CO_2-Formatics: How Do Proteins Bind Carbon Dioxide?

Thomas R. Cundari,* Angela K. Wilson, Michael L. Drummond, Hector Emanuel Gonzalez, Kameron R. Jorgensen, Stacy Payne, Jordan Braunfeld, Margarita De Jesus, and Vanessa M. Johnson

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

Received July 6, 2009

The rising atmospheric concentration of CO_2 has motivated researchers to seek routes for improved utilization, increased mitigation, and enhanced sequestration of this greenhouse gas. Through a combination of bioinformatics, molecular modeling, and first-principles quantum mechanics the binding of carbon dioxide to proteins is analyzed. It is concluded that acid/base interactions are the principal chemical force by which CO_2 is bound inside proteins. With respect to regular secondary structural elements, β-sheets show a marked preference for CO_2 binding compared to α-helices. The data also support the inference that while either or both oxygens of CO_2 are generally tightly bound in the protein environment, the carbon is much less "sequestered." First principles and more approximate modeling techniques are assessed for quantifying CO_2 binding thermodynamics.

INTRODUCTION

Few recent challenges have engaged the imagination as much as the problem of mitigating greenhouse gases. Atmospheric carbon dioxide concentration has increased greatly since the Industrial Revolution.[1] The current per annum anthropogenic CO_2 production estimate is ~6 Gtons;[2] in contrast, the highest volume carbon-based industrial chemical (ethylene) is produced on a Mton per year scale.[3] Hence, greater utilization must be one part of a multifaceted strategy to reduce atmospheric CO_2. Kheshgi[4] has published a paper on the extent to which atmospheric CO_2 could be reduced by moving the Earth's oceans to more alkaline pH. If nothing else, such drastic scenarios highlight the seriousness of the current environmental situation.

Many strategies are needed to ameliorate atmospheric CO_2: source reduction, conservation of fossil fuels, development of improved alternative energy sources (solar, wind, etc.), greater utilization of CO_2 as an industrial feedstock, and improved sequestration. The latter has attracted much attention as well as pilot studies[5] in part spurred by experience with CO_2 in enhanced oil recovery (EOR).[6] The utilization of "stabilization wedges" can bridge technology gaps until practical carbon neutral energy sources become available.[7]

In many technical challenges, inspiration has been sought from biology. Numerous scientists have, for example, sought to develop enzyme mimics that efficiently fix dinitrogen[8,9] and selectively oxidize hydrocarbons.[10,11] Such efforts have received impetus by advances in molecular biology, where bioinformatics has greatly supplemented experiment and physics-based modeling for hypothesis generation and testing and lead compound identification.

Our group has embarked on a study of CO_2 binding in proteins using first-principles methods and bioinformatics. Our central question was recently put forth by Cotelesage et al.[12]—how do enzymes bind carbon dioxide? We seek an answer not for a specific enzyme, as in that work, but rather an answer across as wide a portion of the proteome as is feasible. First-principles quantum mechanics is employed to elucidate the thermodynamics of CO_2-protein interactions. Quantification of CO_2 binding in different chemical/biological environments is an important step in the development of bioinspired systems to sequester/utilize CO_2 more efficiently and can help calibrate multiscale models to predict the consequences of sequestration strategies.

METHODS

1. Molecular Mechanics (MM). Molecular mechanics calculations were performed using the Molecular Operating Environment (MOE 2008.10) program.[13] We chose the CHARMM27 force field[14] for our analysis of amino acid-CO_2 interactions. Once the CO_2 adduct was created, a stochastic conformational search was run to obtain optimal energies and geometries. Different starting geometries were manually constructed, and additional stochastic searches were conducted until there was reasonable confidence that the global minimum was obtained. Using the most stable conformation thus obtained, the interaction energy (in kcal/mol) between the CO_2 and the adduct partner was calculated with the CHARMM27 force field.

2. Density Functional Tight Binding (DFTB). DFTB calculations[15] were carried out using the DFTB+ program.[16] For the conjugate gradient geometry optimization all atoms were optimized until the maximum force component was less than 1.0×10^{-4} Hartrees $Bohr^{-1}$. The self-consistent charge (SCC) approach was used, and the SCC tolerance was set to 1.0×10^{-5} au. For the SCC iterations a Broyden mixer was used with a mixing parameter of 0.2. Slater-Koster files were obtained from online.[17] The orbital filling criteria was set to Fermi with a temperature of 0 K.

