common, unfortunately, is the situation where a mutation has little to no effect on the CO_2 binding capability of PEPCK.[20−22,33] Experiments have also determined that the Mn^{2+} cofactor of PEPCK does not interact with CO_2,[32,36,37] despite the fact that the interaction of CO_2 with divalent cations has been calculated to have a strength of over 60 kcal/mol;[7] this lack of CO_2−Mn^{2+} interaction has also been noted in mammalian G-protein-responsive adenylyl cyclase.[37] One final experimental finding of relevance to PEPCK−CO_2 interactions suggests that, in addition to the Arg residue (Arg 65) known to bind CO_2 in the *E. coli* PEPCK crystal structure,[19] Arg 327 in *Anaerobiospirillum succiniciproducens* (equivalent to Arg 333 in *E. coli* PEPCK) may also play a role in CO_2 binding.[30] This potential role of Arg 333 as a secondary CO_2 binding site will be examined using MD in the current work.

Two MD approaches were applied to investigate PEPCK−CO_2 interactions: one with a solvent model that implicitly accounts for water molecules in an approximate (but rapid) fashion, and one that explicitly incorporates tens of thousands of water molecules in the MD simulation. After elaborating on the computational protocols used, simulations where 18 molecules of CO_2 simultaneously interact with *E. coli* PEPCK will be discussed. Next, simulations illustrating the exit of CO_2 from the known crystallographic binding site will be detailed, both for wild-type PEPCK as well as for five single-point mutant forms of the enzyme. Finally, the information contained in the MD simulations will be used to address some unresolved issues in the experimental PEPCK literature, including the role of an active site lid in achieving optimal enzymatic efficiency.[26] Overall, the results of this work will offer guidance to future efforts designing bioinspired CO_2 capture and sequestration systems.[38]

2. COMPUTATIONAL METHODS

2.1. Implicit Solvent Model MD. All molecular dynamics simulations incorporating an implicit solvent model were performed using the distance-dependent dielectric model, as implemented in the Molecular Operating Environment (MOE) software package, version 2008.10.[39] Coordinates for the PEPCK of *E. coli* were taken from the published crystal structure (Protein Data Bank ID 2OLQ).[19] All crystallographically identified cofactors and substrates (i.e., Mn^{2+}, Mg^{2+}, ATP, and CO_2) were included in the simulation, but all explicit water molecules were removed. The AMBER94 force field[40] was used; any missing parameters were automatically assigned by MOE, using Gasteiger partial charges.[41] We have previously

demonstrated[9] that this protocol yields results for protein-CO_2 interactions similar to other force fields, such as AMBER99,[42] CHARMM27,[43−45] and the TraPPE force field, which was derived from simulations of bulk CO_2.[46] Hydrogen atoms were added with MOE, and their positions were optimized (with all heavy atoms fixed) using a multistep protocol: first, a steepest descent algorithm was used until the root-mean-square (rms) gradient was less than 1000 kcal mol^{-1} $Å^{-1}$, followed by a conjugate gradient algorithm until the rms gradient was less than 10 kcal mol^{-1} $Å^{-1}$, and finally a truncated Newton algorithm was used to bring the rms gradient below 10^{-4} kcal mol^{-1} $Å^{-1}$. After optimizing the hydrogen atoms, the procedure was repeated with all atoms free to move. The optimized enzyme was enclosed in a box-shaped, impassible potential barrier located no closer than 6 Å from the nearest atom along the *x*-, *y*-, and *z*-axes.

To simulate the interaction of PEPCK with free CO_2, eight molecules of CO_2 were added, one in each of the corners of this restraining box. In addition, the program Q-Sitefinder[47] was used to predict the ten most likely CO_2 binding sites in 2OLQ. The most likely binding site was predicted to be the crystallographic binding site, which already had a bound CO_2 from the crystal structure. A molecule of carbon dioxide was therefore added to each of the remaining nine predicted sites, yielding a total of 18 molecules of CO_2 in this PEPCK system. The positions of the backbone atoms of PEPCK were fixed, and all other atoms were optimized after CO_2 placement using the multistep procedure described above. These backbone atoms were also held in place during the MD simulation, and bond lengths to hydrogen were kept constant.

The MD simulations were run in the canonical (NVT) ensemble. The Nosé−Poincaire−Anderson algorithm, as implemented in MOE,[48] was used to solve the equations of motion. The system was heated from 0 to 500 K in 1 ps using a 0.5 fs time step, and then held at 500 K for 15 ns; atomic positions were saved every 0.5 ps. The PEPCK−CO_2 interaction energy for each CO_2 molecule at each time step was calculated as the difference between the potential energy (i.e., the AMBER94 force field energy) of the CO_2−PEPCK system minus the sum of the isolated CO_2 molecule and the remaining PEPCK-17 CO_2 system, averaged over 250 timesteps to smooth out instantaneous variations. Preliminary MD simulations at 300 K showed little motion of the 10 CO_2 molecules placed by Q-Sitefinder, but large movement of the eight CO_2 molecules added to the corners of the restraining box. Elevating the temperature to 500 K enabled most (but not all, vide infra) of the 18 carbon dioxide molecules, even those placed by Q-Sitefinder, to explore the PEPCK environment.

2.2. Explicit Solvent MD. All MD simulations with explicit solvent were performed using the parallelized version of single-precision GROMACS 4.0.7.[49,50] As with the implicit solvent MD simulations, the coordinates of PEPCK, both metals, ATP, and CO_2 were taken from the published (2OLQ) crystal structure;[19] all crystallographic waters were removed. The AMBER94 force field, as implemented in GROMACS by Sorin et al.,[51,52] was used. The only required parameter missing from this force field was for Mn^{2+}, which was treated as a nonbonding ion with parameters adapted from the MM3 force field.[53,54] Topology files for ATP and CO_2 were generated using the acpype interface[55] to Antechamber along with the general AMBER force field (GAFF).[56,57] A total of 18 CO_2 molecules were placed in and around PEPCK at the same positions used above for the implicit solvent simulations. Hydrogen atoms were added to PEPCK using GROMACS; the enzyme was centered in a rhombic dodecahedron periodic cell with at least 10 Å of space between any protein atom and the cell boundary. The cell was then filled with Simple Point Charge (SPC) water molecules[58] and neutralized with randomly placed Na^+ and Cl^- ions to a concentration of 0.15 M. The system was optimized using a conjugate gradient algorithm, with a cycle of a steepest descent algorithm once every 10 steps, until the maximum force acting on any atom in the system was less than 10 kJ mol^{-1} nm^{-1} (or to within the extent allowed by the single precision implementation). The long-range electrostatic interactions were evaluated with a Particle Mesh Ewald algorithm[49] using a cubic interpolation scheme with the default grid spacing and a grid-based neighbor searching algorithm. Following geometry optimization of the system, all nonsolvent molecules were held fixed and the water molecules were allowed to settle around the PEPCK holoenzyme for 20 ps using a leapfrog MD algorithm under the canonical (NVT) ensemble, with a time step of 0.5 fs and the Berendsen thermostat[59] applied to maintain the temperature at 300 K. All bond lengths were held fixed during this water soaking phase via the LINCS algorithm.[60] Finally, the entire system was simulated using leapfrog algorithm-driven MD in the NVT ensemble for 15 ns, again with a time step of 0.5 fs. The SPC water molecules were held rigid with the SETTLE algorithm, but no other bond lengths were constrained.[61] All atomic positions were saved every 0.5 ps for further analysis. Interaction energies among the various components of the system (namely, protein, CO_2, ATP, Mn^{2+}, Mg^{2+}, and water) were calculated by summing the corresponding Lennard-Jones and short-range Coulombic interaction terms, as has been done elsewhere,[62] averaged over 250 timesteps to smooth out instantaneous variations. The temperature in the MD run was increased from 0 to 300 K in 1 ps and then held constant using a Nosé-Hoover thermostat.[63,64] It should be noted that, while preliminary simulations using the implicit solvent model at 300 K showed little movement of most CO_2 molecules, this limitation was not observed in the explicit solvent simulations, and thus, PEPCK was judged to be adequately explored by the 18 CO_2 molecules at 300 K. Moreover, while the protein backbone was significantly deformed if not held in place in the implicit solvent simulation above, this phenomenon was not seen in the explicit solvent simulations, and thus, no protein backbone constraints were included. Overall, the parameters and algorithms discussed above for the two MD solvation protocols were chosen to be similar between the two programs but with minor modifications included to render the explicit solvent simulation less artificial, and thus most comparable to experimental results, but still comparable to the implicit solvent simulation as well.

In addition to the PEPCK-18 CO_2 simulations just described, explicit solvent simulations with a single CO_2 molecule placed at the crystallographic binding site were performed for wild-type (WT) *E. coli* PEPCK, as well as five mutants: Arg 65 replaced by Gln (denoted Arg65Gln), which was also investigated experimentally[19] and found to have little effect on PEPCK's enzymatic activity, Tyr207Phe, Lys212Ala, and Lys213Ala, to remove functionality in the crystallographic binding site, and Glu311Ala, to remove a possible interaction between a negatively charged side chain and the partially positive carbon of CO_2, which was identified as potentially important in preliminary implicit solvent WT simulations. For the mutant simulations, the *E. coli* PEPCK crystal structure

2OLQ (with crystallographic waters removed) was modified at the corresponding amino acid residue, and then subjected to the optimization-water soak-full MD protocol described above. Four separate trajectories of 15 ns each were generated for each mutant, as well as for the WT, yielding 360 ns of independent simulations. Additionally, factoring in the two 15 ns simulations with 18 CO_2 molecules per PEPCK yields a total simulation time of 900 ns of protein−CO_2 interactions. A similar magnitude of MD simulation time was recently found necessary[65] to observe infrequent diffusion events of CO_2 across a biomolecular membrane and pore.

3. RESULTS AND DISCUSSION

3.1. Implicit Solvent Model MD of the PEPCK-18 CO_2 System. The 18 CO_2 molecules that interact with PEPCK using the implicit solvent description can be divided into three rough categories: those that stay bound in essentially one place for the entire 15 ns simulation; those that reside in a binding site for 3−10 ns, but also diffuse and explore other areas of the enzyme; and those that never reside in a single binding site for more than 1−2 ns. A typical member of each category is shown in Figure 2, where the jagged black line depicts the CO_2−PEPCK interaction energy (measured on the left vertical axis) and the amino acids neighboring the CO_2 molecule throughout the simulation are plotted (and listed on the right vertical axis), with thick dots representing interactions within 3 Å and smaller dots for interactions between 3 and 5 Å. For example, in Figure 2a, it can be seen (via the thick lines) that one particular CO_2 molecule binds primarily to Arg 65, Lys 70, Val 191, Glu 210, Lys 213, and Phe 413, with lesser, more distant interactions between CO_2 and Ser 66, Pro 67, Gly 214, and Ala 412. The PEPCK interaction energy for this CO_2 molecule is largely constant, fluctuating only by ∼3 kcal/mol. The CO_2 molecule described by Figure 2a was initially placed by Q-Sitefinder[47] 7.5 Å away from the crystallographic binding site, but well within the active site cleft of PEPCK, and stayed in essentially the same place throughout the entire 15 ns simulation. (Unexpectedly, the CO_2 molecule initially placed by Q-Sitefinder at the crystallographic binding site immediately migrated to another site ∼13 Å away and stayed there throughout the simulation; this finding will be discussed in more detail below, in the context of explicit solvent simulations.) At the other extreme, Figure 2c depicts the behavior of a molecule of carbon dioxide whose interaction energy with PEPCK varies from 0 to −16 kcal/mol, and which does not bind to any specific site for more 1−2 ns. The CO_2 molecule described by Figure 2b exhibits both behaviors: a series of short-lived binding sites in the first 10 ns (and which are generally weak, except for the site between 7 and 8 ns, which will be discussed later), followed by a longer-lived (5 ns) binding site described primarily by Lys91, Asn520, and Tyr524.

