Gas-Phase Geometry Optimization of Biological Molecules as a Reasonable Alternative to a Continuum Environment Description: Fact, Myth, or Fiction?†

Sérgio Filipe Sousa, Pedro Alexandrino Fernandes, and Maria João Ramos*

REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal

Received: March 12, 2009; Revised Manuscript Received: July 13, 2009

Gas-phase optimization of single biological molecules and of small active-site biological models has become a standard approach in first principles computational enzymology. The important role played by the surrounding environment (solvent, enzyme, both) is normally only accounted for through higher-level single point energy calculations performed using a polarizable continuum model (PCM) and an appropriate dielectric constant with the gas-phase-optimized geometries. In this study we analyze this widely used approximation, by comparing gas-phase-optimized geometries with geometries optimized with different PCM approaches (and considering different dielectric constants) for a representative data set of 20 very important biological molecules—the 20 natural amino acids. A total of 323 chemical bonds and 469 angles present in standard amino acid residues were evaluated. The results show that the use of gas-phase-optimized geometries can in fact be quite a reasonable alternative to the use of the more computationally intensive continuum optimizations, providing a good description of bond lengths and angles for typical biological molecules, even for charged amino acids, such as Asp, Glu, Lys, and Arg. This approximation is particularly successful if the protonation state of the biological molecule could be reasonably described in vacuum, a requirement that was already necessary in first principles computational enzymology.

Introduction

The environment plays a vital role in the large plethora of biochemical phenomena, and its effect is often essential for a correct atomistic description of molecular biological systems and for an accurate determination of many of the properties associated. In particular, water, the biological solvent of choice and the most profuse constituent of living organisms, plays a particularly important role in biological processes. The inclusion of the effect of the solvent in computational models is particularly challenging, although several methods able to represent molecules in solution, at different levels of sophistification, have been developed.[1−5] The level of detail used to describe the chemical system, the physical rules underlying the process of interest, and the mathematical formulas employed in describing these rules are among the various features that distinguish between the different alternatives available.[2] While gas-phase predictions can render faster and very accurate results for some chemical processes and molecular properties, there is a whole range of phenomena and molecular features that cannot be accurately addressed by such means. In these cases, the influence of the solvent has to be accounted for.

The polarizable continuum model (PCM) has become during the past decades a strong and widely applied alternative in the gamut of available computational methodologies.[6] PCM methods allow the inclusion of the effect of the solvent in standard ab initio or density functional theory (DFT) calculations, by modeling the solvent as a polarizable continuum dielectric with a specific constant instead of using individual molecules. Over the years, PCM has been successfully applied in the determination of geometries and of a wide range of molecular properties in the study of reaction mechanisms and spectroscopic phenomena (for reviews on this topic please see refs 1, 2, 7, and 8). Nevertheless, and despite the immense computational improvement that has characterized these years, for many applications in computational chemistry and biochemistry, its use is often avoided, particularly for geometry optimization.

The geometry of a molecule determines many of the physical and chemical properties associated with it. In computational chemistry, geometry optimization is used to find the structure of the minima on a potential energy surface, with these minimum energy structures representing equilibrium states. In addition, geometry optimization is also employed in locating transition state structures, which are saddle points on the potential energy surface. During this process the energy of molecules is reduced by adjusting atomic coordinates, yielding optimal bond lengths, bond angles, and dihedrals. High accuracy in determining these values is of great importance in the computational simulation of biological processes as experimental information on such quantities is often scarce and difficult to obtain. Geometry optimizations are therefore a critical component in the wide plethora of computational chemistry applications, as most of these applications are essentially based on the characterization of stationary states. Geometry optimizations with PCM typically involve a higher computational cost (in terms of CPU time) by optimization cycle than gas-phase calculations do. In addition, geometry convergence is normally more difficult to achieve, due to the decreased steepness of the potential energy surfaces.

It is important to take into consideration that, in general, the energy of a molecular system has a rather slow dependence on the particular geometric parameters that define a given conformation. In fact, energetic differences between conformers optimized with several different combinations method/basis set calculated with a single point energy calculation with a single method/basis set will typically vary much less than the differences arising from using single point energy calculations with

† Part of the "Vincenzo Aquilanti Festschrift".

* Corresponding author, mjramos@fc.up.pt.

