JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

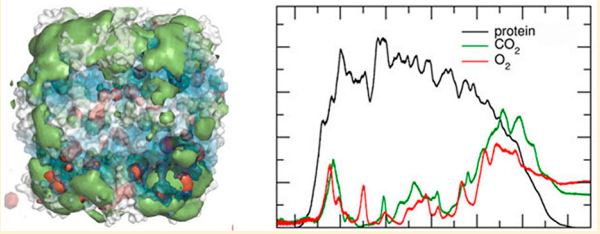
CO_2 and O_2 Distribution in Rubisco Suggests the Small Subunit Functions as a CO_2 Reservoir

Michiel van Lun,[†] Jochen S. Hub,[‡,§] David van der Spoel,[‡] and Inger Andersson*[,‡]

[†]Department of Molecular Biology, Swedish University of Agricultural Sciences, Box 590, S-751 24 Uppsala, Sweden
[‡]Department of Cell and Molecular Biology, Uppsala University, Box 596, S-751 24 Uppsala, Sweden
[§]Institute for Microbiology and Genetics, Georg-August-University Göttingen, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

Supporting Information

ABSTRACT: Protein−gas interactions are important in biology. The enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyzes two competing reactions involving CO_2 and O_2 as substrates. Carboxylation of the common substrate ribulose-1,5-bisphosphate leads to photosynthetic carbon assimilation, while the oxygenation reaction competes with carboxylation and reduces photosynthetic productivity. The migration of the two gases in and around Rubisco was investigated using molecular dynamics simulations. The results indicate that at equal concentrations of the gases, Rubisco binds CO_2 stronger than it does O_2. Amino acids with small hydrophobic side chains are the most proficient in attracting CO_2, indicating a significant contribution of the hydrophobic effect in the interaction. On average, residues in the small subunit bind approximately twice as much CO_2 as do residues in the large subunit. We did not detect any cavities that would provide a route to the active site for the gases. Instead, CO_2 appears to be guided toward the active site through a CO_2 binding region around the active site opening that extends to the closest neighboring small subunits. Taken together, these results suggest the small subunit may function as a "reservoir" for CO_2 storage.



INTRODUCTION

Effective encounter between an enzyme and its substrate(s) is one of the factors that determine the efficiency of enzymatic catalysis. In the case of gaseous substrates there is yet no conclusive picture how small nonpolar ligands move from the surface of an enzyme toward the active site. Diffusion and transport of dioxygen (O_2) in proteins have been investigated to some extent (e.g., refs 1−7), but the migration of carbon dioxide (CO_2) in proteins has attracted less attention. The enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyzes two competing reactions involving CO_2 and O_2 as substrates, respectively. This system therefore offers an excellent opportunity to study not only the diffusion of gases into an enzyme but also potential differences between the two gases CO_2 and O_2 that may influence the partitioning of the two reactions.

Assimilation of atmospheric CO_2 into biomass is fundamental to providing food, feed, and biomaterials. Rubisco catalyzes the rate-limiting step, carboxylation of ribulose-1,5-bisphosphate (RuBP) to produce two molecules of a three-carbon organic acid, 3-phosphoglycerate (3-PGA) from which carbohydrate is derived. Rubisco also catalyzes a competing oxygenation reaction of RuBP, producing equimolar amounts of 3-PGA and 2-phosphoglycolate (2-PG). The atmospheric gases CO_2 and O_2 thus compete directly for the (enolized) common substrate RuBP in the active site of Rubisco. In addition, the further metabolism of 2-PG is an energy-requiring process, which evolves CO_2. There is thus a complex balance between the catalytic rate of the enzyme, the efficiency of the carboxylation reaction, and the concentration of the gaseous substrates in the cell. Because Rubisco fixes some 10^{11} tons of CO_2 worldwide annually, the activity of the enzyme directly influences the atmospheric balance of the two gases with consequences for the concentration of atmospheric CO_2 and, ultimately, the climate. Nearly half of the total carbon dioxide is fixed by photosynthetic marine microorganisms,[8] while tropical forests and savannahs are responsible for 60% of the terrestrial CO_2 uptake.[9]

The efficiency with which CO_2 is able to compete with O_2 in the active site of Rubisco is determined by the CO_2/O_2 specificity factor, V_cK_o/V_oK_c, where V_c and V_o are the maximal velocities of carboxylation and oxygenation, respectively, and K_c and K_o are the Michaelis constants for CO_2 and O_2, respectively.[10] The ratio of carboxylation to oxygenation rates is defined by the product of the specificity factor (often referred to as Ω and sometimes as τ) and the ratio of CO_2 to O_2 concentrations at the active site. The substrate specificity factor of Rubisco enzymes from different origins differs substantially.[11] Nongreen algae have the highest CO_2 specificity factor, anaerobic bacteria have the lowest, while plants, green algae, and cyanobacteria have intermediate specificity factors. However, it appears that selection toward higher specificity factors has occurred at the cost of an overall carboxylation rate,

Received: November 13, 2013
Published: February 4, 2014

dx.doi.org/10.1021/ja411579b | J. Am. Chem. Soc. 2014, 136, 3165−3171

because an inverse correlation has been observed between the specificity factor and turnover rate[11,12] (V_c or k_{cat} for carboxylation) with bacteria displaying low specificities and high turnover numbers, while plants have high specificities coupled to low turnover numbers.

CO_2 also plays a role in regulation of Rubisco activity. To be functional, Rubisco requires prior activation by carbamylation of the ε-amino group of active-site Lys201[13] by a CO_2 molecule, which is distinct from the substrate-CO_2. The carbamylated Lys201 is stabilized by the binding of a magnesium ion to the carbamate group.