It is informative to consider the behavior of the 18 CO_2 molecules in aggregate, rather than individually. Furthermore, for this discussion, only the 8 carbon dioxides initially placed in the corners of the restraining potential box are considered, rather than the 10 initially placed near the protein by Q-Sitefinder, to address the question of how a *free* molecule of CO_2 diffusing in the vicinity of PEPCK can make its way into the active site interior of the enzyme. Including the 10 CO_2 molecules specifically placed in locations predicted by Q-Sitefinder to tightly bind biases the answer to this question. It is possible that a CO_2 placed by Q-Sitefinder, if stable, might prevent a free CO_2 from migrating toward this site, thus, underestimating slightly the CO_2 affinity of the residues in that site. However, consideration of 8 carbon dioxides simultaneously at 30000 individual snapshots will yield sufficient sampling to mitigate this possible oversight. In addition, the interaction of *all* 18 CO_2 molecules with PEPCK will be considered later in the context of the tightest binding sites.

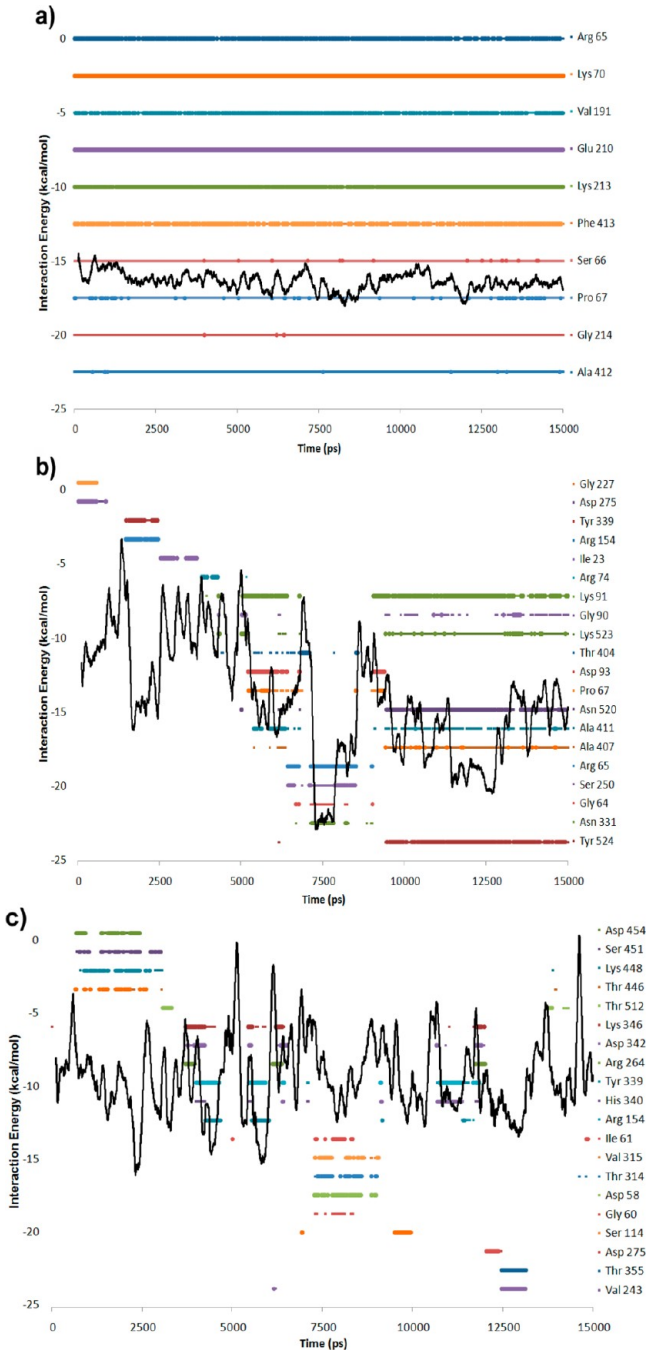


Figure 2. CO_2−PEPCK interaction energies (thick black line, left vertical axis) and neighboring amino acids (colored lines, right vertical axis) for three typical CO_2 molecules studied with the implicit solvent model. Thick points and thick lines correspond to distances within 3 Å, whereas thin points and lines correspond to distances of 3−5 Å. (a) CO_2 molecule that stays bound in one site throughout the entire simulation; (b) CO_2 molecule that briefly explores multiple binding sites on PEPCK before binding to one site for the final third of the simulation; (c) CO_2 molecule that does not bind for an appreciable length of time to any one site during the entire simulation.

dx.doi.org/10.1021/jp304770h | J. Phys. Chem. B 2012, 116, 11578−11593

To address how PEPCK can initially bind freely diffusing CO_2 molecules, each of PEPCK's 535 amino acid residues were assigned, at 0.5 ps intervals, a score of 1 if any amino acid atom came within 3 Å of one of the eight unbiased CO_2 molecules described above. The scores for each amino acid residue were summed across the entire 15 ns simulation, then grouped by amino acid identity, and finally divided by the total number of CO_2 contacts within 3 Å for all amino acids, thereby yielding the percentage of close contacts per amino acid. Finally, these percentages were weighted by population by subtracting the total relative population of each amino acid type in PEPCK from these percentages of close contacts. This final value is labeled "% enhancement" and is given in Table 1 for each of the 20 amino acids.

Table 1. Percent Enhancement for the Eight Unbiased CO_2 Molecules in the Implicit Solvent MD Simulation

residue	% close contacts	% population	% enhancement
Lys	18.89	6.36	12.54
Arg	13.54	4.11	9.43
Leu	10.72	8.60	2.12
Thr	9.92	8.41	1.51
Tyr	3.82	3.18	0.65
Asn	4.94	4.67	0.27
Trp	1.56	1.68	−0.12
His	1.22	1.68	−0.46
Met	1.23	1.87	−0.64
Cys	0.06	0.93	−0.87
Ile	4.04	5.05	−1.00
Gln	1.44	2.80	−1.36
Pro	3.57	5.05	−1.48
Asp	5.45	7.10	−1.66
Gly	6.77	8.79	−2.02
Phe	3.10	5.23	−2.13
Ser	2.18	4.67	−2.49
Glu	2.59	5.61	−3.01
Val	1.94	5.79	−3.85
Ala	2.99	8.41	−5.42

The enhancement values in Table 1 compare reasonably well to a previous bioinformatics study of crystallographically observed amino acid-CO_2 close interactions.[7] In that study, the basic amino acids Arg, His, and Lys, in that order, were found to be especially prevalent in the local environments of the carbon dioxide molecules bound to proteins in protein−CO_2 X-ray crystal structures. As in that study, Arg and Lys are found by the implicit solvent MD approach to be especially "CO_2-phillic." Also noteworthy is the enhancement shown by Thr, which we have previously suggested is an important amino acid in CO_2 migration channels.[10] By contrast, His binds CO_2 far less than found previously;[7] indeed, His shows a negative enhancement in Table 1, meaning that it binds CO_2 less often than would be expected based simply on its abundance in PEPCK. This poor performance of His may be a reflection of the low population of His in PEPCK (9 out of 535 residues), potentially leading to undersampling, particularly when compared to the populations of Arg and Lys in PEPCK (22 and 34, respectively). Another possible complicating factor for His is that it can assume three different protonation states: the imidazole side chain ring can have a hydrogen at the δ nitrogen, at the ε nitrogen, or at both, depending on the local environment. The locations of hydrogen atoms are not revealed by X-ray crystallography, and thus MOE automatically assigns a protonation state based on the local environment of each His residue, which may be incorrect, thereby potentially impacting a proper description of His−CO_2 binding. While a PEPCK−CO_2 crystal structure obtained with neutron diffraction techniques would allow for the precise assignment of His protonation states, no such structure exists.

Besides His, another notable discrepancy between the CO_2 binding affinities given in Table 1 and our previous analysis is for Leu. Although CO_2 can hydrogen bond to the backbone amine group of Leu,[9] only a dispersive interaction is possible between CO_2 and the aliphatic side chain of Leu. In contrast, the other five enhanced amino acids in Table 1 (Lys, Arg, Thr, Tyr, and Asn) all possess side chains capable of forming hydrogen bonds or interacting with CO_2 via a dipole−quadrupole interaction. Ala and Val, which (like Leu) possess solely aliphatic side chains, are the two amino acids least enhanced in Table 1; similarly, aliphatic Ile also underbinds CO_2 relative to its population, albeit to a lesser degree than do Ala and Val. The seemingly anomalous role in binding CO_2 exhibited by Leu may be related to the assumptions made in an implicit solvent model, and thus comparison with explicit solvent MD simulations can help shed light on the role of Leu in binding CO_2 to PEPCK.

3.2. Explicit Solvent MD of the PEPCK-18 CO_2 System.

As in the implicit solvent MD simulation, the 18 molecules of carbon dioxide in the explicit solvent simulation can be divided into three groups: those that are bound at one site throughout the entire simulation; those that explore numerous sites, but very briefly (1−2 ns each); and those intermediate, that explore more than one protein binding site, spending up to 3−5 ns at each site. Typical examples of each are given in Figure SI1 in Supporting Information. Comparison of Figure SI1 with Figure 2 for the implicit solvent simulations shows that, overall, the CO_2 molecules explore fewer protein-based binding sites when explicit water molecules are present, as might be expected due to competition with water. Nevertheless, it is possible, as above, to calculate the percent enhancement values for each amino acid type based on the behavior of the eight unbiased CO_2 molecules, and this data is shown in Table 2.

As in Table 1, the basic amino acids Arg and Lys show the greatest percent enhancement, although Arg is the most enhanced in the explicit solvent simulation (in agreement with available experimental data),[7] whereas Lys showed the most enhancement in the implicit solvent simulation. Also in agreement with the implicit solvent simulation, His binds CO_2 much more poorly than would be expected based on the previous bioinformatics study.[7] The algorithms used to assign His protonation states in the implicit and explicit solvent simulation software packages are different, suggesting that the poor CO_2 binding behavior of His may not be an artifact of this procedure. It appears, therefore, that the *E. coli* PEPCK enzyme studied in the current work represents an outlier when compared to other protein−CO_2 crystal structures, insofar as PEPCK does not seem to utilize its His residues to bind CO_2 with the same frequency as do other proteins.