* Corresponding author e-mail: t@unt.edu.

10.1021/ci9002377 CCC: $40.75
Published on Web 08/25/2009

3. Density Functional Theory (DFT). DFT calculations were performed starting from B3LYP/aug-cc-pVTZ geometry optimizations. Frequency calculations were performed at the same level of theory, with a scale factor of 0.9890 for the harmonic vibrational frequencies to obtain the zero-point energy (ZPE). All calculations were performed with Gaussian 03.[18]

4. Correlation Consistent Composite Approach (ccCA). The correlation consistent Composite Approach (ccCA) was applied to the CO_2 adducts. As detailed in ref 19, the ccCA method begins with a B3LYP/aug-cc-pVTZ geometry optimization of the adduct. The frequency calculation was performed at the same level of theory, with a scale factor of 0.9890 for the vibrational frequencies to obtain the ZPE. MP2/aug-cc-pVxZ energies, where x = D(2),T(3),Q(4), were computed and extrapolated to the complete basis set (CBS) limit via an average of the Peterson—a mixed Gaussian/Exponential extrapolation—and Schwartz 3—an inverse cubic formula—extrapolation schemes, as suggested by DeYonker and co-workers.[19] To account for correlation beyond the MP2 level of theory (eq 1), scalar relativistic effects (eq 2), and core−valence correlation (eq 3) a series of single point energy calculations was performed as follows

$$\Delta E(\mathrm{CC}) = E[\mathrm{CCSD(T)/cc-pVTZ}] - E[\mathrm{MP2/cc-pVTZ}] \quad (1)$$

$$\Delta E(\mathrm{SR-MP2}) = E[\mathrm{MP2/cc-pVTZ-DK}] - E[\mathrm{MP2/cc-pVTZ}] \quad (2)$$

$$\Delta E(\mathrm{CV}) = E[\mathrm{MP2(FC1)/aug-cc-pCVTZ}] - E[\mathrm{MP2/aug-cc-pVTZ}] \quad (3)$$

The ZPE and the additive corrections listed above were combined with the MP2/CBS energy resulting in the ccCA energy. These calculations were performed using Gaussian 03.[18]

5. Ligand Environment Analysis for CO_2. The MOE[13] LigX functionality, the LPC (Ligand-Protein Contacts) program,[20] and the Ligand Explorer program[21] were used in conjunction to isolate amino acid residues within a certain distance from carbon dioxide and to identify CO_2 binding motifs for further analysis with the above listed computational chemistry techniques.

6. MSD Motif Database. The MSD Motif database[22] was used to delineate primary and secondary structural elements in CO_2 binding proteins. In order to have the largest database of interactions to analyze, no limitations were made with respect to protein family, coordination geometry, or interaction type.

7. pI and Hydropathy Calculations. Isoelectric points were calculated using the bisection method of Kozlowski.[23] pK_a's were obtained using the EMBOSS algorithm;[24] hydropathy indices[25] were evaluated using Membrane Protein Explorer (MPEx).[26]

8. PDB Codes for the Proteins Included in This Study. Accession to the RCSB database[27] in early 2009 yielded 21 protein structures that have been resolved to date that have CO_2 listed as a ligand: 1E5H, 1KEK, 1KRC, 1R3Q, 2BGI, 2C3Y, 2D2G, 2D2H, 2D2J, 2FVY, 2H31, 2I61, 2OLQ, 2OLR, 2PXZ, 2UZA, 2VBK, 3B52, 3B9Z, 3D92, and 3D93. The structures fall into seven different classes of proteins, four of which serve as catalysts. Some of these enzymes had multiple bound CO_2 molecules. The four primary enzyme classes for the CO_2 binding enzymes represented in the RCBS are oxidoreductases, hydrolases, lyases, and ligases. Five of the structures did not have a bound metal ion; 8 of the 17 proteins with bound metals had only a transition metal.