10.1021/jp902213t
Published on Web 09/28/2009

different methods/basis set on conformers optimized with a unique combination method/basis set. This fact, together with the difficulties outlined above cause the influence of the environment in geometries to be frequently only included at a subsequent stage, through PCM energy calculations performed on the gas-phase-optimized geometries. In fact, the application of higher-level single point energy calculations with PCM (typically with a dielectric constant of around 4 for an enzymatic reaction or of 80 for a reaction in water) in gas-phase geometries is still the standard approach in first principles quantum mechanical computational enzymology,[9−11] continuing to provide highly valuable insights into the catalytic mechanism of many important enzymes. Recent notable examples include the molybdoenzyme formate dehydrogenase,[12] haloalcohol dehalogenase,[13] the [NiFe] and [FeFe] hydrogenases,[14] the apocarotenoid oxygenase,[15] glutaminyl cyclase,[16] the prostatic acid phosphatase,[17] the farnesyltransferase,[18−22] the dinuclear zinc enzyme dihydroorotase,[23] and several more general Zn biological systems.[24,25]

The big question that arises from the use of this approximation is how large an error are we introducing by admitting that the gas-phase geometries are a reasonable alternative to a description that includes the effect of the solvent in geometry optimization? Logically, the answer will depend on the particular molecule or biological system that one is studying; but is there an intrinsic systematic error for typical biological molecules in this approximation? Is there a significant gain that could be obtained by optimizing the geometries in a continuum environment? To address these questions we have chosen a standard first principles quantum mechanical method and a commonly used basis set for geometry optimization to evaluate the differences in the optimized bond lengths and angles that could be obtained by optimizing the geometry of a data set of typical individual molecules of high biological significance in the gas phase and with a polarizable continuum model with dielectric constants of 4 (protein) and 80 (water), for three different PCM methods. An unsuspected data set comprising the 20 natural amino acid residues in their most stable protonation states at pH 7 was chosen for this effect. This selection of molecules ensured a structurally balanced, chemically diverse, biologically representative, and most of all unbiased choice of test cases, illustrative of the type of chemical bonds and angles normally present in the quantum mechanical models employed in computational enzymology.

The results show that the gain in accuracy in the determination of bond lengths and angles for single biological molecules by optimizing the geometries with PCM methods can be in general quite small, compared with the computational cost associated, particularly if the protonation state of the biological molecule could be reasonably attributed in vacuum.

Computational Methods

The standard approach in quantum-mechanical computational enzymology for the determination of molecular geometries of standard biological molecules is the optimization of the geometries in vacuum, with the introduction of the effect of the environment only in a single-point energy calculation with an appropriate continuum solvation model and dielectric constant at a later stage.[9−11] The purpose of this study was to evaluate the gain, in terms of accuracy, that might be obtained in the determination of geometries for such molecules when performing the corresponding geometry optimizations while using a continuum solvation model instead of the gas-phase optimizations.

Calculations were carried out using the Gaussian 03 suite of programs.[26] Three commonly available continuum solvation methods were considered. These were the Polarizable Conductor Continuum model (C-PCM),[27,28] the Integral-Equation-Formalism Polarizable Continuum Model (IEF-PCM),[8,29−31] and the Self-Consistent Isodensity Polarizable Continuum Model (SCI-PCM),[32] which perform an SCRF calculation using a cavity determined self-consistently from an isodensity surface (standard value of 0.0004 a.u. for the isodensity used). Geometry optimizations at the B3LYP/6-31+G(d,p) level of theory were performed with each one of these three solvation methods in both water ($\epsilon = 78.39$) and protein ($\epsilon = 4.335$). Final optimized geometries were compared with the reference geometries and with the gas-phase geometries optimized at the same level of theory. The popular density functional B3LYP was chosen as it still remains today the most widely used alternative for standard DFT calculations.[33] The 6-31+G(d,p) is also a common choice for geometry optimization in general organic chemistry.

The choice to perform such calculations also in the protein environment was made taking into consideration that the use of a dielectric constant of around 4 is normally taken as an approximation to the effect of the global enzyme environment on a reaction, generally giving a good agreement with experimental results, and accounting for the average effect of both the protein and buried water molecules.[34−36]

The data set considered in these calculations comprised the 20 natural amino acids normally present in proteins, in their most stable protonation states at pH 7, as they represent a very structurally diverse and biologically important set of molecules, typically present in quantum mechanical computational enzymology models. High-quality reference values for the geometries of these molecules were taken from the Cambridge Structural Database (CSD)[37] and chosen from the alternatives with the lowest R-values. Details are presented in Table 1. The reader should however be aware that comparing equilibrium structures in the gas phase or in continuum with thermally averaged crystallographic structures is by itself an approximation and has a certain degree of error associated. However, it should be noticed, that the standard procedure in quantum mechanical computational enzymology also involves comparing the structures obtained for reaction minima or transition states with X-ray or NMR structures of the corresponding enzymes and closest states in the Protein Data Bank.[38] Hence, the protocol followed