In addition to the basic residues Arg and Lys, Tyr and Asn are also found to have enhanced binding of carbon dioxide in both simulations. Indeed, the agreement between the data in Table 1 and the data in Table 2 is generally quite good (correlation coefficient of % close contacts = 0.87). The difference in percent enhancement between the two tables is shown in the far right column of Table 2, and differences

Table 2. Percent Enhancement for the Eight Unbiased CO_2 Molecules in the 18 CO_2 Explicit Solvent MD Simulation[a]

residue	% close contacts	% enhancement	difference from Table1[b]
Arg	11.09	6.98	−2.45%
Lys	12.34	5.98	*−6.55%*
Asn	7.39	2.72	2.45%
Pro	7.38	2.34	*3.81%*
Tyr	4.46	1.28	0.63%
Thr	8.39	−0.02	−1.53%
Asp	7.04	−0.06	1.60%
Leu	8.49	−0.11	−2.23%
Val	5.45	−0.35	*3.51%*
Trp	1.00	−0.68	−0.56%
Cys	0.01	−0.93	−0.06%
His	0.38	−1.30	−0.84%
Ser	3.27	−1.40	1.09%
Gly	7.32	−1.47	0.55%
Ile	3.48	−1.57	−0.56%
Glu	4.00	−1.61	1.40%
Gln	1.02	−1.78	−0.42%
Met	0.01	−1.87	−1.23%
Ala	6.30	−2.12	*3.30%*
Phe	1.19	−4.04	−1.91%

[a]See Table 1 for the % population data. [b]The % enhancement values in this Table minus the corresponding data in Table 1. Italicized values differ by more than 3%

greater than 3% are italicized. Only four values differ by more than this threshold: Lys, which shows less (but still highly enhanced) CO_2 binding ability in the explicit solvent simulation; and Pro, Val, and Ala, which all show greater CO_2 binding ability in the explicit solvent simulation (although of these three, only Pro shows true enhancement relative to its population in PEPCK). Taken together, these four differences indicate greater participation of the hydrophobic residues in CO_2 binding in the explicit solvent simulation, a reflection of *decreased* participation of the *hydrophilic* residues in CO_2 binding due to the more realistic description of the interaction between explicit water molecules and hydrophilic amino acids. Leu now shows negative enhancement, as is observed for the similar aliphatic amino acids Ala, Val, and Ile. Because the balance in CO_2 binding capabilities between the hydrophobic and hydrophilic amino acids is more accurately described in the explicit solvent model simulation, all future results presented, unless otherwise noted, will be based on explicit solvent simulations. However, the fact that the more rapid implicit solvent model generally agrees with the explicit solvent simulations suggests that future insight into protein−CO_2 binding may be gained efficiently by using a more advanced implicit solvent model.[66]

3.3. CO_2 Entrance/Exit Channels. The previous sections detailed which amino acids are used by PEPCK (and thus presumably by other proteins as well) to capture CO_2 diffusing in the nearby environment and concentrate it in the vicinity of PEPCK via protein−CO_2 interactions. These surface-based CO_2 binding sites have previously been labeled as "reservoirs" of CO_2.[67] However, for PEPCK to catalyze the chemistry shown in Figure 1, it is necessary for CO_2 to move from the enzyme surface reservoirs into the active site cleft, a process that has been suggested to occur in the enzyme carbonic anhydrase through both simple diffusion[67] and via a protein-based carbon dioxide relay.[68] As we have previously shown,[10] MD demonstrates that CO_2 preferentially enters (and leaves) the active site through a limited number of pathways, suggesting a *directed* relay mechanism for CO_2 transfer rather than passive diffusion.

Of the 18 CO_2 molecules modeled in the explicit solvent model PEPCK simulation (section 3.2), none made a transition either to or from the crystallographic binding site, because the molecule of CO_2 initially placed there by Q-Sitefinder stayed bound throughout. However, as mentioned in section 3.1, the molecule of carbon dioxide placed at the crystallographic binding site in the *implicit* solvent model simulation diffused to a site ∼13 Å distant, thereby freeing the crystallographic binding site to bind another CO_2 molecule. Throughout the 15 ns implicit solvent simulation, three separate molecules of CO_2 diffused into the active site cleft and bound to the crystallographic site. The path of the first molecule to do so is presented in Figure 2b, which shows that this CO_2 molecule is found within 3 Å of Arg 65, the primary crystallographic binding site residue,[19] between 6400 and 9000 ps (albeit not quite continuously). This molecule leaves the binding site and is replaced by another between 9000 and 12000 ps; the path of this second molecule of CO_2 is shown in Figure 3a. After this molecule leaves the binding site at 12000 ps, a third CO_2 molecule occupies the binding site (Figure 3b) until the end of the simulation. All told, the crystallographic binding site is occupied for more than half of the implicit solvent simulation.

A closer inspection of Figures 2b and 3a,b reveals that all three CO_2 molecules approach the crystallographic binding site by the same route. For example, at approximately 5000 ps in Figure 2b, CO_2 is within 3 Å of Pro 67, Gly 90, Lys 91, Asp 93, and Ala 411. The interaction between CO_2 and PEPCK then becomes less favorable as the CO_2 moves first toward Ser 250 and then Thr 404, but becomes much stronger as the CO_2 binds to Arg 65 and Asn 331. This strong interaction is weakened as the CO_2 molecule again explores the vicinity of Ser 250, followed by a brief time period of close interaction with only Thr 404. CO_2 then binds again near Pro 67, Lys 91, and Asp 93 before moving toward amino acids farther along in that direction (Figure 3c), such as Asn 520 and Tyr 524. Overall, the motion of CO_2 in Figure 2b follows the direction depicted in yellow in Figure 3c; strong PEPCK−CO_2 interactions, in particular, exist with Lys 91 and Asp 93 (but also with Pro 67 and Ala 411) just prior to and immediately after binding to Arg 65. The CO_2 molecule described by Figure 3a shows very similar behavior: close contacts with Lys 91, Asp 93, and Ala 411 exist prior to the very stable binding to Arg 65 and Asn 331 found between 4300 and 5000 ps. CO_2 then leaves the crystallographic binding site, again making close contact with Lys 91, and then binds relatively removed from the active site (as indicated by the role of Asn 520; see Figure 3c). Indeed, CO_2 nearly leaves the enzyme entirely, as indicated by its brief and weak interactions with Ile 518 and Asp 519, before returning once again to the active site at about 9000 ps. When it ultimately leaves the crystallographic binding site at about 12000 ps, it once more leaves via Lys 91, Asp 93, and Ala 411, which, as shown in Figure 3b, also guide the final CO_2 molecule to the crystallographic site.

To see if the CO_2 binding channel of Figure 3, defined primarily by Lys 91, Asp 93, and Ala 411 in the implicit solvent simulation, is also found in the explicit solvent simulations, four additional 15 ns MD runs were performed, starting from the crystal structure geometry; each therefore contained only a single molecule of CO_2. Of these four simulations, CO_2 twice

dx.doi.org/10.1021/jp304770h | J. Phys. Chem. B 2012, 116, 11578−11593

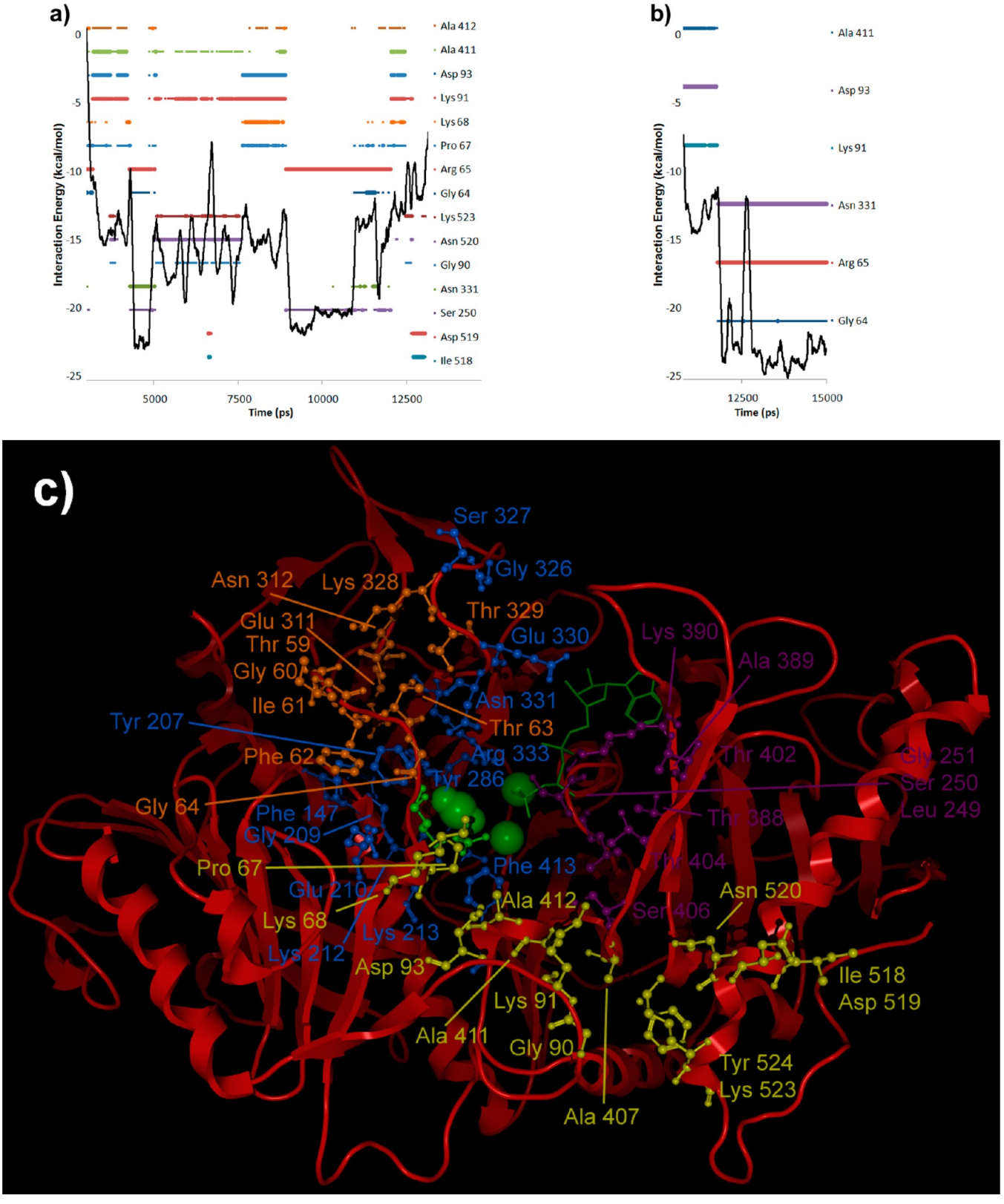


Figure 3. (a) Motion path of the second molecule of CO_2 to occupy the crystallographic binding site (indicated by Arg 65), following the CO_2 described by Figure 2b, during the implicit solvent simulation of PEPCK with 18 molecules of carbon dioxide; see the caption for Figure 2b for an explanation of the various lines in this figure. (b) Motion path of the third molecule of CO_2 to occupy the crystallographic binding site, following the CO_2 described in (a). (c) View of the backbone of PEPCK, with the amino acids of Figures 2b and 3a and b indicated in yellow, those of Figure 4 in purple, those of Figure 5 in blue, and those of Figure 6 in orange. Hydrogen atoms have been omitted. ATP is shown as a green wire structure, and Mg^{2+}, Mn^{2+}, and the crystallographic binding site of CO_2 (and the nearby Arg 65 residue) are shown for reference in green.