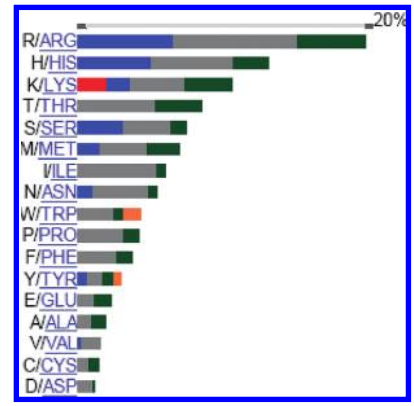

Figure 1. Binding of carbon dioxide to amino acid side chains. 442 samples in the Motif database.[22] Colors indicate nature of bonding: blue = hydrogen bond; gray = undefined; red = covalent; green = van der Waals; orange = atom-plane (aromatic) interactions.

9. Normalized Temperature Factors. In order to enable the comparison of experimentally derived temperature factors across proteins, the normalized temperature factor B_{norm} was used.[28] B_{norm} is equal to $(B - B_{Prot})/\sigma(B_{Prot})$, where B is the experimental temperature factor of the atom in question, B_{Prot} is the average temperature factor for all heavy protein atoms (excluding water, ligands, cofactors, etc.), and σ is the standard deviation.

RESULTS AND DISCUSSION

In early 2009, the RCSB[27] yielded ~20 protein structures in which CO_2 is designated as a ligand. Some contain several bound CO_2 molecules, either in a single chain or in multimeric enzymes. A bioinformatics analysis of CO_2 in proteins (see Methods for proteins investigated and packages employed) reveals a diversity of binding elements: metals, hydrogen bonds, hydrophobic interactions, etc. Chief among these in frequency is hydrogen bonding—or, more accurately, acid/base motifs. *The most basic amino acids like arginine (pI~10.8), lysine (pI~9.5,) and histidine (pI~7.6) are those most commonly found in CO_2 protein binding sites, Figure 1*. However, despite the greater basicity of Lys, His is somewhat more common in CO_2 binding sites. Also, although CO_2 has been described[29] as "hydrophobic", the electronegativity difference between carbon and oxygen yields a $C^{\delta+}=O^{\delta-}$ bond polarity that can be exploited to bind CO_2. Thus, although serine (pI~5.7) and threonine (pI~5.9) are far less basic than Arg, His, and Lys, they show relatively large affinities for CO_2. Negatively charged Asp and Glu show very low binding propensities, suggesting that CO_2 binding does not usually occur via the $C^{\delta+}$. Finally, the similar binding characteristics of Tyr and its less polar counterpart Phe suggest that other chemical factors are important as well; molecular modeling (*vide infra*) can provide some insight into these factors.

The nature of binding sites proximal to and distal from protein-bound CO_2 was assessed. The MOE package[13] was

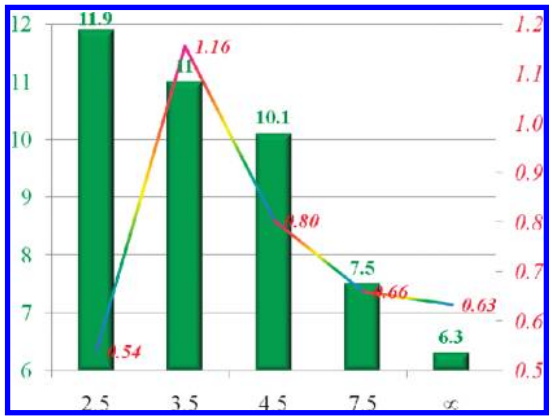

Figure 2. Calculated isoelectric point (pI, bar graphs, left *y*-axis) and hydropathy index (ΔG (kcal/mol), line, right *y*-axis) versus distance (Å, *x*-axis) from CO_2 binding sites. ∞ refers to all amino acid residues within sampled proteins. Distance is taken as the closest distance between any atom of CO_2 and the nearest protein atom.

utilized to identify residues within 2.5 Å, 3.5 Å, 4.5 Å, and 7.5 Å of protein-bound CO_2 and then analyzed to gauge their chemical and secondary structure properties. Figure 2 delineates a striking alteration in the nature of protein CO_2 binding sites that have been crystallographically characterized. Figure 2 shows the calculated[23,24] isoelectric point (pI) for residues at specified distances from bound CO_2. Given a lack of structural detail and ignorance of, for example, the role of bound waters in modifying acid/base properties of amino acids, the aggregate pI metric is best viewed as a qualitative measure of acid/base properties of the binding site and its changing behavior as a function of distance from the bound carbon dioxide.