During catalysis, Rubisco undergoes a conformational change, which closes the active site and prevents access of solvent during the reaction. The transition between open and closed forms involves rigid-body domain movements, movements of several loops and of the carboxy-terminal strand.[14−16] Structures of the closed form have been obtained with a tight-binding transition-state analogue, 2-carboxyarabinitol-1,5-bisphosphate (CABP) or with inhibitors and substrates binding tightly to the nonactivated form, whereas open forms include the activated unliganded enzyme or complexes with more loosely bound substrates or products (reviewed in ref 17). Presumably, opening is necessary for entry and exit of sugar substrate and product, although the timing of the conformational changes and what triggers these changes are still matters of speculation. Rubisco has been investigated in detail, but very little is known on how the gaseous substrates migrate through the protein and more specifically how they reach the active site of Rubisco.

The transport of oxygen in proteins has been studied to some extent. In some cases, oxygen transport has been detected along well-defined pathways or transiently formed channels and internal cavities.[3,18,19] Alternatively, entry and migration of the gas through liquid-like diffusion have been proposed.[2] The migration of CO_2 in proteins has been studied for membrane channels[20,21] but much less for globular proteins (see, e.g., refs 22 and 23). It is also not known what determines the partition between the gases and what influences the (favorable) competition of CO_2 with O_2.

Here, we investigate the diffusion of CO_2 and O_2 in Rubisco in water using molecular dynamics (MD) simulations. This approach allows us to calculate the binding affinity of CO_2 and O_2 for residues in Rubisco and map pathways that CO_2 may take from the solvent to the active site. Our data allow us to identify the parts of the enzyme with a positive influence on the carboxylation efficiency of the hexadecameric form of Rubisco.

MATERIALS AND METHODS

Starting structures used for the simulations were complexes of the hexadecameric form of Rubisco (Figure 1): activated and unliganded Rubisco from spinach[15] (Protein Data Bank (PDB) access code 1aus), activated Rubisco from spinach in complex with 2-carboxyarabinitol-1,5-bisphosphate (CABP) (1ruc),[24] nonactivated Rubisco from spinach (1rcx),[25] and activated Rubisco from *Chlamydomonas reinhardtii* complexed with CABP (1gk8).[26] Each Rubisco complex was solvated separately in a dodecahedron with an edge length of 12 nm (corresponding to a volume of 2000 nm^3) with TIP4P[27] explicit water, adding Na^+ or Cl^- to neutralize the overall protein charge. A typical simulation box contains 300 000 particles, of which two-thirds is water. Subsequently a range of simulation boxes with varying amounts of gas molecules (either CO_2 or O_2) was created, where the gas molecules were distributed randomly in each box to reach concentrations of 25, 50, 100, 200, or 400 mM. High concentrations were chosen to speed up sampling and hence keep calculation time within reasonable limits (each trajectory calculation took ∼4.2 single CPU years on an AMD Opteron core running at 2.1 GHz). It was verified that no gas clustering occurred in the solvent at these concentrations.

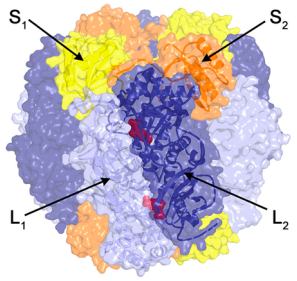


Figure 1. Overall structure of hexadecameric Rubisco. Eight large subunits in light and dark blue and eight small subunits in yellow and orange. The large subunits form tight dimers with two active sites located at the subunit interface. The substrate mimic CABP is displayed as red spheres in the two active sites of the L_1/L_2 dimer facing the front. The small subunit intercalates two dimer pairs at the top and bottom.

From previous calculations it was noted that a large protein complex like Rubisco needs around 4−5 ns to relax in its solvent condition.[28] The distribution of CO_2 or O_2 gas molecules proceeds in a matter of nanoseconds through the entire volume. Each box with its own concentration of gas was therefore run for at least 45 ns to make sure that an equilibrium gas distribution around the protein was obtained. Radial distribution functions were calculated for different simulation lengths, which saturated after around 20 ns. From each trajectory the last 20 ns were taken for radial distribution function calculations. Glutathione peroxidase 5 (PDB code 2p5q) was used as a control. It was simulated under the same conditions as Rubisco, for 50 ns with 30 gas molecules (CO_2 or O_2) corresponding to a concentration of 70 mM.

MD calculations were performed with Gromacs 4.5.[29,30] Interactions of the protein were described by the OPLS all-atom force field,[31] while CO_2 and O_2 parameters were taken from refs 32 and 33, respectively. Particle-mesh Ewald[34] was used for long-range electrostatics with a Fourier grid spacing of 0.12 nm. The temperature was kept at 300 K and the pressure at 1 bar using the weak-coupling algorithm.[35] The systems were prepared and energy minimized and then run with the nonsolvent atoms restrained harmonically with a force constant of 1000 kJ mol^{-1} nm^{-2} for 500 ps before starting the production simulations that were used for analysis. Atomic coordinates were saved every ps for analysis. Cumulative radial distribution data were collected and averaged per residue, subunit, or protein complex. A linear regression analysis was done of the average number of CO_2 and O_2 gas molecules within 0.6 nm of protein residues as a function of gas concentration, and from this analysis the gas distribution at physiological concentrations could be estimated.