dx.doi.org/10.1021/jp304770h | J. Phys. Chem. B 2012, 116, 11578–11593

stayed entirely within the active site cleft (i.e., generally near the Arg 65, Tyr 207, and Lys 213 residues that define the crystallographically determined[19] binding site), and thus, these two simulations are not useful for locating potential CO_2 entrance/exit channels. In the other two simulations, however, CO_2 *does* leave the crystallographic binding site and via the same pathway each time, but this path differs dramatically from that shown in Figure 3. Figure 4a,b details the amino acids

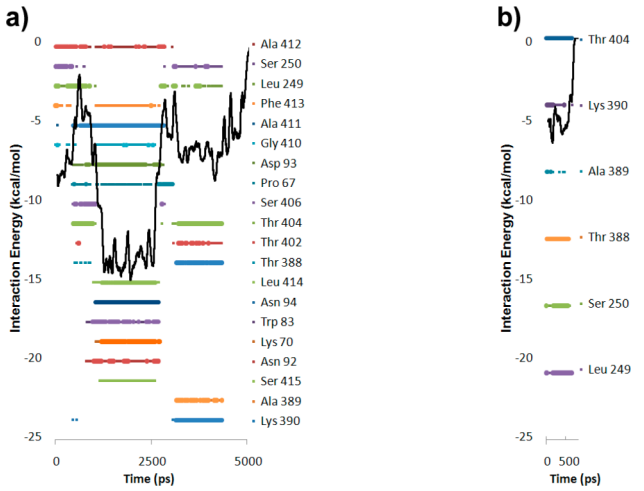


Figure 4. (a, b) Amino acids and protein−CO_2 interaction energies (see Figure 2) for the exit pathway directed toward the lid, observed in explicit solvent simulations where a single molecule of carbon dioxide is originally located at the crystallographic binding site. Amino acids involved in the exit of CO_2 are illustrated in purple in Figure 3c.

involved with this new CO_2 exit pathway, which is illustrated in purple in Figure 3c. In Figure 4a, the initial interaction with Arg 65 is so short-lived that it does not appear in this diagram, which illustrates only the 20 most prevalent CO_2−amino acid interactions in the interest of clarity. Rather than binding to Arg 65, the CO_2 in Figure 4a quickly moves to the vicinity of Leu 249, Ser 250, and Ala 412, followed by a transition to a binding site located near Lys 70, Trp 83, Asp 93, Asn 94, Ala 411, and Leu 414. This binding site is quite stable; moreover, it is located in the direction of the CO_2 entrance/exit channel described in Figure 3, as indicated by the presence of Asp 93 and Ala 411 in both pathways. However, rather than continuing along the yellow path of Figure 3c toward Asn 520 (as in Figures 2b and 3a), the CO_2 described by Figure 4a moves toward Thr 388, Ala 389, Lys 390, Thr 402, and Thr 404 before diffusing away from PEPCK entirely. These five amino acids, shown in purple in Figure 3c, flank a flexible lid segment of PEPCK, consisting of residues 391−401; this lid can be seen in Figure 3c as the loop above the green adenosine moiety of ATP. More will be said about this lid later, but for now it suffices to say that the lid is not closed when the CO_2 molecule of Figure 4a leaves PEPCK, nor does CO_2 contact these lid residues for an appreciable amount of time as it diffuses away from the enzyme. The exit path just described is nearly identical to the much more rapid exit depicted in Figure 4b, where once again the lid is not closed at the time of CO_2's exit.

3.4. CO_2 Entrance/Exit Channels in PEPCK Mutants. At this point, two pathways have been described: the one shown in Figure 3, defined by Lys 91, Asp 93, Ala 411, and so on, which was utilized three times in the implicit solvent simulation; and the one shown in Figure 4, defined by the amino acids flanking the lid of PEPCK, which was observed in two (out of four) explicit solvent simulations, where a single CO_2 was placed at the crystallographic binding site at the start of the simulation. It is therefore tempting to state that the two different pathways are a result of the two different models used to treat solvent effects. However, undersampling (and overinterpretation) is the bête noire of MD, due to the inherently stochastic nature of the simulations. As an illustration, a recent MD study[65] of CO_2 permeation through protein channels and lipid membranes found that one diffusion event was observed only twice in 240 ns. In other words, it is premature to conclude that the two CO_2 paths described thus far result from differences in the solvation model, because sampling has not yet been adequately addressed; additional simulations are therefore required. However, rather than simply repeat the simulations described above, trajectories were instead generated using mutants of PEPCK, where a single amino acid component of the crystallographic CO_2 binding site was modified (or, in the case of Glu 311, this modification was made to an amino acid near the crystallographic binding site). These mutant simulations enable simultaneous investigation of both potential CO_2 exit pathways and an evaluation of the relative roles of these amino acids in binding CO_2 within the active site cleft. Four independent 15 ns trajectories were generated, incorporating explicit water molecules, for each single-point PEPCK mutant, and the overall results are summarized in Table 3.

Table 3. [69] Summary of the Exit Pathways of CO_2 in the Explicit Solvent Mutant PEPCK Simulations

mutation[a]	run 1	run 2	run 3	run 4
none (WT)	Figure 4	bound	bound	Figure 4
Arg65Gln	Figure 3	Figure 3	bound	Figure 5
Tyr207Phe	Figure 5	Figure 3	Figure 3	bound
Lys212Ala	bound	Figure 5	bound	bound
Lys213Ala	Figure 4	Figure 3	bound	Figure 6
Glu311Ala	bound	Figure 5	bound	Figure 4

[a]Amino acid in unmutated PEPCK is given on the left and its replacement on the right. WT = wild-type (i.e., unmutated PEPCK).

It can be seen in Table 3 that the entrance pathway taken by CO_2 in the implicit solvent simulation, that is, Figure 3a,b, is observed in many of the mutant simulations, even though it was not seen in any wild-type (WT) explicit solvent simulation. Indeed, the exit path of Figure 3a,b is the most common route observed among all trajectories, occurring five times in the explicit solvent simulations in addition to the three events observed in the implicit solvent simulation. However, considering only the more realistic explicit solvent simulations, this pathway is not utilized by the WT. Moreover, the mutation site in Glu311Ala is over 8 Å distant from the CO_2 in the crystal structure of PEPCK, and so, with respect to the crystallographic binding site, these simulations can be considered as "pseudo-WT"; the pathway in Figure 3a,b is not utilized in these simulations, either. Thus, in effect, there are eight explicit solvent simulations where the crystallographically observed CO_2 binding site was left essentially unperturbed—that is, the first and last rows of Table 3—and the exit pathway of Figure 3 was not observed for any of these trajectories. Ultimately, the most certain way to determine if this pathway is relevant is through experimental work: for example, Lys 91 could be replaced and the kinetics of this mutant could be measured. Lys 91 is located on the opposite side of the active site cleft from

dx.doi.org/10.1021/jp304770h | J. Phys. Chem. B 2012, 116, 11578−11593

ATP, Mn^{2+}, and Mg^{2+}, so this change should not unduly influence the binding characteristics of these cofactors, but replacing Lys 91 with a bulky but chemically inert amino acid, such as Phe or Leu, should have a measurable effect on enzyme efficiency by obscuring CO_2's motion, if this migration channel is indeed relevant.

In addition to the channel described in Figure 4, which Table 3 indicates is utilized in four trajectories, there is another channel equally prevalent in the mutant simulations. Run 2 of Glu311Ala and Run 2 of Lys212Ala are two typical examples of this exit pathway, and the amino acids located near CO_2 as it exits via this channel are detailed in Figure 5a and b, respectively, and illustrated in blue in Figure 3c. This pathway is similar to that shown in Figure 4, but CO_2 exits along the opposite side of the cleft, rather than toward the base of the lid as was shown in Figure 4. Indeed, the lid is partially (i.e., not to a degree sufficient to block CO_2) closed between 3000 and 4500 ps for the Glu311Ala run, as indicated by the close contacts in Figure 5a between CO_2 and the lid residues Lys 390 and Ala 392. (The lid is not closed at all in the other three simulations [Table 3] described by Figure 5.) The primary amino acids involved in this exit pathway are Asn 331 and Arg 333, the last amino acids to make contact with CO_2 before it diffuses away from PEPCK in Figure 5b; the CO_2 described by Figure 5a passes Asn 331 and Arg 333, then Glu 330, and finally, Gly 326 and Ser 327, all of which lie along the same direction as it exits PEPCK.

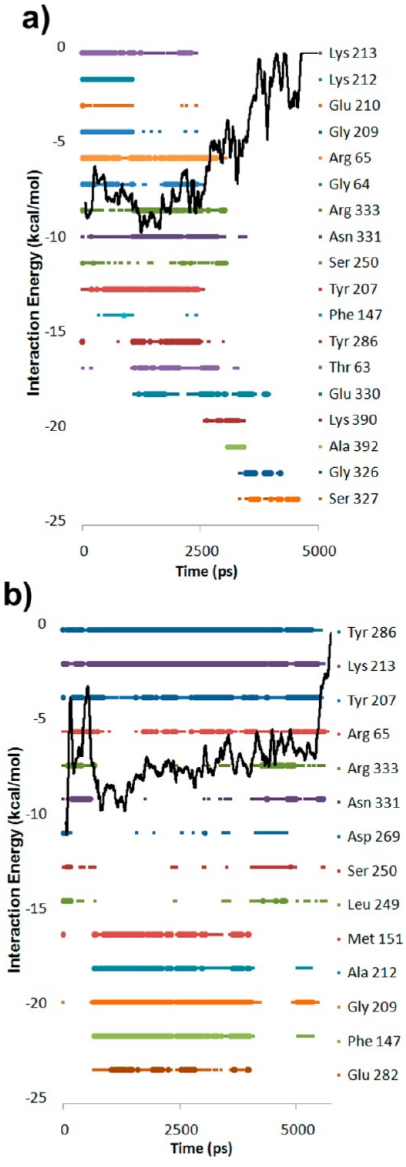


Figure 5. Amino acids and protein−CO_2 interaction energies (see Figure 2) for the exit pathway of (a) run 2 of the Glu311Ala mutant and (b) run 2 of the Lys212Ala mutant simulations. The amino acids involved in this exit pathway are illustrated in blue in Figure 3c.

The final CO_2 exit pathway found in the mutant simulations (Table 3) occurs only once, for run 4 of Lys213Ala. The amino acids involved in this channel are listed in Figure 6 and their location in PEPCK is shown in orange in Figure 3c. The CO_2 in this trajectory is bound for most of the simulation (albeit with gradually diminishing strength) not within the active site cleft of PEPCK, but rather within a smaller pocket just off the cleft defined primarily by two loops: Thr 59 to Gly 64 and Lys 328 to Asn 331. CO_2 diffuses out of this pocket on the face bordering the active site cleft, and hence, close contacts are seen with Gly 64 and Arg 65 in Figure 6 (but see below for details of the orientation of the latter), but it does not diffuse into the active site cleft before exiting PEPCK nor does it follow any of the above channels as it leaves.