Figure 2 highlights the importance of acid/base chemistry in the binding of CO_2 in proteins. From a highly basic environment, the binding site gets progressively and significantly less alkaline farther from CO_2. Furthermore, inspection of various CO_2 binding pockets using tools such as LPC[20] and Ligand Explorer[21] implies that while either or both oxygens of CO_2 are generally tightly bound in the protein environment, carbon is less "sequestered". As carbon is often the site at which chemistry is to be effected in CO_2 catalysis (e.g., R-H + CO_2 → RCO_2H or CO_2 → CO_3^{2-}), this difference in accessibility of carbon versus oxygen in protein systems has important implications with respect to biomimetic catalysis for CO_2 utilization.

The line plotted in Figure 2 denotes the average hydropathy index[25] calculated using the MPEx code[26] for amino acids within a specified distance from the CO_2 binding site. Originally conceived as a gauge of the energetic preference for a protein segement to embed itself into a membrane as a transmembrane helix, the hydropathy index can also be used to assess the local hydrophilicity, where more positive values are more hydrophilic. As seen by the line in Figure 2, in contrast to the steady decrease in basicity (as measured by pI) just discussed, there is a turnover in the hydropathy index, changing from hydrophobic residues in the immediate vicinity of CO_2 to hydrophilic sites up to a 3.5 Å distance cutoff, followed by a decrease in hydrophilicity farther from bound CO_2. The clear contrast in the two trends shown in Figure 2 thus imply that acid/base interactions are more important than hydrophobic/philic interactions with respect to protein-based CO_2 bonding.

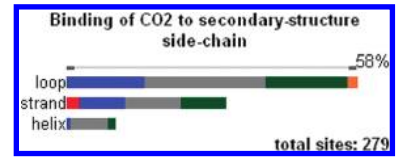

Figure 3. Binding of carbon dioxide to secondary structure side chains. 279 samples in the MSD Motif database.[22] See caption of Figure 1 for color coding.

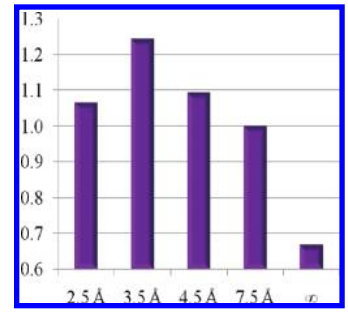

Figure 4. Ratio of *β*-sheets to α-helices (*y*-axis) versus distance (Å, *x*-axis) from CO_2 binding sites. ∞ refers to all amino acid residues within sampled proteins. Distance is taken as the closest distance between any atom of CO_2 and the nearest protein atom.

Despite tremendous advances in synthetic supramolecular chemistry, a major advantage of biological over biomimetic systems remains the superiority of the former in incorporating higher-order architectures—secondary structure elements (SSEs), protein folds, etc.—to achieve desired biological outcomes. Thus, CO_2 binding in proteins was analyzed for simple SSEs, i.e., loop (random coil), *β*-sheet ("strand" in Figure 3), and α-helix. Loops dominate (∼58%), as expected given their flexibility,[22] Figure 3. More interestingly, among true SSEs, there is a marked preference for *β*-sheets over α-helices in protein-CO_2 binding sites.

We have considered four hypotheses to explain the preferential binding of CO_2 to *β*-sheets over α-helices. First, *β*-sheets simply may be more prevalent in the ∼20 CO_2 binding proteins. However, Figure 4 reveals that not only are *β*-sheets less common overall in these proteins than α-helices (*β*:α∼0.67, entry ∞, Figure 4), but it also shows that there is a distinct local preference (up to ∼7.5 Å from CO_2) for *β* over α.

Turning to the second hypothesis, it is known that certain amino acids show a propensity for either α-helix or *β*-sheet formation.[30] Thus, it is possible that the basic amino acids (R, K, and H) identified in Figure 1 as being most common in the CO_2 binding site tend to favor *β*-sheet rather than α-helix formation. However, the literature does not support this hypothesis. For example, the most common amino acid in CO_2 binding pockets, Arg (Figure 1), displays a preference for α over *β* structures. His and Lys are the next most common residues, and these display opposing preferences: for His *β* and α for Lys. The next two most common amino acid residues (Figure 1), Thr and Ser, show *β* over α preferences. In short, there is no obvious correlation between α-vs-*β* amino acid propensities[30] and CO_2-binding affinity.