From each Rubisco model and at each gas concentration, a snapshot of the cumulative gas density distribution over 20 ns was made to show the areas where residues have the highest cumulative number of contacts (distance <0.6 nm) of CO_2 or O_2 (Figure 2). The gas distribution figures were created using Pymol[36] by taking the gas coordinates from every hundredth time frame (i.e., with 100 ps spacing) and constructing an average gas-density map, which was plotted with a 2σ cutoff for both CO_2 and O_2. The densities for CO_2 and O_2 were computed from separate simulations but are shown in the same figure for comparison (Figure 2 and Supporting Figure 1).

Hydration free energy ΔG_s for the solutes O_2 and CO_2 were computed using thermodynamic integration (TI).[37] ΔG_s was computed in three steps: (i) All interactions of the solute in vacuum turned off; (ii) Lennard-Jones interactions between solute and solvent in a box of water turned on; (iii) in addition, the Coulomb interactions between the solute and water turned on. Steps (ii) and (iii) were simulated in a cubic simulation box of 887 water molecules and step

dx.doi.org/10.1021/ja411579b | J. Am. Chem. Soc. 2014, 136, 3165−3171

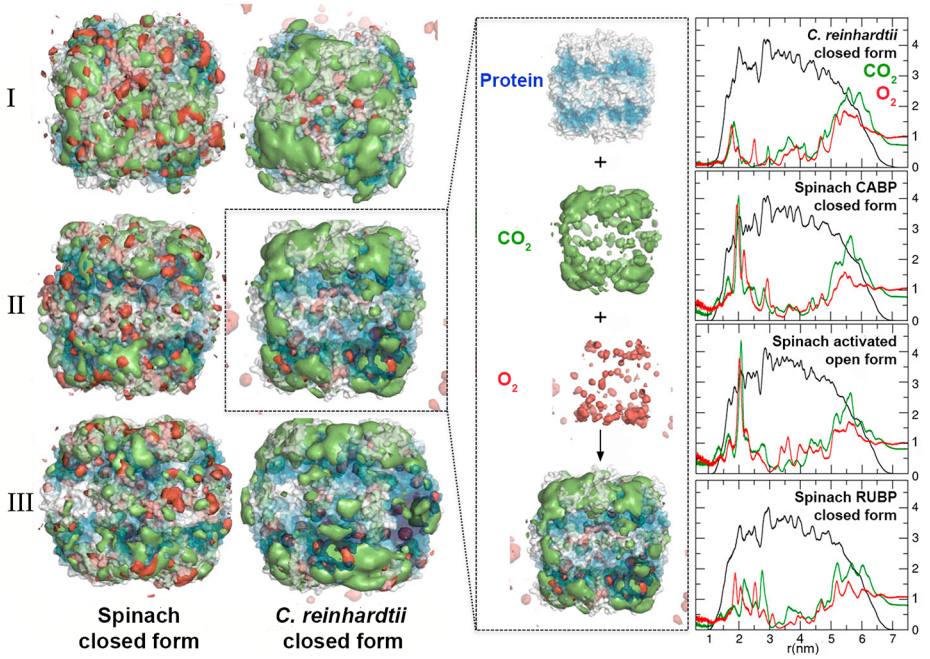


Figure 2. Density map of CO_2 and O_2 in Rubisco and surrounding solvent after 20 ns of simulation. The protein surface is shown as a gradient from white to blue, with darkest blue indicating the center of the active sites. The protein surface is transparent to show the location of CO_2 and O_2 within the molecule (green and red, respectively, 2σ cutoff). The three views for each complex are: (I) down the 4-fold axis, (II) perpendicular to the 4-fold axis on the 2-fold axis, and (III) perpendicular to the 4-fold axis and 45° from the 2-fold axis. The large−large subunit dimer is at the center of view III. Images were constructed by merging fitted density maps for CO_2 and O_2 with the starting protein structure, as shown in the inset. These images are a representative set depicting the highest gas concentration simulations for Rubisco from spinach (PDB code 8ruc) and *C. reinhardtii* (PDB code 1gk8). The complete set of maps from all simulated Rubisco complexes and concentrations are presented in Supporting Figure 1. The last column shows the radial distribution function of protein with respect to its center of mass (black) and the radial distribution function of CO_2 (green) and O_2 (red). From top to bottom is shown: *C. reinhardtii* Rubisco (activated closed CABP complex, PDB code 1gk8), spinach Rubisco (activated closed CABP complex, PDB code 8ruc), spinach Rubisco (activated unliganded open complex, PDB code 1aus), spinach Rubisco (nonactivated closed RuBP complex, PDB code 1rcx). These plots represent radial functions, therefore peaks at the right side of each graph constitute a much larger volume and hence more molecules than peaks of equal size on the left side.

(i) in a simulation box of the same volume (27 nm^3) but containing the solute only. The transitions were carried out along an alchemical reaction coordinate λ, where $\lambda = 0$ and 1 correspond to the initial and final states, respectively. Steps (i−iii) were decomposed into 5, 21, and 11 equally spaced λ-steps, respectively, and each λ step was simulated for 200, 1000, and 1000 ps, respectively. The temperature was controlled via a stochastic dynamics[38] integration scheme ($\tau = 0.1$ ps), and step (i) was simulated at constant volume. A time step of 4 fs was applied. Free-energy differences for each step were subsequently computed by integrating $\langle \partial H/\partial\lambda \rangle$ from $\lambda = 0$ to 1. Here, $\langle \cdot \rangle$ denotes the average computed from the respective trajectory, after removing the first 50 ps for equilibration, and H is the Hamiltonian of the system. ΔG_s is given by the sum of the free energy differences of the three steps. Statistical errors for $\langle \partial H/\partial\lambda \rangle$ were computed using binning analysis, which subsequently yield the error for ΔG_s via Gaussian error propagation.