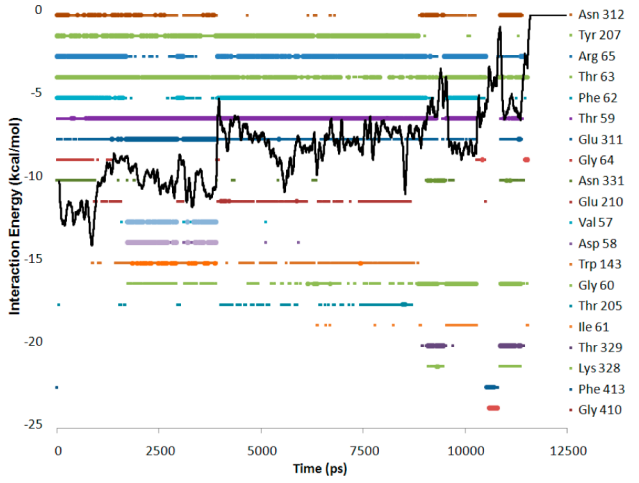


Figure 6. Amino acids and protein−CO_2 interaction energies (see Figure 2) for the unusual exit pathway seen in run 4 of the Lys213Ala mutant. These amino acids are indicated in orange in Figure 3c.

Two questions naturally arise concerning this CO_2 binding channel: why is this CO_2 exit channel observed only once, in a simulation of the Lys213Ala mutation, and is it relevant or merely a computational irregularity? To answer the first question, the positions of all α-carbons were averaged across all timesteps within each of the 24 trajectories in Table 3, these average Cα-only structures were aligned, and the root-mean-square difference (RMSD) of each compared to all others was then calculated. The full results are given as Figure SI2 in Supporting Information, but run 4 of Lys213Ala, which leads to the unusual CO_2 exit path of Figure 6, shows the greatest RMSD, yielding an average deviation of 2.52 Å when compared against all other trajectories. Thus, the protein backbone in this particular trajectory is the most dissimilar of all of the simulations listed in Table 3. Furthermore, this run shows the most movement of the main-chain atoms during the simulation. Figure 7a illustrates the RMSD between each time

dx.doi.org/10.1021/jp304770h | J. Phys. Chem. B 2012, 116, 11578−11593

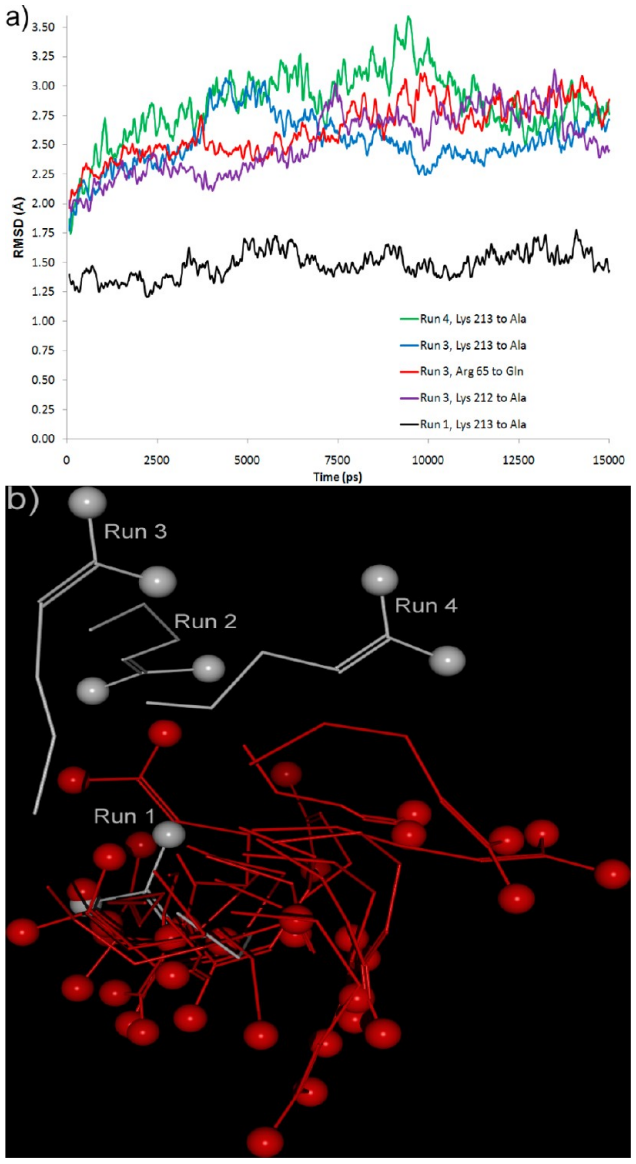


Figure 7. (a) RMSD of the main-chain atoms at 0.5 ps intervals in the listed trajectories, measured against the initial time step for each trajectory, for the four trajectories that show the most mainchain flexibility, as well as for the trajectory (in black) showing the least main-chain flexibility. (b) Side chains of Arg/Gln 65 of the 24 aligned average structures for the trajectories listed in Table 3. The white structures are for the Lys213Ala mutation. The balls indicate the NH1 and NH2 atoms of Arg (and the Nε2 atom of Gln). Hydrogen atoms are not shown. The active site cleft is located just past the bottom edge of this figure.

step in a trajectory and the first time step in that trajectory, for the four trajectories that show the most main-chain atom movement (run 4 of Lys213Ala, run 3 of Arg65Gln, run 3 of Lys213Ala, and run 3 of Lys212Ala, in descending order), as well as (for comparison) the trajectory that shows the *least* movement of the main-chain atoms, run 1 of Lys213Ala. Closer inspection of all trajectories reveals that the Lys213Ala mutation generally leads to considerable flexibility for Arg 65. In order to illustrate this increased flexibility, the average structures for each of the 24 trajectories in Table 3 were aligned against each other using only the alpha carbons; Figure 7b displays the average position of each Arg 65 (or Gln 65, for the four trajectories with that mutation) after this alignment procedure. Three of the Arg 65 side chains among the four Lys213Ala mutant trajectories (shown in white in Figure 7b) are directed *away* from the active site cleft (i.e., away from the bottom in Figure 7b). Taken together, the data indicate that replacement of Lys 213 by Ala has a dramatic impact on the shape and flexibility of the protein backbone, particularly for the key CO_2 binding Arg 65 residue.

Experimental site-directed mutagenesis studies[20−22,24] have revealed that Lys 213 in PEPCK is bound to Mn^{2+}, implying that the side chain of Lys 213 probably does not bear the positive charge characteristic of protonated Lys; the micro-environment of Lys 213 is thought to be carefully balanced to allow this unusual protonation state, thereby enabling Mn^{2+} binding without charge−charge repulsion. Disrupting this delicate environment by replacing Lys 213 with Ala, as in these simulations, has significant consequences on protein structure and apparently can lead to unusual behavior, such as the CO_2 pathway of Figure 6. While experimental mutation of Lys 213 is not a promising direction for further exploration of these unusual and potentially useful behaviors, as it disrupts binding of the Mn^{2+} cofactor necessary for efficient catalysis, it is possible that other mutations may allow access to other similar alternative substrate migration pathways. Alternatively, it is possible that collective low frequency protein motions may also open up unusual substrate entrance/exit pathways like those shown in Figure 6, even in WT proteins. Controlling such motions via temperature, mutation, solvent, or chemical modification may lead to a more efficient enzyme or perhaps one that specifically yields a stereochemically modified product,[16] which would be of great use in the generation of bioderived products.[18]

One final comment about the simulations summarized in Table 3 is that none of the single-point mutations completely destroy the ability of the active site cleft of PEPCK to bind CO_2. That is, CO_2 stays bound in this pocket in at least one of the four simulations for each mutant. This result is perhaps unsurprising; CO_2 is much smaller than a typical, drug-like ligand and, thus does not necessarily require a precise, complementary binding site that can be ruined by a single mutation. While it may be tempting to assign importance to the number of fully bound trajectories for each mutation in Table 3, such conclusions should only be regarded as preliminary. Bearing that caveat in mind, however, it appears that, for example, replacing Tyr 207 with Phe may have a detrimental effect on CO_2 binding, whereas replacing Lys 212 with Ala might improve CO_2 binding efficacy, as it led to three fully bound trajectories, compared to two for the WT. Although the mutations pursued in this work were intended to diagnose the active site cleft's ability to bind CO_2 by *removing*, rather than adding, CO_2 binding character, it has been shown[22] that point mutations can indeed improve the binding of carbon dioxide: in yeast PEPCK, replacing wild-type Glu 272 (analogous to Glu 270 in *E. coli* PEPCK) with Gln improved bicarbonate (equivalent to CO_2) binding, and replacing Glu 284 and Lys 213 (Glu 282 and Lys 213 in *E. coli*) with Gln also each yielded improved CO_2 binding. To reiterate, however, the ratios of outcomes listed in Table 3 should not be taken as statistically significant absent further verification. For example, an experimental study[19] where Arg 65 was replaced by Gln found essentially no change in CO_2 binding behavior, whereas a rudimentary analysis of Table 3 would suggest poorer CO_2 binding in this mutant. All in all, therefore, it appears that the most solid conclusion that can be drawn based on the aggregate

dx.doi.org/10.1021/jp304770h | J. Phys. Chem. B 2012, 116, 11578−11593

data of Table 3 is that all of these mutants can, to varying extents, still bind CO_2. It is therefore of interest to compare the respective CO_2 binding abilities of these 24 trajectories in greater detail through consideration of the strongest CO_2 binding sites.

3.5. Strongest Binding CO_2 Sites among All Simulations. The time step with the strongest PEPCK−CO_2 interaction was identified for each of the 18 CO_2 molecules in the implicit and explicit solvent model simulations, as well as for all trajectories listed in Table 3 (see the Computational Methods section for details on how the interaction energies between CO_2 and the PEPCK system were calculated). The positions of all amino acids and water molecules with any atom within 5 Å of the CO_2 molecule in these tightest binding sites were extracted. The CO_2 molecules were superimposed and fixed in space, and then a slightly modified version of the SEAL algorithm[70] was used to align hydrogen bond donors with other donors and hydrogen bond acceptors with other acceptors; atoms without either functionality were disregarded. In addition, to establish a baseline, a random snapshot was taken from each of the trajectories generated in this work, with the sole selection criterion being that at least one amino acid must be within 5 Å of CO_2 (to prevent the random selection of a CO_2 fully dissociated from PEPCK); these snapshots were similarly aligned. The resulting alignments for these two scenarios are given in Figure 8, where potential hydrogen bond donors are represented by blue points and potential hydrogen bond acceptors are shown by green points. Water molecules are not shown in these images, for the sake of clarity, but they were included in the alignment procedure above and their absence from Figure 8 does not qualitatively change the depictions. Immediately obvious in Figure 8 is that the strongest binding CO_2 sites (Figure 8a) contain more amino acid functional groups than do the randomly selected snapshots (Figure 8b). In addition, there is a large concentration of hydrogen bond acceptors (green points) near the carbon of CO_2 in the nonrandom sites. CO_2, of course, does not possess a hydrogen atom that can be donated to these acceptors, but a general characteristic of hydrogen bond acceptors is a partially negative charge, which is attracted to the partially positive charge located on carbon. The oxygen atoms of CO_2 can, however, act as a hydrogen bond acceptor for a properly positioned hydrogen bond donor, and, as indicated in Figure 8a, there is indeed a strong concentration of blue points along the O═C═O axis. The ratio of hydrogen bond donors to acceptors along the axis and near the carbon were calculated and are given in Table 4 for various subsets of the points in Figure 8.