A third possibility is that the SSE preference in Figure 3 might correspond to a marked difference in flexibility, which could correspond to a difference in CO_2 affinity. However, the experimentally derived B-factors, which provide a measure of thermal motion of the protein atoms, seem to suggest against this as a potential explanation for the SSE

Table 1. Calculated Binding Energies for CO_2 and Amino Acid Side Chain Motifs and Metals at Four Levels of Theory[a]

motif	MM	DFTB	DFT	ccCA
Arg	−10.6	−6.9	−4.9	−6.1
His	−8.4	−5.8	−5.8	−5.7
benzene	−2.8	−0.5	−0.1	−1.3
phenol	−7.5	−1.15	−1.15	−1.9
Cl^-	−3.6	n.p.	−6.1	−7.3
F^-	−9.7	n.p.	−30.1	−29.9
Mg^{2+}	−37.7	n.p.	−63.5	−62.3
Na^+	−10.7	n.p.	−12.9	−11.9
H_2S	−2.8	−0.9	−0.1	−1.1
H_2O	−5.9	−1.6	−1.1	−2.1
H_3O^+	−19.5	−13.3	−15.0	−14.5
NH_3	−4.7	−0.7	−1.1	−2.1
NH_4^+	−11.5	−9.6	−8.2	−8.4
CH_4	−0.7	−0.21	0.0	0.0

[a] n.p. = no DFTB (Slater-Koster) parameters existed for the following elements: F, Mg, Na, Cl; MM = molecular mechanics; DFTB = density function tight binding; DFT = density functional theory; ccCA = correlation consistent Composite Approach. See Methods for full computational details.

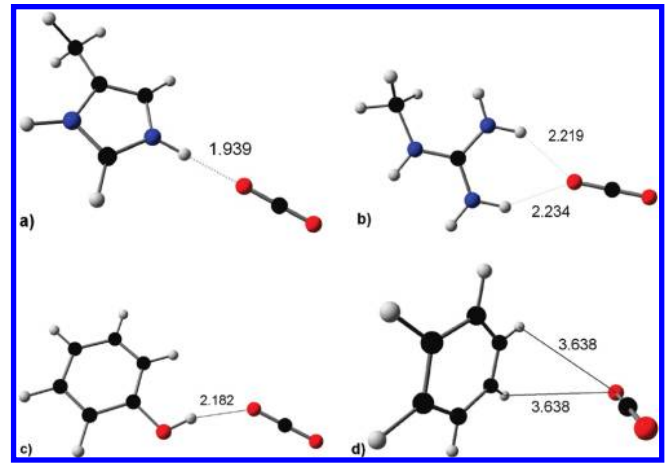

Figure 5. CO_2 adducts of histidine (a), arginine (b), tyrosine (c), and phenylalanine (d) side chain models. His and Arg are shown in their protonated forms. Calculated via DFT methods; other methodologies yield similar calculated geometries. Bond lengths denoted are in Å.

preferences of Figure 3. When normalized to enable comparison between proteins,[28] α-helix residues within 4.5 Å of bound CO_2 have an average, normalized temperature factor of −0.59, compared to −0.63 for residues belonging to β-sheets and −0.61 for those residues without a regular SSE (termed "loop" in Figure 3). Thus, there is little difference in flexibility for any of the SSEs near bound CO_2; all are significantly more stationary than the rest of the protein atoms.

A final hypothesis concerns differing availabilities of hydrogen-bond donors and acceptors. It has been suggested[31] that most of the hydrogen bonding in α-helices goes toward stabilizing the tightly coiled helix, whereas in contrast the edge of a β-sheet is open to hydrogen bonding. Although further work is required to offer a definitive answer, preliminary results from large scale molecular dynamics (MD) simulations on CO_2-binding proteins in our laboratory tentatively support this explanation.

Closer analysis of protein-bound CO_2 interactions can provide important leads with respect to identifying chemical motifs best able to ligate this greenhouse gas. To afford a more quantitative description of the thermodynamics of these interactions, first principles *ab initio* techniques and approximate modeling approaches were employed. The former not only are the most computationally expensive but also the most accurate; the latter rely on parametrization versus experimental data, which can be sparse or of limited accuracy for many biological systems.