RESULTS

Solubility of CO_2 and O_2 in Water. Free energy perturbation calculations for the two gases were performed in order to validate the gas models. The models accurately reproduce the experimental data (Table 1), suggesting that the partitioning between solvated and protein-bound gas molecules is correct, thereby allowing us to directly compare the binding of the two gas molecules to Rubisco.

Gas Diffusion into Rubisco. Gas molecules were placed randomly in the solvent at the start of the simulations and gradually diffused into the protein during the simulation. Figure 2 illustrates the distribution of CO_2 and O_2 in Rubisco during the last 20 ns of the simulation by showing the most frequently visited locations as a density map. Gas diffusion into the protein did not lead to an increase in protein size, suggesting that CO_2 and O_2 occupy existing internal cavities and channels. CO_2 migrates through solvent and protein and is found at virtually any location inside Rubisco, although the largest densities of CO_2 in any simulation were found just below or around the protein surface. The central solvent channel located on Rubisco's 4-fold axis provides an additional entry area for gas molecules. The most frequently visited sites for CO_2 are at the interface of the large and small subunits and close to the active site entrance. For each simulation only a handful of residues (out of almost 5000) were not visited by CO_2 within 40 ns.

The distribution of O_2 differs from that of CO_2. Whereas CO_2 resides mostly subsurface, a significantly higher number of O_2 molecules remains in solvent in all simulations. This is illustrated by the difference of CO_2/O_2 count in the region between 5 and 7.5 nm from the center of the simulation box (0 nm) in Figure 2, last column (radial distribution function). CO_2 localizes to defined areas in the protein, whereas O_2 is spread

Table 1. Gibbs Energy of Hydration (kJ/mol) for Gas Molecules in Water from Experimental Data[39] and Simulation

molecule	experiment	simulation
CO_2	0.66	0.7 ± 0.2
O_2	8.8	7.8 ± 0.2

dx.doi.org/10.1021/ja411579b | J. Am. Chem. Soc. 2014, 136, 3165−3171

out less specifically with a larger proportion remaining in the solvent, as shown in the distribution maps.

Affinity of CO_2 and O_2 for Rubisco. At equal concentration of the gases, on average 40% more CO_2 is bound to Rubisco than O_2 (Figure 3). The difference is similar,

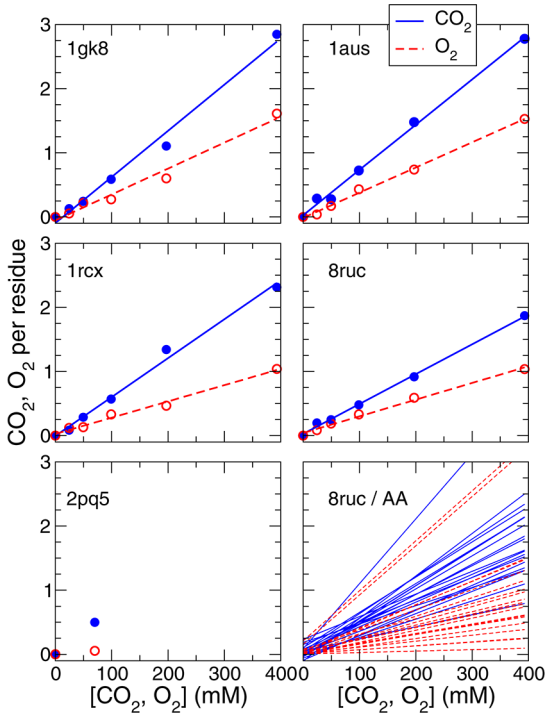


Figure 3. Cumulative distribution of CO_2 (blue solid lines) and O_2 (red dashed lines) per residue in a 0.6 nm radius in Rubisco as a function of the concentration of the gas. The protein complexes are denoted by their PDB code, see Materials and Methods Section for details. Data points for CO_2 and O_2 are represented by closed and open circles, respectively. Data for glutathione peroxidase 5 (PDB code 2p5q) at 70 mM were used as a control. On average, Rubisco has a preference for CO_2 over O_2 at equal concentration. The bottom right plot shows a typical example (CO_2 in 8ruc) of the spread of preference for all 20 types of amino acid.

but not identical, for the four tested complexes. The variation in CO_2 preference for the different Rubisco complexes arises because it is not tractable to get equilibrated CO_2 distributions for each amino acid. The values are accumulated counts of contacts over 20 ns of simulation in each system. Theoretically, a similar structure with an identical sequence (as is the case for the three spinach Rubisco ligand complexes) is expected to have a similar contact count. The starting coordinates for the gas molecules are randomized for each simulation, and therefore the results from two independent simulations with the same gas and number of gas molecules will differ slightly. As shown in the plot per amino acid in Figure 3, there is a large spread in CO_2 binding preference for each type of amino acid, and this will be elaborated on further below.