As seen in Table 4, for these strongest CO_2 binding sites, the ratio of hydrogen bond donors to hydrogen bond acceptors increases as the C═O···X angle (where X is either a donor or acceptor) becomes more and more obtuse, culminating in a total lack of hydrogen bond acceptors in the 170° cone (i.e., that subset closest to end-on binding to CO_2). This pattern mirrors those found in a survey[71] of protein−ligand hydrogen bonds, illustrating the importance of the hydrogen bond in the binding of CO_2 to PEPCK. By contrast, this dependence on the cone angle is not observed for the baseline of randomly selected protein−CO_2 snapshots. It should be noted, however, that even for this random set, hydrogen bond donors outnumber hydrogen bond acceptors at the most obtuse orientation (within the 170° cone) by a count of five to two, providing a further reflection of the strength of these hydrogen bond-based protein−CO_2 arrangements: because they are more stable than other, nonhydrogen bonding arrangements, they last longer in the simulations and were therefore more likely to be selected in the random sample.

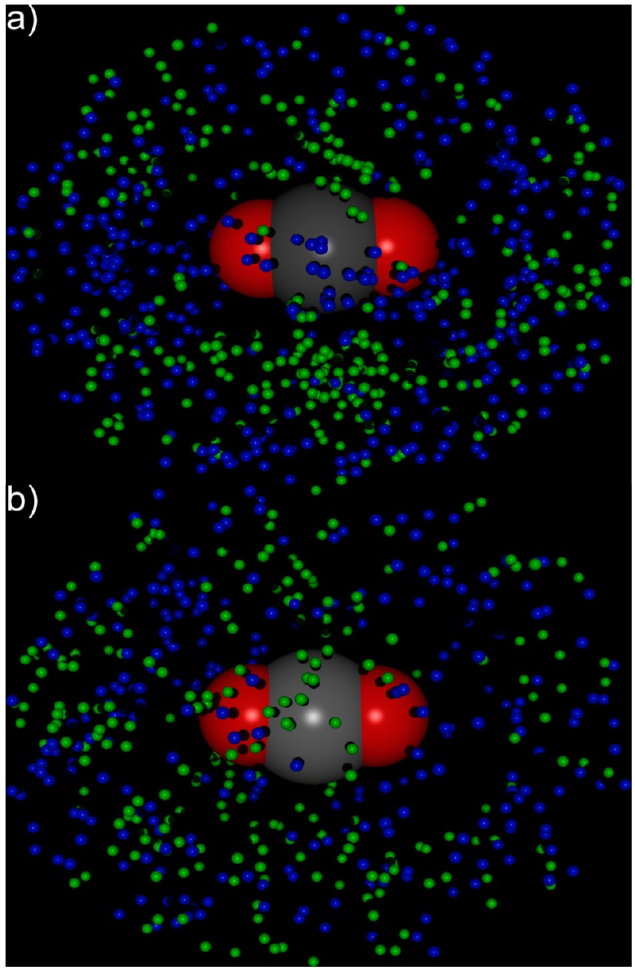


Figure 8. Hydrogen bond donors (blue points) and hydrogen bond acceptors (green points) within 5 Å of CO_2 for (a) the strongest binding CO_2 sites and (b) randomly selected snapshots from each of the MD trajectories generated in this study. See text for details on how the positions of the donors and acceptors were aligned.

Table 4. Hydrogen Bond Donor to Acceptor Ratios for the Strongest CO_2 Binding Sites and the Randomly Selected Sites

strongest sites within cone[a]		random sites within cone[a]
2.34	130°	1.30
2.80	140°	1.34
4.30	150°	1.46
5.78	160°	0.68
15.00[b]	170°	2.50
0.83	near carbon[c]	0.73
1.52	overall	1.17

[a]A hydrogen bond donor/acceptor (X) is within the given cone if the C═O···X angle is greater (more obtuse) than the indicated value. Cones are defined for both of the oxygen atoms in CO_2 and summed. [b]This ratio is actually infinite, as there are 15 hydrogen bond donors, but no hydrogen bond acceptors within this narrowest cone. [c]Points in between the two oxygen atoms of CO_2 and within 5 Å of any CO_2 atom.

dx.doi.org/10.1021/jp304770h | J. Phys. Chem. B 2012, 116, 11578−11593

Also quantified in Table 4 is the enrichment of hydrogen bond *acceptors* in the vicinity of the carbon atom. That acceptors outnumber donors in this region for both sets indicates that the presence of a hydrogen bond donor, with its partially positive hydrogen atom, is rather destabilizing with respect to the protein−CO_2 interaction, a reasonable result given the nearly unit partial charge on carbon in carbon dioxide: calculated to be +0.91*e* with ChelpG[72] at the MP2/6-31G* level and assigned as +0.97*e* by the AMBER94 force field.

3.6. Interaction of CO_2 with Mn^{2+} or Mg^{2+}. The only interaction observed in the MD simulations that is stronger than the protein−CO_2 binding interaction is that between carbon dioxide and Mn^{2+}. In one instance, run 2 of the explicit solvent WT simulations (Table 3), the interaction energy between CO_2 and Mn^{2+} reaches about 15 kcal/mol, comparable in strength to the tightest protein-only CO_2 binding sites. Our previous work,[7] using the highly accurate ab initio correlation consistent Composite Approach (ccCA)[73] method, found an interaction energy between CO_2 and divalent Mg^{2+} of 62.3 kcal/mol; that value is much higher than the currently observed value of 15 kcal/mol not because of the difference between Mg^{2+} and Mn^{2+}, but rather because the ccCA calculations were performed in the gas phase, whereas the charge of Mn^{2+} in these MD simulations has largely been diminished through protein− and solvent−Mn^{2+} interactions. Nevertheless, despite the strength of this interaction, CO_2 is observed to interact with Mn^{2+} in the 24 trajectories listed in Table 3 only twice: in run 2 of the WT simulations and in run 2 of the Lys212Ala simulations, and not at all with the Mg^{2+} cation.

The lack of interaction between CO_2 and Mn^{2+} has been noted experimentally, both in PEPCK[32,36] and in mammalian G-protein-responsive adenylyl cyclase.[37] Why is this interaction missing, given its strength? Inspection of the two aforementioned trajectories reveals that CO_2 initially interacts with Mn^{2+} very early in the simulations, starting within the first 10 ps. This interaction lasts for roughly 6000 ps in the WT simulation but for only 200 ps in the Lys212Ala simulation. Once the interaction is disrupted, through either protein motion or solvent interference, it is not reformed during the remainder of the simulation time. Indeed, for this WT simulation, CO_2 never again comes within 4 Å of Mn^{2+} after its initial close contact of 3−3.5 Å. In other words, once Mn^{2+} assumes its protein- and water-based octahedral coordination environment, it is exceedingly stable and CO_2 cannot interject itself. In the simulations of this work, Mg^{2+} starts even farther away from CO_2's initial location, and thus, CO_2 cannot diffuse toward this cation before the metal has established its coordination environment, leading to a total lack of observed Mg^{2+}−CO_2 interactions.

It should be noted, however, that, in situations where the metal ions are not strongly coordinated by the enzyme, such as in low metal environments, in naturally occurring or engineered mutants, or because the cations have not yet diffused to their usual places in PEPCK as a consequence of the random sequential kinetic mechanism of PEPCK,[74,75] it may be possible for CO_2 to interact with the metal ions and thereby impact the catalytic ability of the enzyme. One extreme example of this possibility can be found in the crystal structure[19] (PDB code 2OLR) of *E. coli* PEPCK, where Mn^{2+} is missing altogether. In this structure, essentially identical in every other respect to the 2OLQ PEPCK structure used in the current work, a second molecule of CO_2 is bound in the site where Mn^{2+} sits in 2OLQ. Cotelesage et al.,[19] who analyzed this crystal structure, stated that they were uncertain about the biological relevance of this binding site, which is located near His 232. To attempt to address this question, the MD simulations of this work were analyzed with respect to CO_2 binding to His 232.

3.7. Binding of CO_2 to a Putative Site Near His 232. To determine if a possible role for His 232 in CO_2 binding exists, Table 5 presents data on the close contacts (<3 Å) between CO_2 and the amino acids of PEPCK for those trajectories in Table 3, where CO_2 stays bound to PEPCK for the entire simulation. (Although one might expect the amino acids that comprise the crystallographically known CO_2 binding site[19] in PEPCK—Arg 65, Tyr 207, and Lys 213—to be the three most prevalent amino acids, it should be point out that most of the MD simulations in this work involved mutations [listed in Table 3] explicitly intended to weaken the CO_2 binding ability of the crystallographic binding site.) In addition, it can be seen in Table 5 that the amino acids most utilized in binding CO_2 are all, with the exception of the last two, within 9 Å of the crystallographic binding site. The resiliency of the entire active site cleft in holding on to CO_2 can be seen, which is essential considering that PEPCK operates by a random sequential kinetic mechanism:[74,75] that is, the substrates used by PEPCK can bind in any order, which requires distinct binding sites for each substrate. More specifically, Table 5 indicates that His 232 does seem to play a significant role in binding CO_2. It is not possible, due to the inability of the classical MD force fields used in this study to address bond breaking events, to determine if binding of CO_2 near His 232 leads to a different stereochemical product[16] compared to binding near Arg 65. At the very least, however, the simulations of this work do seem to suggest that the crystal structure obtained by Cotelesage et al.[19] showing CO_2 bound near His 232 is not anomalous; indeed, it appears that this binding site may be relevant even in the holoenzyme, where Mn^{2+} is bound in its proper location, as was the case in these simulations. Finally, note that Arg 333 is one of only three amino acids (along with Lys 213 and Arg 65) that is used in each of these ten trajectories to bind CO_2, thereby confirming a role in CO_2 binding for this residue, as was postulated previously.[30]

Table 5. Prevalent Amino Acids Used to Bind CO_2 among the 10 Trajectories in Table 3 That Have CO_2 Bound for the Entire Simulation

amino acid[a]	No. of close contacts	% of trajectories with given AA	distance[c] from crystallographic CO_2 (Å)
Tyr 286	52906	70	4.48
Asn 331	26321	80	4.93
Lys 213[b]	24539	100	3.33
Ser 250	24298	60	5.38
His 232	21350	60	5.89
Tyr 207[b]	18688	60	2.92
Arg 333	16257	100	4.87
Leu 249	14238	70	8.07
Phe 413	12230	70	3.84
Gly 410	10342	60	8.97
Arg 65[b]	5366	100	2.93
Ala 411	1581	60	10.17
Thr 404	1493	60	12.05

[a]Only amino acids that appear in at least 60% of the 10 noted trajectories are listed. [b]Also including Gln 65, Phe 207, and Ala 213 for those WT amino acids that have been mutated. [c]Defined as the minimum distance between any atom of the listed amino acid and any atom of CO_2 in the 2OLQ crystal structure of *E. coli* PEPCK.

dx.doi.org/10.1021/jp304770h | J. Phys. Chem. B 2012, 116, 11578−11593

3.8. Role of Lid Closure in PEPCK's Catalysis. One final experimental issue analyzed based on the MD simulations of this work concerns the role of the lid in PEPCK catalysis. It has recently been proposed by Holyoak et al.,[26,76] based on rat PEPCK (which utilizes GTP rather than ATP as an energy source and differs in some respects in structure from *E. coli* PEPCK[12,13]), that when the lid of PEPCK closes, catalysis is accomplished not only through excluding bulk solvent,[30] but also by properly orienting the various substrates. Additionally, via X-ray crystallography, lid closure was found to be triggered by substrate binding in an induced fit-type mechanism.[76] To verify this proposed coupling between substrate binding and lid closure, Holyoak et al. incorporated a point mutation into the lid of rat PEPCK, which led to fewer crystal structures exhibiting a closed lid, poorer overall catalytic efficiency, and even different products in some cases.[26] The authors postulate, based on their crystallographic data, that because their point mutation does not affect the number or nature of contacts between the lid and the rest of the enzyme, the mutated PEPCK must close its lid less often due to an entropic, rather than an enthalpic, effect. While we cannot unequivocally state that these conclusions are wrong and while we do not disagree with the overall induced fit mechanism proposed, the results of the MD simulations in this work suggest that there may be an additional complicating factor: the lid, at least in *E. coli* PEPCK, does sometimes come into close contact with both the rest of the protein and with ATP, and thus, the poorer enzymatic profile of Holyoak's mutant may simply be because these contacts have been disrupted by the mutation.

