THE JOURNAL OF PHYSICAL CHEMISTRY Letters

Letter

pubs.acs.org/JPCL

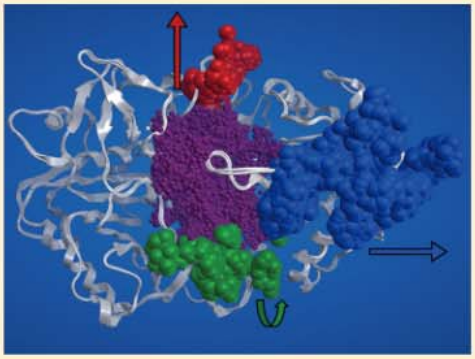
Carbon Dioxide Migration Pathways in Proteins

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

Center for Advanced Scientific Computing and Modeling (CASCaM), Department of Chemistry, University of North Texas, Denton, Texas 76203-5070, United States

Supporting Information

ABSTRACT: Some of the most important biological processes, such as carbon fixation, are dependent on protein−gas interactions. The motion of CO_2 through the enzyme phosphoenolpyruvate carboxykinase was investigated using extensive all-atom molecular dynamics simulations. Three discrete migration pathways were located, suggesting the protein directs the movement of CO_2. The chemical nature of these pathways is discussed, as are their biotechnological ramifications.



SECTION: Macromolecules, Soft Matter

Many important reactions in nature involve the interactions of small gas molecules with proteins. For example, N_2 is fixed by bacterial nitrogenases,[1] O_2 and CO_2 are transported in the blood by hemoglobin,[2] NO controls vasodilation,[3] and CO_2 is fixed in the Calvin cycle by the enzyme RuBisCO.[4] This last, the enzymatic utilization of carbon dioxide, not only represents the ultimate energy source for all higher life, but has also recently been scrutinized in the context of developing biotechnological approaches[5−8] to ameliorate rising levels of atmospheric CO_2.[9] Recent research detailing protein−CO_2 interactions has revealed patterns in primary,[10] secondary,[11] and tertiary[12] protein structure that influence carbon dioxide binding affinity. However, one crucial element not contained in previous work is a description of the dynamics of protein-CO_2 binding, that is, how substrate CO_2 navigates across and through a protein to reach favorable binding sites.

Extensive research, both experimental[13−21] and computational,[22−36] has been devoted to the motion of small gas molecules through proteins, particularly the diffusion of O_2, CO, and NO through hemo-, myo-, and neuroglobin. However, despite this ongoing research, a general consensus about the relevance of proposed ligand migration pathways has not yet been reached.[24] Nevertheless, important insight has been gained from these studies on ligand migration pathways, including support for Ni as the initial H_2 binding site in an Ni−Fe cluster in hydrogenase,[37] the role of these pathways in determining evolutionary similarity[38] (or lack thereof)[39] between proteins, and assignment of unknown enzymatic function.[31] Overall, a complete understanding of many important biological processes requires an understanding of the motion of ligands through proteins.

For CO_2, in the early 1990s, two groups[40,41] studied the movement of carbon dioxide into the active site cleft of carbonic anhydrase using molecular dynamics (MD) simulations. However, these simulations were not lengthy (≤126 ps) and utilized either approximate models of the protein[40] or restraints on the position of CO_2.[41] For the development of a more accurate and comprehensive description of protein−CO_2 interactions, extensive all-atom, explicit solvent MD simulations were performed. (See the Supporting Information for technical details.) The system investigated in this work is the enzyme phosphoenolpyruvate carboxykinase (PEPCK), a 540-reside enzyme found in the gluconeogenesis pathway that, in one direction, can utilize CO_2 as a substrate (Figure 1). It was

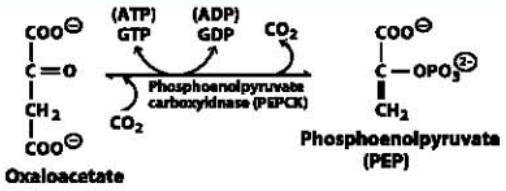


Figure 1. Reaction catalyzed by PEPCK.

shown in 1980[42] that CO_2 binds with a semirandom kinetic mechanism to the PEPCK in *E. coli*; fully random kinetics were found in 1992 for plant (*C. gayana*) PEPCK,[43] implying CO_2-specific binding sites in both species. Subsequent X-ray crystallography (PDB ID: 2OLQ)[44] confirmed this finding

Received: January 27, 2012
Accepted: March 4, 2012
Published: March 4, 2012

for *E. coli*, revealing a CO_2 binding site located deep in a lidded active site cleft; this work also confirmed CO_2 and not HCO_3^- as the most active substrate. However, details of the transit of CO_2 into and out of this cleft are not yet known.