The affinity of the two gases for Rubisco relative to the solvent was estimated by computing the fraction of bound gas molecules during the last 20 ns of the simulations. We find that 54% of the CO_2 molecules bind to the protein compared to 39% of the O_2 molecules. Using $\Delta G = -k_B T \ln k_{eq}$, where k_{eq} was taken to be equal to the ratio of gas molecules bound/not bound to protein, we can convert the binding preferences to energy differences of 0.4 kJ/mol for CO_2, favoring binding to the protein, and −1.1 kJ/mol for O_2, favoring aqueous solution. The affinity for Rubisco compared to the gas phase is then (using Table 1) +0.3 kJ/mol for CO_2 and +10 kJ/mol for O_2. Hence, the intrinsic affinity of Rubisco for CO_2 is ∼10 kJ/mol stronger than for O_2. It is interesting to consider that the hexadecane/water partition coefficients of CO_2 and O_2 are 1.5 and 6 respectively, corresponding to Gibbs energies of solvation in hexadecane of 0.2 kJ/mol (CO_2) versus 4.3 kJ/mol (O_2). This indicates that O_2 is hardly soluble in any environment, whereas CO_2 is.

Residues That Have a High Affinity for CO_2. Virtually all residues are in contact with a gas molecule within 20 ns, but the interior of the subunit is less accessible than the surface. This needs to be considered when attempting to determine the absolute CO_2 preference for a residue. From each simulation the cumulative radial distribution of gas (CO_2 or O_2) within 0.6 nm from the protein, roughly corresponding to the first minimum in the radial distribution function, was calculated for each residue and then averaged over all large and small subunits in the hexadecamer. The values were divided by the average solvent accessible surface area per residue during simulation (based on the size of accessibility area of each residue at every fourth ns) and were normalized per residue (Supporting Figure 2). The resulting binding strength is dependent on simulation time and gas concentration.

On average, amino acids with a small and hydrophobic side chain (alanine, valine, leucine, isoleucine) or sulfur-containing amino acids are preferred by CO_2 in all tested Rubisco complexes. Alanine and cysteine are the most frequently visited residues. The larger size of the hydrophobic residues phenylalanine, tyrosine, and tryptophan does not add extra affinity for CO_2. Amino acids with a net charge (arginine, lysine, aspartic acid, glutamic acid) are less visited than amino acids with a polar but uncharged side chain (serine, threonine, histidine). O_2 binds with a similar distribution but less frequently to all amino acid types, especially the amino acids with charged side chains. A separate case is cysteine, that attracts O_2 almost as well as CO_2.

Cross Section Images and Routes to the Active Site. Figure 4 shows a thin cross section of the density maps through four active sites of Rubisco. The immediate access point for CO_2 appears as a groove that forms an opening to the solvent and indicates the shortest route for CO_2 from the solvent to the active site. In the simulation, CO_2 accumulates in three out of

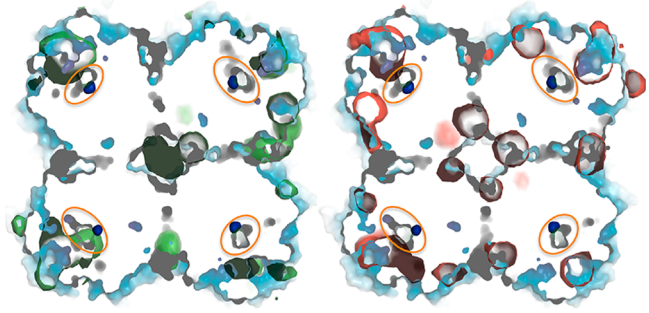


Figure 4. CO_2/O_2 distribution after simulation with highest (C)O_2 concentrations in a cross section of Rubisco (spinach enzyme, PDB code 8ruc). The view is along the 4-fold axis and sections four active sites (circled in orange). The protein surface is white/blue, and the activator CO_2 is shown as a dark blue dot inside the circles. Highest CO_2/O_2 densities are in green and red, respectively.

dx.doi.org/10.1021/ja411579b | J. Am. Chem. Soc. 2014, 136, 3165−3171

four active sites. From this we may conclude that an entry point for CO_2 exists directly from the solvent to the active site in the open conformation. O_2 accumulates at that same point as well. The complete picture is more complex, however, and is illustrated in Figure 5 that shows a close-up of a CO_2/O_2 density map around the active site.

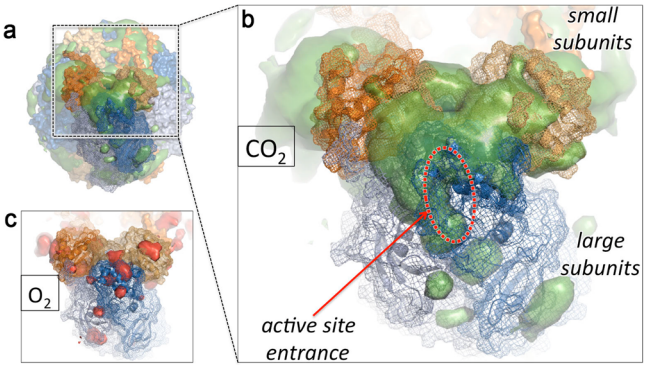


Figure 5. Cumulative CO_2 and O_2 density around the active site of Rubisco. (a) Hexadecameric enzyme (*C. reinhardtii* Rubisco closed complex, PDB code 1gk8) with large subunits in blue and small subunits in orange. The highest density of CO_2 after 20 ns simulation is shown as a green smooth surface. (b) Close-up around the active site with residues from large (blue) and small (orange) subunits represented as a mesh. CO_2 density in green has a direct connection between the large and small subunits that may aid in attracting and guiding CO_2 toward the active site. The active site area (circled in red) is covered by flexible loops. (c) Cumulative O_2 density (red) after 20 ns simulation. The orientation is the same as in (b), and the simulation was run at the same concentration as for CO_2.