We have applied a parameter-free *ab initio* technique—the correlation consistent Composite Approach[19] (ccCA)—to quantify the thermodynamics of interaction between CO_2 and amino acid side chain motifs deemed significant in protein binding of CO_2 (Figure 1). These binding energies are organized in Table 1. The amino acid side chains of arginine and histidine (first two rows of Table 1) bind more strongly to CO_2 (via the oxygen atom of the latter, Figure 5) than do other side chain mimics, e.g., methane (Ala), phenol (Tyr), or benzene (Phe).

Chemical motifs like ammonium (mimicking the side chain of Lys) also bind strongly to CO_2, even more strongly than Arg and His. On the other hand, neutral species like ammonia and water are only weakly bound to CO_2, Table 1. Metal ions like Mg^{2+} and Na^+ have very strong CO_2 binding affinities. Interestingly, aromatic side-chain mimics (benzene for Phe and phenol for Tyr) have similar binding enthalpies despite different coordination modes: hydrogen bonding through the hydroxyl moiety for the latter, the former binding through its aromatic π-cloud (see Figure 5c,d). Note that MM and DFT indicate a discrepancy in binding energies for Phe and Tyr models (Table 1, entries for benzene and phenol, respectively), whereas the more approximate density functional tight binding (DFTB) result and the more sophisticated ccCA result predict commensurate CO_2 binding propensities, Figure 1.

From the perspective of modeling larger and more realistic chemical and biological systems with respect to CO_2 sequestration, it is interesting to observe the agreement among the various techniques with respect to CO_2 binding (Table 1). There are differences in magnitudes of CO_2 binding energies between the quantum-based techniques and the faster and more approximate (but also more applicable to large biological systems) method—molecular mechanics (using the CHARMM27[14] force field). However, the ordering is uniform across classical (MM) or quantum (DFTB,[15,16] DFT,[18] or ccCA[19]) techniques. Qualitatively, the results for DFTB and DFT versus the most accurate of the techniques (ccCA) are also very good, Table 1. Hence, in the design/analysis of CO_2-enzyme interactions and CO_2-biomimic sequestration, the nature of the bonding lends itself to description by a host of modeling techniques. Moreover, one may expect fruitful results from hybrid techniques that describe crucial interactions with CO_2 using a technique like ccCA and a more approximate technique like DFTB or MM to describe chemical interactions distal to bound CO_2. Work along these lines is underway in our laboratory.

CONCLUSION

The binding of CO_2 to proteins was analyzed through a combination of bioinformatics, molecular modeling, and first-principles quantum mechanics. It is concluded that acid/base interactions are the dominant chemical force by which

proteins bind CO_2. Acid/base interactions are deemed more significant than hydrophobicity/philicity. Also, among the two types of regular secondary structural elements, β-sheets are more amenable to CO_2 binding than α-helices. The data support the inference that while either or both oxygens of CO_2 are generally tightly bound in the protein, the carbon is much less "sequestered". As carbon is often the most desirable site for reactivity in CO_2 catalysis, this difference in accessibility has important implications with respect to biomimetic catalysis. First principles *ab initio* techniques and more approximate modeling techniques are assessed for quantifying CO_2 binding thermodynamics; the latter are deemed appropriate for describing protein-CO_2 interactions, which may thus open up new vistas for the design and analysis of protein-CO_2 interactions via computational chemistry/biology, as well as bioinspired materials for enhanced CO_2 sequestration.

ACKNOWLEDGMENT

The authors acknowledge the support of the United States Department of Energy (BER-08ER64603) and the National Science Foundation REU program (CHE-0648843) for summer scholar support for M.D.J. and V.M.J. and NSF-CRIF (CHE-0741936) for equipment support. T.R.C. also acknowledges the Chemical Computing Group for generously providing the MOE software suite.