Four independent trajectories, each 15 ns, were generated starting from the (energy minimized) crystallographically determined protein−CO_2 complex geometry[44] for wild-type (WT) PEPCK. Additionally, four 15 ns trajectories were generated for two single-point PEPCK mutants, where the residues that hydrogen bond CO_2 in the crystal structure, Arg 65 and Tyr 207, were changed in silico to Gln and Phe, respectively. These 12 trajectories, each 15 ns in length, yield 180 ns of independent simulations. While shorter than simulations that have been performed for smaller proteins (e.g., the 7 s performed for 151-residue myoglobin[35]), the MD sampling in this work is, adjusting for the larger size of PEPCK, equivalent to very recent MD work investigating ligand migration.[23]

In 4 of these 12 trajectories − two for WT and one each for the Arg65Gln and Tyr207Phe mutants − the CO_2 remained in PEPCK's active site cleft for the entire 15 ns simulation. It is suggestive, although not conclusive absent simulation of many additional trajectories,[30] to note that modifying the WT residues responsible[44] for binding carbon dioxide resulted in a greater propensity for CO_2 to exit the protein. It should also be noted that experimentally the Arg65Gln mutation did not affect the binding affinity of PEPCK for CO_2,[44] further suggesting that understanding the relative effects of point mutation requires additional trajectories. In any event, because the focus of this work is ligand transport to and from the active site cleft, further details of trajectories with CO_2 bound throughout will be offered in a future publication, in the interest of space.

In the eight remaining trajectories, CO_2 left the crystallographically determined active site by three paths (Figure 2). Making reference to the orientation shown in Figure 2 and the position of the lid capping the active site (indicated by an asterisk), these three paths are labeled as proceeding out (a) the back, (b) the top, or (c) the right side of the active site cleft. There is no obvious connection between the Arg65Gln or Tyr207Phe mutations and CO_2 exit pathway: CO_2 in the WT simulations exits via the back in one and via the top in another; for Arg65Gln, CO_2 exits via each of the three paths in three independent trajectories; and for Tyr207Phe, CO_2 exits via the top twice and via the right once. As before, it is tempting to assign precedence to the top exit (Figure 2b) given its abundance relative to the other paths, but this conclusion is premature absent further simulations. Nevertheless, the fact that these eight exit trajectories can be clustered into three distinct channels rather than being randomly distributed throughout the open space of the cleft suggests that interaction with the protein strongly influences the egress of CO_2 from the active site.

To further characterize the directing influence of the protein, Table 1 lists the amino acids found along each pathway. It is interesting to compare the physicochemical properties of the CO_2-directing residues lining these paths to the CO_2-binding amino acids previously identified in a study of all known protein-CO_2 X-ray crystal structures found in the PDB.[10] The latter group is dominated by the basic amino acids Arg, His, and Lys. Indeed, for the specific example of the PEPCK-CO_2 complex, there are three lysines, two arginines, and one histidine within 5 Å of CO_2 in the experimentally determined