The high-CO_2 density area comprises the entrance to the active site, the area close to the nearest large−large subunit interface (dimer interface) and continues toward the nearest small subunits on the large−small subunit interfaces (Rubisco features 48 subunit interfaces and 7 unique interfaces, described in ref 28). The area is continuous and suggests a possible route for CO_2 from the solvent via the large−small subunit interface to the active site. CO_2 does not seem to travel along a narrow path along selected surface residues, but resides in an area that extends from the surface into the protein. This area is predominantly subsurface and CO_2 can either drift along the protein surface or may migrate below the surface. The amino acid side chains are flexible and may create enough space to allow a small molecule like CO_2 to pass. The electrostatic interactions of CO_2 and protein, or rather the lack thereof compared to water, favor CO_2 to be located close to small hydrophobic residues and near the main chain, rather than in the solvent on the surface, reflecting the slight hydrophobic character of CO_2. The CO_2 binding residues (defined as residues that have a <0.5 nm contact distance) differ in some detail in the eight individual active sites. However, the region preferred by CO_2 close to the active site shows the same characteristics at all active sites comprising the active site opening (even in the closed conformation) and regions of the adjacent large−large and large−small subunit interfaces.

Gas Distribution Per Subunit Type. The relative preferences of CO_2 and O_2 for the large and small subunits were examined. As shown in Table 2, residues in the small subunit bind ∼80% more CO_2 per unit of area compared to residues in the large subunit. The relative difference between

Table 2. Relative Distribution of CO_2 and O_2 to Small (S) and Large (L) Subunits Averaged over All Concentrations

S/L	1gk8	8ruc	1aus	1rcx	average
CO_2	2.13	1.81	1.59	1.68	1.8
O_2	1.34	1.60	0.94	1.79	1.4

attracting O_2 to the small subunit in favor of the large subunit is 40% (Table 2, Figure 6); the number of O_2 bound to either subunit type is lower in absolute terms as well.

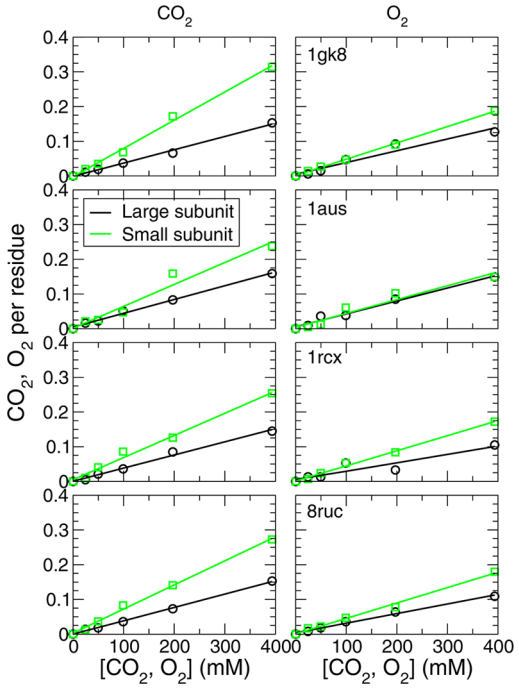


Figure 6. Affinity for CO_2 andO_2 of Rubisco large (black, circles) and small (green, squares) subunits. The ratio of large over small subunit affinities is given in Table 2.

DISCUSSION

The overall efficiency of carbon fixation by Rubisco may be influenced by the ability of the enzyme to attract CO_2, and therefore we have investigated the distribution of gas molecules around Rubisco by MD simulations. The simulations of CO_2 and O_2 binding show a preference for CO_2 over O_2 in all Rubiscos tested at equal concentrations. This is quite remarkable since the solvation free energy of CO_2 in water is in fact 8 kJ/mol lower than that of O_2 (Table 1). We hence estimate that the solvation free energy of CO_2 to the protein is ∼10 kJ/mol lower than that of O_2 (stronger binding). One reason for this may be that CO_2 is somewhat more polar than O_2 (the molecular quadrupoles of the two gas molecules are ∼19.4 and 10.6 Buckingham, respectively) implying that CO_2 may more readily participate in weak polar interactions with both solvent and protein.

Our simulations show a preference of CO_2 for amino acids with small hydrophobic side chains, such as alanine, valine, leucine, and isoleucine, and for the sulfur-containing cysteine residue. Recent surveys of the PDB for CO_2 binding to proteins show basic amino acids like arginine, lysine, and histidine to be the most commonly found in CO_2 binding sites.[22,40] The dominating interaction in the crystal structures exploit the

dx.doi.org/10.1021/ja411579b | J. Am. Chem. Soc. 2014, 136, 3165−3171

difference in electronegativity between carbon and oxygen and its capacity for transient local electrostatic and hydrogen-bonding interactions. An example is the crystal structure of phosphoenolpyruvate carboxykinase that coordinates CO_2 via arginine and tyrosine interactions.[41] The apparent contradiction between these results may be rationalized by considering the dual nature of CO_2 binding, governed by both Coulombic and van der Waals forces. In crystal structures only strong binding CO_2 will be detected, and in these cases binding will typically be dominated by electrostatic interactions, whereas the migration of CO_2 investigated here is dominated by hydrophobic interactions. Similar conclusions were reached by El-Hendawy et al.[23] concerning the forces driving the migration of CO_2.