Close contacts between the closed lid of PEPCK and the rest of the protein have been conclusively demonstrated via X-ray crystallography in the PEPCK of *A. succiniciproducens*,[30] an ATP-utilizing PEPCK very similar to the *E. coli* enzyme. Hydrogen bonds of 3 Å between the backbone of Arg 60 and two lid residues, Thr 388 and Arg 390, were observed, which would be analogous to hydrogen bonds between Arg 65 and Thr 394 and Arg 396 in the *E. coli* PEPCK of this work. Figure 9 shows the crystal structure of *A. succiniciproducens*' PEPCK, overlaid with both the crystal structure of *E. coli*'s PEPCK[19] and the structure from the simulation in this work where the lid is most closed (run 1 of WT). While the lid in *E. coli*'s PEPCK never quite closes in the simulations to the extent seen in *A. succiniciproducens*, it is clearly closed to a considerable degree: the backbones of Arg 65 and Glu 395 are 5.6 Å apart at their closest (cf. >18 Å in the crystal structure of *E. coli*'s PEPCK). The MD simulations reveal that the lid makes even more significant contacts with ATP: in run 3 of the WT, the backbone oxygen of Leu 391 at one point makes a hydrogen bond 2.18 Å long with the amine group of adenine; this group in ATP also makes an even shorter hydrogen bond (1.62 Å) with the side chain of Glu 395 in the same trajectory. Although the lid of the rat PEPCK differs in both sequence and structure from the lid of *E. coli* PEPCK used in the present work and although the guanine in GTP differs slightly from the adenine in ATP, the fact that these contacts are so close certainly suggests that a mutation to the lid of rat PEPCK may indeed destroy similar contacts in that system.

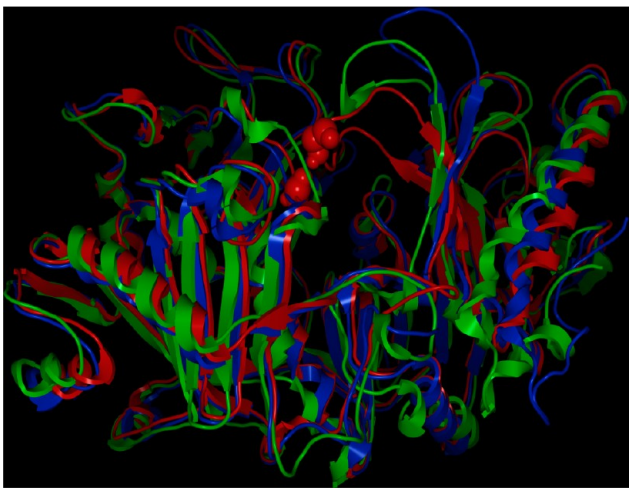


Figure 9. Overlay of the closed lid PEPCK crystal structure from *A. succiniciproducens* (red),[30] the most closed structure in the simulations of this work, from run 1 of WT *E. coli* PEPCK (green), and the crystal structure of *E. coli* (blue).[19] The red space-filling amino acids near the center of the figure indicate the position of the backbone atoms of Arg 60 and Arg 390 in the PEPCK of *A. succiniciproducens*.

It should also be noted that these close contacts between the lid and the rest of the protein (or between the lid and ATP) were observed in simulations *without* any bound phosphoenolpyruvate or oxalacetic acid. Moreover, even if CO_2 is considered a nominal substrate, a putative link between CO_2 binding and lid closure is contraindicated by the MD results: CO_2 had diffused away from PEPCK over 6000 ps before the closed lid structure (the green enzyme in Figure 9) was observed. The current work also does not support the structural explanation offered by Holyoak et al.:[26] that lid closure is linked to the rotomeric state of Tyr 235 (analogous to Tyr 207 in this work). Figure 10 plots the change in the $C-C_\alpha-C_\beta-C_\gamma$ dihedral angle (illustrated in Figure 10a) of Tyr 207 for the following: run 1 of the WT (Figure 10b), where the lid made its closest contact with the rest of the protein; run 3 of the WT (Figure 10c), where the lid made its closest contact with ATP; and run 2 of Lys213Ala (Figure 10d), where the lid did not come within 10 Å of either the rest of the protein or ATP. As can be seen, the greatest change in the rotomeric state of Tyr 207 can be seen in the Lys213Ala trajectory, in which the lid never approaches a closed state. By contrast, when the lid contacted ATP most closely, the Tyr 207 angle (Figure 10c) was mostly constant. Finally, although there is some change in the dihedral angle of Tyr 207 in run 1 of the WT (Figure 10b), which peaks at around 2000 ps, plateaus until about 8000 ps, and then diminishes, it should be noted that the lid in this simulation did not close until 11000 ps, well after this variation in the dihedral angle occurred. Additionally (data not shown), the hydration environment of Tyr 207 was not found to be correlated to the lid's state, as was also suggested to be linked to lid closure.[26]

4. CONCLUSIONS

In this study, the interaction between CO_2 and the enzyme PEPCK was explored using molecular dynamics. Simulations of 18 CO_2 molecules simultaneously interacting with PEPCK yielded similar results with both an implicit solvent model and an explicit solvent model, but slight differences in the results led to the adoption of the more realistic explicit solvent description for the bulk of this work. Using this approach, it was shown that CO_2 preferentially follows three major (and one very minor) channels as it enters or exits the active site cleft of PEPCK. The repeated observation of these binding channels suggests that CO_2 does not merely passively diffuse around a protein but is

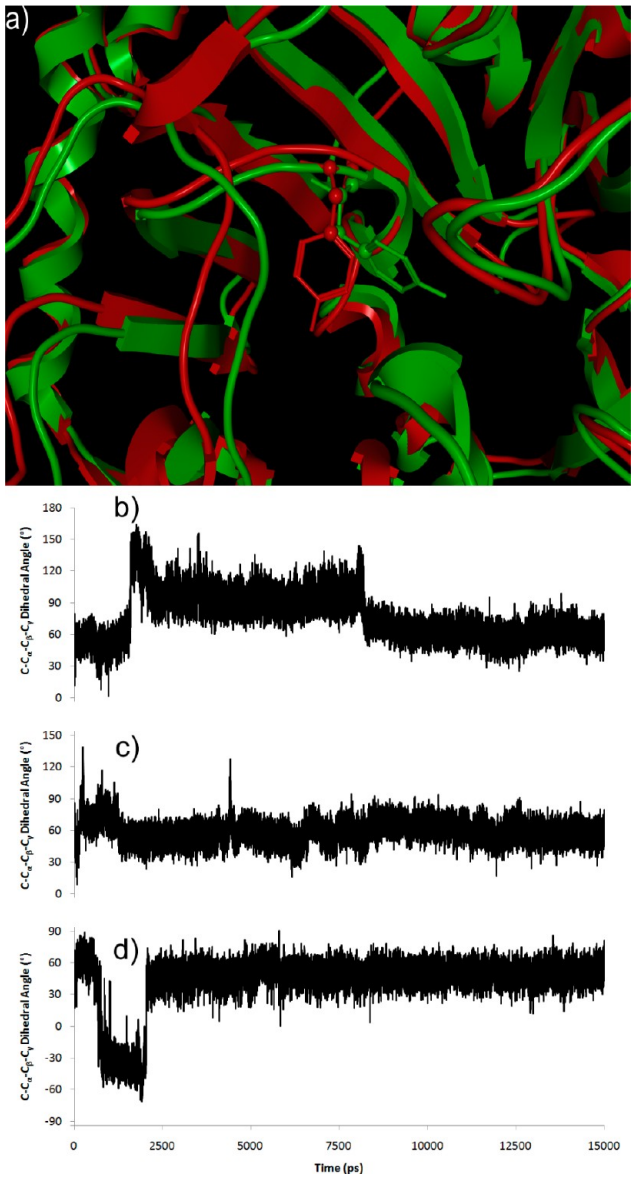


Figure 10. (a) Illustration of two different rotomeric states of Tyr 207 (stick figure); the red structure is the minimum value in (d) (at 1901.25 ps), and the green structure is the maximum value in (d) (at 5788 ps). The atoms used to measure the dihedral angle in Tyr 207 are indicated by balls. (b–d) C–C_α–C_β–C_γ dihedral angle of Tyr 207 for (b) run 1 of the WT trajectory, (c) run 3 of the WT trajectory, and (d) run 2 of the Lys213Ala trajectory.

instead directed by the amino acids of the protein. Additionally, enzymes were simulated where key active site amino acid residues were replaced with chemically inert amino acids, but the ability of PEPCK to bind CO_2 for 15 ns was never removed completely, demonstrating the resiliency of this binding site. Comparison of the strongest CO_2 binding sites revealed the importance of hydrogen bonding in binding carbon dioxide to PEPCK; suitable placement of a partially negative hydrogen bond acceptor, to balance out the nearly unit partially positive charge on carbon in CO_2, was also found to be prevalent in the tightest CO_2 binding sites. The roles of Mn^{2+}, Mg^{2+}, and His 232 in binding carbon dioxide—or the lack thereof, for the first two—was investigated, as was a recently proposed mechanism for lid closure, a process that, based on the MD results, is likely too rapid to be effectively described by X-ray crystallography.