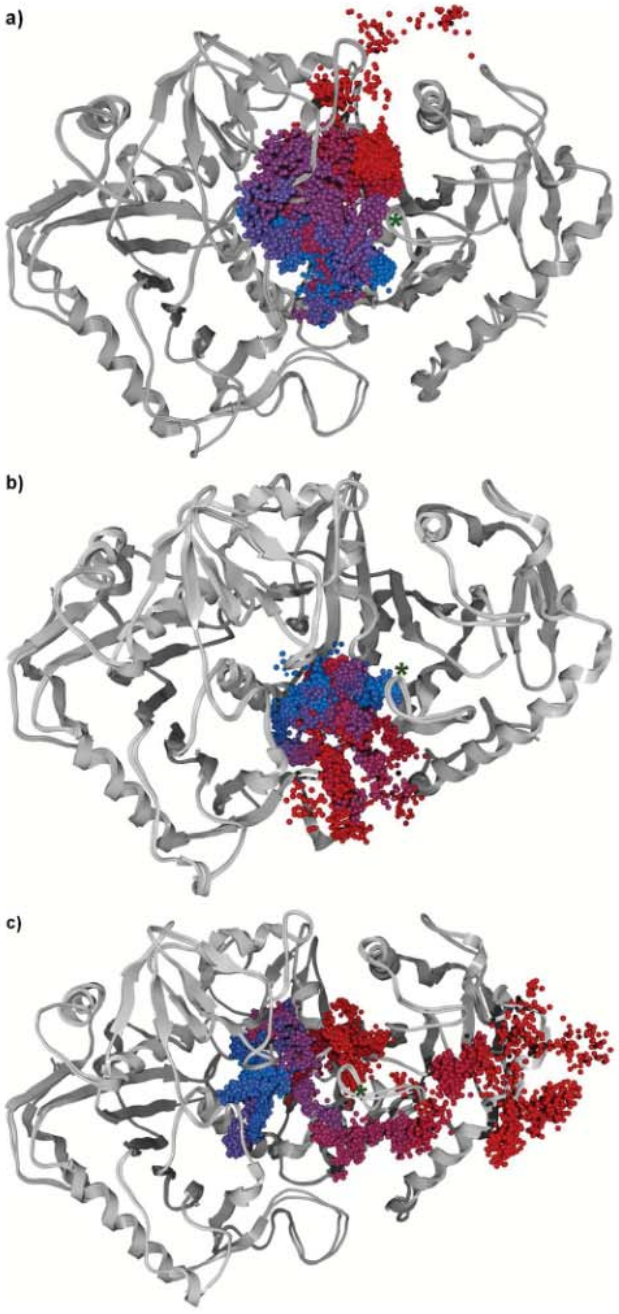


Figure 2. Trace of the carbon of CO_2 (spheres) as it moves from the crystallographically known binding site of PEPCK (ribbon) out (a) the back, (b) the top, or (c) the right side of the active site cleft; the number of superimposed trajectories, respectively, is 2, 4, and 2. Progress over time is indicated by transition from blue through purple to red. The mobile active site lid is marked with an asterisk.

CO_2 binding site.[44] By contrast, the only basic residue found along any of the CO_2 migration paths (Table 1) is Arg 449. (The series of red dots at the top of Figure 2a represent interactions between CO_2 and Arg 449.) This residue is quite distant (22.0 Å) from the crystallographic binding site and indeed is located at the solvent-accessible end of the path; it could thus, perhaps, instead be categorized as a surface residue forming a relatively long-lived interaction with CO_2 (Supporting Information) rather than a true pathway-lining amino acid. Nomenclature aside, it seems that Arg 449 is serving as a CO_2

dx.doi.org/10.1021/jz3001085 | J. Phys. Chem. Lett. 2012, 3, 830 833

Table 1. Key Amino Acids along the CO_2 Migration Pathways and their Distance from the Experimental CO_2 Binding Site[a]

path (a)		path (b)		path (c)	
Thr 404	15.3 Å	Thr 404	15.3 Å	Thr 404	15.3 Å
Asn 92	15.5 Å	Thr 388	15.6 Å	Phe 405	16.3 Å
Thr 388	15.6 Å	Ala 407	15.8 Å	Pro 403	18.3 Å
Asn 94	15.9 Å	Phe 405	16.3 Å	Phe 365	23.5 Å
Trp 83	19.9 Å			Leu 516	24.9 Å
Arg 449	22.0 Å			Thr 512	28.8 Å

[a]See the Supporting Information for definition.

reservoir, analogous to Merz's previous characterization of a histidine-rich CO_2 binding site near the mouth of the active site cleft of carbonic anhydrase.[40] Essentially, these solvent-accessible reservoir sites serve to increase the local concentration of CO_2 around the enzyme, trapping CO_2 diffusing in the solvent environment and thereby facilitating efficient turnover for enzymes that utilize CO_2 as a substrate. Basic surface amino acids have also recently been invoked to explain the permeation of CO_2 through a membrane channel.[45]