The transport of CO_2 (and O_2) inside the plant leaf or the cell is governed by complex forces,[42,43] and in vivo gas concentrations are difficult to assess. The current atmospheric CO_2 and O_2 concentrations are 395 ppm (0.04%) and 209 460 ppm (20.95%), respectively. CO_2 availability in aquatic environments is much lower because of a high pH and the slower CO_2 diffusion in seawater. The concentration of CO_2 around Rubisco varies from species to species depending, e.g., on the presence of carbonic anhydrases and carbon concentrating mechanisms. In addition, cross-membrane diffusion, temperature, and carbon transporters all influence the final CO_2 concentration at Rubisco's active site. CO_2 uptake also varies from organism to organism and can depend on light availability. A reasonable estimate of the local CO_2 concentration is in the 2−10 μM range. O_2 presence is furthered by the high O_2 concentration in air and also because plants themselves produce O_2, but the exact concentration around Rubisco is unknown. The intracellular concentration of CO_2 in the plant leaf has been estimated at 60−70% of ambient concentrations.[43,44] It is difficult however to extrapolate the values across species and habitats.

We show here that at equal concentrations, Rubisco provides an environment that favors CO_2 over O_2. While this may partly be a property of proteins in general (as suggested by the control protein), it seems that Rubisco is capable of specific CO_2 capture and is able to localize CO_2 to certain regions of the protein. The areas where CO_2 binds are continuous and connected to the active sites, whereas the areas where O_2 binds tend to be isolated. The CO_2 binding region contains a number of hydrophobic residues close to the surface. Genkov et al.[45] have identified a set of highly conserved small-subunit residues that influence large-subunit catalysis. It is interesting to note that a majority of these residues are located in the CO_2 binding region described above. Changes in specificity on mutation of these residues may be explained by decreases in holoenzyme stability but could also be a result of disturbing the CO_2 binding region.

We did not detect cavities that provide special access to the active site apart from the active site solvent entrance present in the open conformation. This contrasts results from the simulations of diffusion of O_2 in, e.g., myoglobin/hemoglobin[5,18] flavoenzymes[1] or phosphoenolpyruvate carboxykinase.[22] These papers describe migration of O_2 along discrete channels to the active site. In these cases the active site is deeply buried in the protein, and therefore diffusion in a channel or cavity will facilitate O_2 migration significantly. In addition, the lower general specificity for O_2 compared to CO_2 that we find here may explain, to some extent at least, the requirement for O_2 tunnels in these proteins. In Rubisco the active site is located in a shallow groove on the surface and is covered by flexible loops and the C-terminus during catalysis.[15] Our results show that it does not make a difference whether the active site is open or closed for CO_2 to be able to enter. Direct entry from the solvent is possible, but unlikely since the protein is the preferred solvent for the gas and more suitable to guide CO_2 to the active site opening. Instead, CO_2 diffuses from the solvent to a random site on the surface. The time of residence in this initial site depends on the local affinity for the gas. Initial binding may be brief, especially at a low-affinity site, and CO_2 will diffuse further. CO_2 appears to be guided toward the active site through a CO_2 binding region around the active site opening that extends to the closest neighboring small subunits (Figure 5). CO_2 migrating in this area may eventually move into the active site. It is important to note that this movement is not limited to the protein−water interface. While CO_2 is small enough to migrate along and between flexible amino acid side chains, the protein is a relatively viscous solvent (compared to water), and therefore the residues close to the surface are more accessible than those in the core of the protein. Although CO_2 diffusion in protein is slower than in water, it can still be considered fast compared to the catalytic turnover of the enzyme.

The difference in CO_2 binding preference of open and closed states of Rubisco is not measurable. This is not surprising as the amino acid sequences are identical and the open state does not expose a large number of extra residues. The difference between Rubisco from *C. reinhardtii* and spinach in terms of overall gas affinity are also minor, suggesting that our findings may hold for other Rubiscos as well.

The simulations presented here show that the small subunit has a higher affinity for CO_2 than the large subunit in all Rubiscos. The role of the Rubisco small subunit is poorly understood. The arrangement of the small subunits covering a substantial area at two opposite ends of the L-subunit octamer indicates a structural function of the small subunit, namely to stabilize the holoenzyme structure and concentrate the large catalytic subunits.[46] It also appears that the small subunit may be involved in the assembly of Rubisco into the algal pyrenoids.[47] Aside from this structural role, the small subunit is required for maximal catalytic efficiency[48] and can influence specificity.[45,46] The nuclear encoded small subunit has higher sequence diversity than the large subunit. The results presented here indicate that an additional effect of the sequence variability is that the small subunit has evolved to store CO_2.

■ ASSOCIATED CONTENT

Supporting Information

Density maps of CO_2 and O_2 in Rubisco and surrounding solvent and total number of CO_2/O_2 molecules per residue. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

inger@xray.bmc.uu.se

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

MD simulations were performed at the PDC Center for High Performance Computing of KTH (Royal Institute of

Technology) in Stockholm, Sweden. This work was supported by the Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning (FORMAS) and SNIC 023/07-13. J.S.H. was supported by a Marie-Curie Intra-European Fellowship within the 7th European Community framework program.