In addition to the PEPCK-specific results, this work also revealed a number of findings that should be applicable to protein–CO_2 interactions in general, and which are therefore of great potential use in climate change amelioration research that requires an especially CO_2-philic protein. For example, the amino acids in Table 1 and especially Table 2 that were identified as most strongly enhancing the affinity of a protein for CO_2 should be applicable to any amino acid-based biomaterial, be it a naturally occurring enzyme, a modified protein, an amino acid salt solution,[6] or even a freshly designed polypeptide. While the pathways of Figures 3–6 are PEPCK-specific, the very fact that favorable channels for CO_2 migration exist can be exploited in other proteins of interest. Moreover, once these channels have been identified via the simulation procedure described in this work (or a similar approach), it may be possible to engineer a more efficient enzyme, or one that yields a slightly different product, by shutting off one of these channels or opening up another. The importance of hydrogen bonding depicted in Figure 8a certainly applies to all protein–CO_2 interactions, and indeed should apply to nonbiological materials (such as hydroxylated silica[77] or MOFs,[78] for example). Finally, as exhibited by the discussion of PEPCK lid closure, computational techniques will continue to prove invaluable in exploring dynamic aspects of the interactions between CO_2 and potential capture materials, which are crucial not only in CO_2 adsorption, but also in CO_2 desorption, be it at high temperatures to regenerate the carbon capture material (as in power plants),[79] or at high pressures in deep earth or deep sea carbon sequestration environments.[80]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Typical examples of CO_2 motion paths in explicit solvent and the average C_α RMSD matrix for the simulations in Table 3 are provided. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: michael.drummond@unt.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the support of the U.S. Department of Energy (BER-08ER64603) and the National Science Foundation (CHE-0741936) for equipment support. T.R.C. also acknowledges the Chemical Computing Group for generously providing the MOE software suite.

■ REFERENCES

(1) Rockström, J.; Steffen, W.; Noone, K.; Persson, Å.; Chapin, F. S., III; Lambin, E. F.; Lenton, T. M.; Scheffer, M.; Folke, C.; Schellnhuber, H. J.; et al. *Nature* **2009**, *461*, 472–475.
(2) Tans, P. NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg/trends/).
(3) Drummond, M. L.; Cundari, T. R.; Wilson, A. K. *Greenhouse Gas Sci. Technol.* **2012**, DOI: 10.1002/ghg.1287).
(4) Schneider, G.; Lindqvist, Y.; Brändèn, C.-I. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 119–143.
(5) Tcherkez, G. G. B.; Farquhar, G. D.; Andrews, T. J. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7246–7251.
(6) McCoy, M. *Chem. Eng. News* **2010**, *88* (48), 14.

(7) Cundari, T. R.; Wilson, A. K.; Drummond, M. L.; Gonzalez, H. E.; Jorgensen, K. R.; Payne, S.; Braunfeld, J.; De Jesus, M.; Johnson, V. M. *J. Chem. Inf. Model.* **2009**, *49*, 2111−2115.
(8) Drummond, M. L.; Wilson, A. K.; Cundari, T. R. *Energy Fuel* **2010**, *24*, 1464−1470.
(9) Drummond, M. L.; Wilson, A. K.; Cundari, T. R. *J. Mol. Model.* **2011**, *18*, 2527−2541.
(10) Drummond, M. L.; Wilson, A. K.; Cundari, T. R. *J. Phys. Chem. Lett.* **2012**, *3*, 830−833.
(11) Delbaere, L. T. J.; Sudom, A. M.; Prasad, L.; Leduc, Y.; Goldie, H. *Biochim. Biophys. Acta* **2004**, *1697*, 271−278.
(12) Case, C. L.; Concar, E. M.; Boswell, K. L.; Mukhopadhyay, B. *J. Biol. Chem.* **2006**, *281*, 39262−39272.
(13) Tari, L. W.; Matte, A.; Goldie, H.; Delbaere, L. T. J. *Nat. Struct. Biol.* **1997**, *4*, 990−994.
(14) Sudom, A. M.; Walters, R.; Pastushok, L.; Goldie, D.; Prasad, L.; Delbaere, L. T. J.; Goldie, H. *J. Bacteriol.* **2003**, *185*, 4233−4242.
(15) Valera, A.; Pujol, A.; Pelegrin, M.; Bosch, F. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9151−9154.
(16) Pérez, E.; Espinoza, R.; Laivenieks, M.; Cardemil, E. *Biochimie* **2008**, *90*, 1685−1692.
(17) Bomgardner, M. *Chem. Eng. News* **2011**, *89*, 7.
(18) Zeikus, J. G.; Jain, M. K.; Elankovan, P. *Appl. Microbiol. Biotechnol.* **1999**, *51*, 545−552.
(19) Cotelesage, J. J. H.; Puttick, J.; Goldie, H.; Rajabi, B.; Novakovski, B.; Delbaere, L. T. J. *Int. J. Biochem. Cell Biol.* **2007**, *39*, 1204−1210.
(20) Yévenes, A.; Espinoza, R.; Rivas-Pardo, J. A.; Villarreal, J. M.; González-Nilo, F. D.; Cardemil, E. *Biochimie* **2006**, *88*, 663−672.
(21) Yévenes, A.; González-Nilo, F. D.; Cardemil, E. *Protein J.* **2007**, *26*, 135−141.
(22) Sepúlveda, C.; Poch, A.; Espinoza, R.; Cardemil, E. *Biochimie* **2010**, *92*, 814−819.
(23) Villarreal, J. M.; Bueno, C.; Arenas, F.; Jabalquinto, A. M.; González-Nilo, F. D.; Encinas, M. V.; Cardemil, E. *Int. J. Biochem. Cell Biol.* **2006**, *38*, 576−588.
(24) Krautwurst, H.; Roschzttardtz, H.; Bazaes, S.; González-Nilo, F. D.; Nowak, T.; Cardemil, E. *Biochemistry* **2002**, *41*, 12763−12770.
(25) Ravanal, M. C.; Flores, M.; Pérez, E.; Aroca, F.; Cardemil, E. *Biochimie* **2004**, *86*, 357−362.
(26) Johnson, T. A.; Holyoak, T. *Biochemistry* **2010**, *49*, 5176−5187.
(27) Stiffin, R. M.; Sullivan, S. M.; Carlson, G. M.; Holyoak, T. *Biochemistry* **2008**, *47*, 2099−2109.
(28) Carlson, G. M.; Holyoak, T. *J. Biol. Chem.* **2009**, *284*, 27037−27041.
(29) Sullivan, S. M.; Holyoak, T. *Biochemistry* **2007**, *46*, 10078−10088.
(30) Cotelesage, J. J. H.; Prasad, L.; Zeikus, J. G.; Laivenieks, M.; Delbaere, L. T. J. *Int. J. Biochem. Cell Biol.* **2005**, *37*, 1829−1837.
(31) Sudom, A. M.; Prasad, L.; Goldie, D.; Delbaere, L. T. J. *J. Mol. Biol.* **2001**, *314*, 83−92.
(32) Encinas, M. V.; González-Nilo, F. D.; Goldie, H.; Cardemil, E. *Eur. J. Biochem.* **2002**, *269*, 4960−4968.
(33) Jabalquinto, A. M.; González-Nilo, F. D.; Laivenieks, M.; Cabezas, M.; Zeikus, J. G.; Cardemil, E. *Biochimie* **2004**, *86*, 47−51.
(34) Jabalquinto, A. M.; Laivenieks, M.; González-Nilo, F. D.; Encinas, M. V.; Zeikus, J. G.; Cardemil, E. *J. Protein Chem.* **2003**, *22*, 515−519.
(35) Jabalquinto, A. M.; Laivenieks, M.; González-Nilo, F. D.; Yévenes, A.; Encinas, M. V.; Zeikus, J. G.; Cardemil, E. *J. Protein Chem.* **2002**, *21*, 393−400.
(36) Hebda, C. A.; Nowak, T. *J. Biol. Chem.* **1982**, *257*, 5515−5522.
(37) Townsend, P. D.; Holliday, P. M.; Fenyk, S.; Hess, K. C.; Gray, M. A.; Hodgson, D. R. W.; Cann, M. J. *J. Biol. Chem.* **2009**, *284*, 784−791.
(38) Mahato, S.; De, D.; Dutta, D.; Kundu, M.; Bhattacharya, S.; Schiavone, M. T.; Bhattacharya, S. K. *Microb. Cell Fact.* **2004**, *3*, 7.
(39) *Molecular Operating Environment (MOE)*; Chemical Computing Group: Montreal, Canada, 2008.
(40) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(41) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *26*, 3219−3228.
(42) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(43) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(44) MacKerell, A. D., Jr.; Banavali, N. *J. Comput. Chem.* **2000**, *21*, 105−120.
(45) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86−104.
(46) Potoff, J. J.; Siepmann, J. I. *AIChE J.* **2001**, *47*, 1676−1682.
(47) Laurie, R. T.; Jackson, R. M. *Bioinformatics* **2005**, *21*, 1908−1916.
(48) Bond, S. D.; Benedict, J. L.; Laird, B. B. *J. Comput. Phys.* **1999**, *151*, 114−134.
(49) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(50) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(51) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472−2493.
(52) DePaul, A. J.; Thompson, E. J.; Patel, S. S.; Haldeman, K.; Sorin, E. J. *Nucleic Acids Res.* **2010**, *38*, 4856−4867.
(53) Lii, J.-H.; Allinger, N. L. *J. Comput. Chem.* **1991**, *12*, 186−199.
(54) Buts, L., personal communication.
(55) Sousa da Silva, A. W.; Vranken, W. F.; Laue, E. D. ACPYPE - AnteChamber PYthon Parser Interface. Manuscript submitted.
(56) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25*, 247−260.
(57) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(58) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J., Interaction models for water in relation to protein hydration. In *Intermolecular Forces*; Pullman, B., Ed.; D. Reidel: Dordrecht, The Netherlands, 1981; pp 331−342.
(59) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(60) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(61) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.
(62) Hub, J. S.; de Groot, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1198−1203.
(63) Nosé, S. *Mol. Phys.* **1984**, *52*, 255−268.
(64) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(65) Hub, J. S.; Winkler, F. K.; Merrick, M.; de Groot, B. L. *J. Am. Chem. Soc.* **2010**, *132*, 13251−13263.
(66) Gallicchio, E.; Paris, K.; Levy, R. M. *J. Chem. Theory Comput.* **2009**, *5*, 2544−2564.
(67) Merz, J., K. M. *J. Am. Chem. Soc.* **1991**, *113*, 406−411.
(68) Liang, J.-Y.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3675−3679.
(69) It should be noted that, compared to our previous work (ref 10), we have moved away from the somewhat vague manner of labeling CO_2 migration pathways based on the direction (i.e., back, top, and right) and towards a more precise label based explicitly on the amino acid residues encountered.
(70) Kearsley, S. K.; Smith, G. M. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615−633.
(71) Sarkhel, S.; Desiraju, G. R. *Proteins: Struct., Funct., Bioinf.* **2004**, *54*, 247−259.
(72) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.
(73) DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. *J. Chem. Phys.* **2006**, *124*, 114104.
(74) Jabalquinto, A. M.; Cardemil, E. *Biochim. Biophys. Acta* **1993**, *1161*, 85−90.

(75) Arnelle, D. R.; O'Leary, M. H. *Biochemistry* **1992**, *31*, 4363−4368.
(76) Sullivan, S. M.; Holyoak, T. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13829−13834.
(77) Qin, Y.; Yang, X.; Zhu, Y.; Ping, J. *J. Phys. Chem. C* **2008**, *112*, 12815−12824.
(78) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. *Science* **2010**, *330*, 650−653.
(79) Rochelle, G. T. *Science* **2009**, *325*, 1652−1654.
(80) Orr, F. M., Jr. *Science* **2009**, *325*, 1656−1658.