In contrast with the basic residues located at the ends of the CO_2 migration paths, the amino acids along the ligand channels (Table 1) are either hydrophobic (Phe, Ala, Leu, Pro, and Trp) or polar but not basic (Thr and Asn). Interestingly, Thr has previously been noted[46] to play a subtle effect in determining the orientation of CO_2 bound by the hydrophobic pocket in the active site of carbonic anhydrase. Furthermore, based on previous ab initio calculations[10] using the highly accurate correlation-consistent composite approach (ccCA),[47] the Thr···O C O hydrogen bond should only be about one-third the strength of the Arg···O C O hydrogen bond. Therefore, the fact that the amino acids in these channels are either hydrophobic (as is CO_2)[48] or at best capable of forming only weak interactions with CO_2 suggests that they may be optimized to interact with carbon dioxide strongly enough to facilitate the entrance of CO_2 into these channels but not tightly enough to bind this substrate within the channel and thus hamper enzymatic turnover. In this respect, therefore, the CO_2 migration pathways depicted in Figure 2 for PEPCK are intermediate between two previously identified extremes: the pipe for NH_3 transport proposed by Wang et al.[49] in the enzyme GPATase, where MD simulations with multiple copies of NH_3 found ammonia to be evenly distributed, implying a lack of strong, localized protein−ligand interactions, and the self-opening mechanism found by Tomita et al.,[21] where CO interacts with the protein with sufficient strength to induce a change in the local protein structure, thereby opening its own passage between cavities. The issue of ligands influencing protein geometry is of special interest for PEPCK, as Holyoak and coworkers[50−52] have investigated the link between closure of the lid domain (indicated by the asterisk in Figure 2), substrate binding (Figure 1), and stabilization of the enolate intermediate. Although a full exploration of this issue is beyond the scope of this Letter, our results do not indicate a correlation between CO_2 and the lid state (open and closed). That is, in various MD trajectories, the lid can either open or close whether CO_2 is still in the active site cleft or if it has already completed its exit from the enzyme.

Further characterization of the balanced role proposed for the amino acids along these CO_2 paths could be readily provided by experiment. For example, replacing Thr residues (e.g., 388 or 404) with Cys via site-directed mutagenesis would minimally affect the steric profile of the pathways (by 8 Å^3)[53] but would cut the strength of the already weak Thr-CO_2 interaction in half,[10] presumably affecting the transport of CO_2 to the active site and thereby influencing enzymatic efficiency. Alternatively, Thr388 or Thr404 could be replaced by Val to judge the effect of weakly polar residues vis-a-vis hydrophobic residues in directing CO_2. To test the hypothesis that a basic residue located along a path could serve as a CO_2 trap, thus hindering catalysis, Phe 365 (or 405) could be replaced with His. Finally, mutations designed to simply add steric bulk to these pathways and thus restrict CO_2 motion could be pursued, such as replacement of Asn with Gln or Ala with Val. Blocking migration channels with steric bulk has previously been shown to be an effective way to probe ligand diffusion pathways in experiments conducted on hemoglobin,[16] myoglobin,[17] and neuroglobin.[14] Indeed, it may prove to be beneficial to block selectively one or even two of the proposed CO_2 migration paths to artificially favor the direction of ligand approach that most efficiently produces OAA with the desired stereochemistry.[54] Such an enhancement could lead to a new means of producing bioderived succinic acid, which has myriad current and potential uses in pharmaceutical, polymer, and surfactant industries.[55]

The CO_2 migration paths elucidated by this work, as well as the computational protocol used herein to study them, can also be used to inform bioinspired carbon capture approaches based on an enzymatic platform.[6,7] For example, if PEPCK is to be used, the efficiency of carbon capture should improve if steric bulk is simply removed from the CO_2 migration paths shown in Figure 2; indeed, such modifications may open up new pathways for CO_2 transfer to the active site. Moreover, carbon capture efficacy should improve if basic reservoir residues (such as Arg, His, or Lys) can be grafted onto the solvent-exposed end of CO_2 transfer pathways.