REFERENCES AND NOTES

(1) Feely, R. A.; Sabine, C. L.; Hernandez-Ayon, J. M.; Ianson, D.; Hales, B. Evidence for upwelling of corrosive "acidified" water onto the Continental Shelf. *Science* **2008**, *320*, 1490–1492.
(2) Boden, T. A.; Marland, G.; Andres, R. J. In *CO_2 Emission Calculations and Trends (EPA/600/R-96/072)*; U.S. Environmental Protection Agency: Washington, DC, 1996.
(3) McCoy, M.; Reisch, M. S.; Tullo, A. H.; Short, P. L.; Tremblay, J.-F.; Storck, W. J.; Voith, M. Facts & Figures of the Chemical Industry. *Chem. Eng. News* **2008**, *86*, 35.
(4) Kheshgi, H. S. Sequestering atmospheric carbon dioxide by increasing ocean alkalinity. *Energy* **1995**, *20*, 915–922.
(5) Oelkers, E. H.; Schott, J. Geochemical aspects of CO_2 sequestration. *Chem. Geol.* **2005**, *217*, 183–186.
(6) Orr, F. M., Jr.; Taber, J. J. Use of Carbon Dioxide in Enhanced Oil Recovery. *Science* **1984**, *224*, 563–579.
(7) Pacala, S.; Socolow, R. Stabilization wedges: Solving the climate problem for the next 50 years with current technologies. *Science* **2004**, *305*, 968–972.
(8) Howard, J. B.; Rees, D. C. How many metals does it take to fix N_2? A mechanistic overview of biological nitrogen fixation. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17088–17093.
(9) Münck, E.; Bominaar, E. L. Bringing stability to highly reduced Iron-Sulfur clusters. *Science* **2008**, *321*, 1452–1453.
(10) Mas-Ballesté, R.; Que, L., Jr. Targeting specific C-H bonds for oxidation. *Science* **2006**, *312*, 1885–1886.
(11) Das, S.; Incarvito, C. D.; Crabtree, R. H.; Brudvig, G. W. Molecular recognition in the selective oxygenation of saturated C-H bonds by a Dimanganese catalyst. *Science* **2006**, *312*, 1941–1943.
(12) Cotelesage, J. J. H.; Puttick, J.; Goldie, H.; Rajabi, B.; Novakovski, B.; Delbaere, L. T. J. How does an enzyme recognize CO_2? *Int. J. Biochem. Cell Biol.* **2007**, *39*, 1204–1210.
(13) Chemical Computing Group, Montreal, Canada.
(14) 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.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3615.
(15) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260–7268.
(16) Density Functional based Tight Binding and more (DFTB+). http://www.dftb-plus.info (accessed June 18, 2009).
(17) Density Functional based Tight Binding. http://www.dftb.org/parameters/download/ (accessed June 18, 2009).
(18) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02*; Gaussian, Inc.: Pittsburgh, PA, 2003.
(19) DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The correlation consistent composite approach (ccCA): An alternative to the Gaussian-n methods. *J. Chem. Phys.* **2006**, *124*, 114104.
(20) Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M. Automated analysis of interatomic contacts in proteins. *Bioinformatics* **1999**, *15*, 327–332.
(21) Zhang, C. http://www.kukool.com/ligand (accessed June 18, 2009).
(22) Golovin, A.; Henrick, K. MSDmotif: Exploring protein sites and motifs. *BMC Bioinf.* **2008**, *9*, 312.
(23) Kozlowski, L. http://isoelectric.ovh.org (accessed June 18, 2009).
(24) Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **2000**, *16*, 276–277.
(25) White, S. H.; Wimley, W. C. Membrane protein folding and stability: Physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 319–365.
(26) Jaysinghe, S.; Hristova, K.; Wimley, W.; Snider, C.; White, S. H. http://blanco.biomol.uci.edu/mpex (accessed June 18, 2009).
(27) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(28) Lu, Y.; Wang, R.; Yang, C.-Y.; Wang, S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.* **2007**, *47*, 668–675.
(29) Leung, K.; Nielsen, I. M. B.; Kurtz, I. Ab initio molecular dynamics study of carbon dioxide and bicarbonate hydration and the nucleophilic attack of hydroxide on CO_2. *J. Phys. Chem. B* **2007**, *111*, 4453–4459.
(30) Costantini, S.; Colonna, G.; Facchiano, A. M. Amino acid propensities for secondary structures are influenced by the protein structural class. *Biochem. Biophys. Res. Commun.* **2006**, *342*, 441–451.
(31) Bartlett, G. J.; Porter, C. T.; Borkakoti, N.; Thornton, J. M. Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.* **2002**, *324*, 105.

CI9002377