TABLE 1: Experimental Reference Structures for the 20 Amino Acids Considered in This Study, with Indication of the Corresponding *R* Factor Value and Original Reference

amino acid	*R* factor	ref
alanine	0.0203	39
arginine	0.0349	40
asparagine	0.175	41
aspartic acid	0.0106	42
cysteine	0.0375	43
glutamic acid	0.026	44
glutamine	0.043	45
glycine	0.0129	46
histidine	0.0296	47
isoleucine	0.040	48
leucine	0.058	49
lysine	0.0333	50
methionine	0.084	51
phenylalanine	0.084	52
proline	0.0208	53
serine	0.020	54
threonine	0.094	55
tryptophan	0.088	56
tyrosine	0.026	57
valine	0.0435	58

TABLE 2: Overview of the Mean Unsigned Error (MUE) and Mean Signed Error (MSE) Calculated for the Three Data Sets Considered in This Study: BL1 (all bond lengths); BL2 (all bond lengths except the ones associated to the amino acid charged NH3 terminus); AA (all angles)

			water ($\epsilon = 78.39$)			protein ($\epsilon = 4.335$)		
data sets	error type	gas phase	C-PCM	IEF-PCM	SCI-PCM[a]	C-PCM	IEF-PCM	SCI-PCM[a]
bond lengths BL1 (Å)	MUE	0.077	0.037	0.037	0.014	0.041	0.041	0.022
	MSE	0.065	0.030	0.030	0.007	0.033	0.033	0.012
bond lengths BL2 (Å)	MUE	0.040	0.035	0.035	0.014	0.038	0.038	0.018
	MSE	0.027	0.029	0.029	0.010	0.031	0.031	0.011
angles AA (deg)	MUE	3.32	2.02	2.02	2.99	2.47	2.55	4.03
	MSE	−0.11	−0.03	−0.03	−1.00	−0.11	−0.12	−1.62

[a] With the SCI-PCM method, optimized geometries could only be obtained for Gly.

for this data set of high-quality amino acid structures reproduces to a significant extent the approach normally used in first principles computational enzymology.

Results and Discussion

To evaluate the ability of the different approaches in geometry optimization for the data set considered, the bond lengths and angles of the resulting structures were compared with the corresponding reference values in the structures from the Cambridge Structural Database[37] (Table 1). The mean signed error (MSE) and the mean unsigned error (MUE) were used in this regard. The MSE is taken as the difference between the values calculated with the approach tested and the correspondent values reported in the CSD structures. A negative MSE indicates that the application of given methodology to the type of systems considered underestimates the value of a given parameter (in this case, bond lengths and angles), whereas a positive MSE indicates that the value is overestimated. The mean unsigned error (MUE) is the module of the difference between the value calculated for a given geometric parameter and the value reported in the CSD structure. The MUE gives an indication of the accuracy of the approach tested in calculating the geometric parameters indicated for the systems considered. Table 2 presents an overview of the mean unsigned error (MUE) and mean signed error (MSE) calculated for all the bond lengths and angles present in the 20 natural amino acids in the gas phase and with the three PCM methods tested for dielectric constants of 78.39 (water) and 4.335 (protein).

The results were grouped into three data sets. The first data set (BL1) is a blind database of the 323 amino acid bond lengths evaluated in this study. BL2 is a data set of 279 amino acid bond lengths selected from BL1 by removing all the bond lengths associated to the charged amino acid NH_3^+ terminus. This second data set was prepared from the first, because all the amino acids considered in this study and the corresponding experimental reference structures are for the zwitterionic state, i.e., with charged NH_3^+ and COO^-. This is certainly not a reasonable protonation state for the gas phase. Gas-phase optimizations of an amino acid residue in the zwitterionic state typically results in an H^+ transfer from the NH_3^+ group to the COO^- group. The distance of this proton to the initial N-terminus almost doubles, increasing the error associated to the gas-phase description in comparison to the description based in the use of a continuum model. BL2 was therefore prepared to ensure a fairer comparison of the gas-phase description with the remaining alternatives, taking into consideration that the use of gas-phase-optimized geometries of biological molecules as an alternative to a continuum description is only attempted when a coherent gas-phase protonation state can be attributed. Finally, the third data set—AA—comprises all the 469 bond angles present in the amino acid molecules evaluated.

The results show that all the methods tested tend to overestimate the experimental bond lengths, particularly the gas-phase optimizations in the BL1 data set, a feature that can be attributed to the proton transfer from the NH_3^+ group to the COO^- taking place during the optimization in the gas phase. An important observation concerns the SCI-PCM method. With this methodology, optimized geometries were only obtained for the smallest amino acid, glycine. All the other geometry optimizations with this method systematically failed, while the optimization with the C-PCM and IEF-PCM methods was successful for all the amino acids. In addition, for glycine the computational cost associated to the SCI-PCM approach was much higher than the one associated to the C-PCM and IEF-PCM models (for example in water SCI-PCM calculation were 8 times longer), with the resulting optimized geometry displaying higher MUE and MSE values than the alternative methods for this particular amino acid residue (data not shown).