In summary, using extensive MD simulations, CO_2 was found to exit the experimentally known binding site of the enzyme PEPCK through three distinct pathways, suggesting that the protein is actively involved in the transport of substrate CO_2 to the active site. The residues along these paths are not the highly CO_2-philic amino acids Arg, His, or Lys but rather are hydrophobic or only weakly polar (e.g., Thr), possibly because basic residues would serve as localized traps for migrating CO_2, which could reduce catalytic efficiency. Such CO_2-philic residues, however, may be useful at the solvent-accessible ends of these pathways. Further work to explore this possibility is underway.

ASSOCIATED CONTENT

Supporting Information

Technical details of the MD simulations. 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

We acknowledge the U.S. Department of Energy (BER-08ER64603) for their generous support of this research.

REFERENCES

(1) *Catalysts for Nitrogen Fixation: Nitrogenases, Relevant Chemical Models and Commercial Processes*; Smith, B. E., Richards, R. L., Newton, W. E., Eds.; Kluwer Academic Publishers: Norwell, MA, 2004.

(2) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*; W. H. Freeman: New York, 2005.

(3) *The Role of Nitric Oxide in Heart Failure*; Jugdutt, B. I., Ed.; Springer-Verlag: New York, 2007.

(4) Schneider, G.; Lindqvist, Y.; Branden, 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) Ozdemir, E. *Energy Fuels* **2009**, *23*, 5725−5730.

(7) Bond, G. M.; Stringer, J.; Brandvold, D. K.; Simsek, F. A.; Medina, M.-G.; Egeland, G. *Energy Fuels* **2001**, *15*, 309−316.

(8) Mahato, S.; De, D.; Dutta, D.; Kundu, M.; Bhattacharya, S.; Schiavone, M. T.; Bhattacharya, S. K. *Microb. Cell Fact.* **2004**, *3*, 7.

(9) Tans, P. NOAA/ESRL. (www.esrl.noaa.gov/gmd/ccgg/trends/) accessed March 7, 2012.

(10) 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.

(11) Drummond, M. L.; Wilson, A. K.; Cundari, T. R. *J. Mol. Model.* **2011**, DOI: 10.1007/s00894-011-1276-0, accepted for publication.

(12) Drummond, M. L.; Wilson, A. K.; Cundari, T. R. *Energy Fuel.* **2010**, *24*, 1464−1470.

(13) Knapp, J. E.; Pahl, R.; Cohen, J.; Nichols, J. C.; Schulten, K.; Gibson, Q. H.; Srajer, V.; Royer, W. E. Jr. *Structure* **2009**, *17*, 1494−1504.

(14) Lutz, S.; Nienhaus, K.; Nienhaus, G. U.; Meuwly, M. *J. Phys. Chem. B.* **2009**, *113*, 15334−15343.

(15) Nienhaus, K.; Lutz, S.; Meuwly, M.; Nienhaus, G. U. *ChemPhysChem* **2010**, *11*, 119−129.

(16) Birukou, I.; Soman, J.; Olson, J. S. *J. Biol. Chem.* **2011**, *286*, 10515−10529.

(17) Brunori, M.; Vallone, B.; Cutruzzola, F.; Travaglini-Allocatelli, C.; Berendzen, J.; Chu, K.; Sweet, R. M.; Schlichting, I. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2058−2063.

(18) Chu, K.; Vojtchovsky, J.; McMahon, B. H.; Sweet, R. M.; Berendzen, J.; Schlichting, I. *Nature* **2000**, *403*, 921−923.

(19) Nienhaus, K.; Knapp, J. E.; Palladino, P.; Royer, W. E. Jr.; Nienhaus, G. U. *Biochemistry* **2007**, *46*, 14018−14031.

(20) Ostermann, A.; Waschipky, R.; Parak, F. G.; Nienhaus, G. U. *Nature* **2000**, *404*, 205−208.

(21) Tomita, A.; Sato, T.; Ichiyanagi, K.; Nozawa, S.; Ichikawa, H.; Chollet, M.; Kawai, F.; Park, S.-Y.; Tsuduki, T.; Yamato, T.; Koshihara, S.-Y.; Adachi, S.-I. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2612−2616.

(22) Anselmi, M.; Brunori, M.; Vallone, B.; Di Nola, A. *Biophys. J.* **2007**, *93*, 434−441.

(23) Anselmi, M.; Di Nola, A.; Amadei, A. *J. Phys. Chem. B.* **2011**, *115*, 2436−2446.