REFERENCES

(1) Baron, R.; Riley, C.; Chenprakhon, P.; Thotsaporn, K.; Winter, R. T.; Alfieri, A.; Forneris, F.; van Berkel, W. J. H.; Chaiyen, P.; Fraaije, M. W; Mattevi, A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 10603−10608.
(2) Calhoun, D. B.; Vanderkooi, J. M.; Woodrow, G. V., III; Englander, S. W. *Biochemistry* **1983**, *22*, 1526−1532.
(3) Hofacker, I.; Schulten, K. *Proteins* **1998**, *30*, 100−107.
(4) Saam, J.; Ivanov, I.; Walther, M.; Holzhuetter, H.; Kuhn, H. *Proc. Nat. Acad. Sci. U.S.A.* **2007**, *104*, 13319−13323.
(5) Shadrina, M. S.; English, A. M.; Peslherbe, G. H. *J. Am. Chem. Soc.* **2012**, *134*, 11177−11184.
(6) Wang, Y.; Cohen, J.; Boron, W. F.; Schulten, K.; Tajkhorshid, E. *J. Struct. Biol.* **2007**, *157*, 534−544.
(7) Winter, M. B.; Herzik, M. A.; Kuriyan, J.; Marletta, M. A. *Proc. Nat. Acad. Sci. U.S.A.* **2011**, *108*, 881−889.
(8) Field, C. B.; Behrenfeld, M. J.; Randerson, J. T.; Falkowski, P. *Science* **1998**, *281*, 237−240.
(9) Beer, C.; Reichstein, M.; Tomelleri, E.; Ciais, P.; Jung, M.; Carvalhais, N.; Roedenbeck, C.; Arain, M. A.; Baldocchi, D.; Bonan, G. B.; Bondeau, A.; Cescatti, A.; Lasslop, G.; Lindroth, A.; Lomas, M.; Luyssaert, S.; Margolis, H.; Oleson, K. W.; Roupsard, O.; Veenendaal, E.; Viovy, N.; Williams, C.; Woodward, F. I.; Papale, D. *Science* **2010**, *329*, 834−838.
(10) Laing, W. A.; Ogren, W. L.; Hageman, R. H. *Plant Physiol.* **1974**, *54*, 678−685.
(11) Jordan, D. B.; Ogren, W. L. *Nature* **1981**, *291*, 513−515.
(12) Bainbridge, G.; Madgwick, P.; Parmar, S.; Mitchell, R.; Paul, M.; Pitts, J.; Keys, A. J.; Parry, M. A. J. *J. Exp. Bot.* **1995**, *46*, 1269−1276.
(13) Lorimer, G. H.; Miziorko, H. M. *Biochemistry* **1980**, *19*, 5321−5328.
(14) Schreuder, H. A.; Knight, S.; Curmi, P. M. G.; Andersson, I.; Cascio, D.; Brändén, C.-I.; Eisenberg, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9968−9972.
(15) Taylor, T. C.; Andersson, I. *Nat. Struct. Biol.* **1996**, *3*, 95−101.
(16) Duff, A. P.; Andrews, T. J.; Curmi, P. M. G. *J. Mol. Biol.* **2000**, *298*, 903−916.
(17) Andersson, I.; Backlund, A. *Plant Physiol. Biochem.* **2008**, *46*, 275−291.
(18) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.
(19) Riistama, S.; Puustinen, A.; Garcia-Horsman, A.; Iwata, S.; Michel, H.; Wikström, M. *Biochim. Biophys. Acta* **1996**, *1275*, 1−4.
(20) Hub, J. S.; de Groot, B. L. *Biophys. J.* **2006**, *91*, 842−848.
(21) Hub, J. S.; Winkler, F.; Merrick, M.; de Groot, B. L. *J. Am. Chem. Soc.* **2010**, *132*, 13251−13263.
(22) Drummond, M. L.; Wilson, A. K.; Cundari, T. R. *J. Phys. Chem. Lett.* **2012**, *3*, 830−833.
(23) El-Hendawy, M. M.; Garate, J.-A.; English, N. J.; O'Reilly, S.; Mooney, D. A. *J. Chem. Phys.* **2012**, *137*, 145103.
(24) Andersson, I. *J. Mol. Biol.* **1996**, *259*, 160−174.
(25) Taylor, T. C.; Andersson, I. *J. Mol. Biol.* **1997**, *265*, 432−444.
(26) Taylor, T. C.; Backlund, A.; Spreitzer, R. J.; Björhall, K.; Andersson, I. *J. Biol. Chem.* **2001**, *276*, 48159−48164.
(27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(28) van Lun, M.; van der Spoel, D.; Andersson, I. *J. Mol. Biol.* **2011**, *411*, 1083−1098.
(29) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−477.
(30) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845−854.
(31) Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665−6670.
(32) Merz, K. M. *J. Am. Chem. Soc.* **1991**, *113*, 406−411.
(33) Hub, J. S.; de Groot, B. L. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1198−1203.
(34) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8592.
(35) Berendsen, H. J. C.; Postma, J. P. M.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(36) DeLano, W. L. *Plymol*; DeLano Scientific: San Carlos, CA, 2002.
(37) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300−313.
(38) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Simul.* **1988**, *1*, 173−185.
(39) Wilhelm, E.; Battino, R.; Wilcock, R. J. *Chem. Rev.* **1977**, *77*, 219−261.
(40) 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.
(41) 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.
(42) Pons, T. L.; Flexas, J.; von Caemmerer, S.; Evans, J. R.; Genty, B.; Ribas-Carbo, M.; Brugnoli, E. *J Exp. Bot.* **2009**, *60*, 2217−2234.
(43) Sharkey, T. D. *Physiol. Plantarum* **1988**, *73*, 147−152.
(44) Peterson, R. B. *Plant Physiol.* **1989**, *90*, 1322−1328.
(45) Genkov, T.; Spreitzer, R. J. *J. Biol. Chem.* **2009**, *284*, 30105−30112.
(46) Spreitzer, R. J. *Arch. Biochem. Biophys.* **2003**, *414*, 141−149.
(47) Genkov, T.; Meyer, M.; Griffiths, H.; Spreitzer, R. J. *J. Biol. Chem.* **2010**, *285*, 19833−19841.
(48) Andrews, T. J. *J. Biol. Chem.* **1988**, *263*, 12213−12219.