(24) Anselmi, M.; Di Nola, A.; Amadei, A. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 867−879.

(25) Bidon-Chanal, A.; Martí, M. A.; Estrin, D. A.; Luque, F. J. *J. Am. Chem. Soc.* **2007**, *129*, 6782−6788.

(26) Bocahut, A.; Bernad, S.; Sebban, P.; Sacquin-Mora, S. *J. Phys. Chem. B.* **2009**, *113*, 16257−16267.

(27) Bossa, C.; Amadei, A.; Daidone, I.; Anselmi, M.; Vallone, B.; Brunori, M.; Di Nola, A. *Biophys. J.* **2005**, *89*, 465−474.

(28) Ceccarelli, M.; Anedda, R.; Casu, M.; Ruggerone, P. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 1231−1236.

(29) Cohen, J.; Arkhipov, A.; Braun, R.; Schulten, K. *Biophys. J.* **2006**, *91*, 1844−1857.

(30) Elber, R.; Gibson, Q. H. *J. Phys. Chem. B.* **2008**, *112*, 6147−6154.

(31) Heroux, M. S.; Mohan, A. D.; Olsen, K. W. *IUBMB Life* **2011**, *63*, 214−220.

(32) Mishra, S.; Meuwly, M. *Biophys. J.* **2009**, *96*, 2105−2118.

(33) Nishihara, Y.; Hayashi, S.; Kato, S. *Chem. Phys. Lett.* **2008**, *464*, 220−225.

(34) Nishihara, Y.; Kato, S.; Hayashi, S. *Biophys. J.* **2010**, *98*, 1649−1657.

(35) Ruscio, J. Z.; Kumar, D.; Shukla, M.; Prisant, M. G.; Murali, T. M.; Onufriev, A. V. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9204−9209.

(36) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161−9175.

(37) Teixeira, V. H.; Baptista, A. M.; Soares, C. M. *Biophys. J.* **2006**, *91*, 2035−2045.

(38) Swails, J. M.; Meng, Y.; Walker, F. A.; Martí, M. A.; Estrin, D. A.; Roitberg, A. E. *J. Phys. Chem. B* **2009**, *113*, 1192−1201.

(39) Zhang, Y.; Lu, M.; Cheng, Y.; Li, Z. *J. Mol. Graph. Model.* **2010**, *28*, 814−819.

(40) Merz, J., K. M. *J. Am. Chem. Soc.* **1991**, *113*, 406−411.

(41) Liang, J.-Y.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3675−3679.

(42) Krebs, A.; Bridger, W. A. *Can. J. Biochem.* **1980**, *58*, 309−318.

(43) Arnelle, D. R.; O'Leary, M. H. *Biochemistry* **1992**, *31*, 4363−4368.

(44) 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.

(45) Du, Q.-S.; Fan, X.-W.; Wang, C.-H.; Huang, R.-B. *PLoS ONE* **2011**, *6*, e24264.

(46) Domsic, J. F.; McKenna, R. *Biochim. Biophys. Acta* **2010**, *1804*, 326−331.

(47) DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. *J. Chem. Phys.* **2006**, *124*, 114104.

(48) Leung, K.; Nielsen, I. M. B.; Kurtz, I. *J. Phys. Chem. B.* **2007**, *111*, 4453−4459.

(49) Wang, X. S.; Roitberg, A. E.; Richards, N. G. J. *Biochemistry* **2009**, *48*, 12272−12282.

(50) Sullivan, S. M.; Holyoak, T. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13829−13834.

(51) Carlson, G. M.; Holyoak, T. *J. Biol. Chem.* **2009**, *284*, 27037−27041.

(52) Johnson, T. A.; Holyoak, T. *Biochemistry* **2010**, *49*, 5716−5187.

(53) Zamyatnin, A. A. *Prog. Biophys. Mol. Biol.* **1972**, *24*, 107−123.

(54) Perez, E.; Espinoza, R.; Laivenieks, M.; Cardemil, E. *Biochimie* **2008**, *90*, 1685−1692.

(55) Zeikus, J. G.; Jain, M. K.; Elankovan, P. *Appl. Microbiol. Biotechnol.* **1999**, *51*, 545−552.