Table 2 further shows that the gas-phase optimizations resulted in a MUE of 0.077 Å for the BL1 data set, whereas MUEs of 0.037 Å were obtained for both C-PCM and IEF-PCM in water. The same methods in the protein environment gave MUEs of 0.041 Å. These results show that even with unreasonable gas-phase protonation states affecting the bond lengths included in the data set, gas-phase optimizations provide a very reasonable approximation to the C-PCM and IEF-PCM optimization, particularly if these optimizations were to be performed in a protein environment (MUEs of 0.077 Å in the gas phase against 0.041 Å in the protein environment), as it would probably be the case in first principles computational enzymology, for which a dielectric constant of 4 is generally taken as an approximation to the global effect of the enzyme and buried water molecules.

For the more reasonable bond lengths data set (BL2, 279 bond lengths), which is expected to reflect better the type of models and molecules used in standard computational enzymology, as in such models the saturation of the N- and C-terminus is normally preferred, the MUE obtained for the gas phase was of 0.040 Å, while MUEs of 0.038 and 0.035 Å were obtained with both C-PCM and IEF-PCM for the protein environment and water, respectively. These differences are remarkably small (only 0.002 Å difference between protein environment optimizations and gas-phase calculations), showing how strong an approximation of gas-phase optimizations can be. For the data set of 469 amino acid bond angles, MUEs of 3.32, 2.55, and 2.02° were obtained for the gas-phase, protein, and water optimization, respectively, once again with C-PCM and IEF-PCM exhibiting very similar performance.

Table 3 presents the MUEs for the BL2 bond-length and AA bond-angles data set by type of amino acid. The purpose of this analysis was to determine how the success of this approximation could vary between polar and nonpolar molecules

TABLE 3: Overview of the Mean Unsigned Error (MUE) for the BL2 Bond-Length Data Set and AA Bond-Angle Data Set Grouped by Type of Amino Acid

by amino acid type	property	no. of cases	gas phase	water ($\epsilon = 78.39$)			protein ($\epsilon = 4.335$)		
				C-PCM	IEF-PCM	SCI-PCM[a]	IEF-PCM	C-PCM	SCI-PCM[a]
polar	bond lengths (Å)	138	0.034	0.029	0.029		0.034	0.032	
	angles (deg)	246	3.21	2.03	2.03		2.47	2.52	
nonpolar	bond lengths (Å)	141	0.046	0.041	0.041	0.014	0.043	0.043	0.018
	angles (deg)	223	3.44	2.00	2.01	2.99	2.47	2.56	4.03
charged	bond lengths (Å)	67	0.020	0.015	0.015		0.020	0.013	
	angles (deg)	130	2.96	1.86	1.84		2.27	2.22	
noncharged	bond lengths (Å)	212	0.046	0.041	0.041	0.014	0.042	0.042	0.018
	angles (deg)	339	3.46	2.08	2.08	2.99	2.53	2.62	4.03

[a] With the SCI-PCM method optimized geometries could only be obtained for Gly.

TABLE 4: Mean Signed Error (MSE) for the Bond Lengths in the 20 Natural Amino Acid Residues by Bond Type

by amino acid type	no. of cases	gas phase	water ($\epsilon = 78.39$)			protein ($\epsilon = 4.335$)		
			C-PCM	IEF-PCM	SCI-PCM[a]	C-PCM	IEF-PCM	SCI-PCM[a]
C−C	88	0.011	0.011	0.011	0.031	0.013	0.013	0.038
C−N	32	−0.006	0.015	0.015	0.021	0.018	0.017	0.021
C−O	48	0.018	0.006	0.006	0.000	0.005	0.005	−0.001
C−S	3	0.030	0.031	0.031		0.032	0.032	
C−H	92	0.061	0.062	0.062	0.004	0.067	0.066	0.004
N−H (BL1)	58	0.233	0.038	0.038	0.003	0.046	0.045	0.015
N−H (BL2)	14	0.021	0.032	0.031		0.039	0.028	
O−H	2	−0.004	0.016	0.016		0.010	0.008	

[a] With the SCI-PCM method, optimized geometries could only be obtained for Gly.

TABLE 5: Mean Unsigned Error (MUE) for the Bond Lengths in the 20 Natural Amino Acid Residues by Bond Type

by amino acid type	no. of cases	gas phase	water ($\epsilon = 78.39$)			protein ($\epsilon = 4.335$)		
			C-PCM	IEF-PCM	SCI-PCM[a]	C-PCM	IEF-PCM	SCI-PCM[a]
C−C	88	0.015	0.016	0.016	0.031	0.019	0.018	0.038
C−N	32	0.016	0.021	0.021	0.021	0.022	0.023	0.021
C−O	48	0.056	0.023	0.023	0.011	0.025	0.025	0.020
C−S	3	0.030	0.031	0.031		0.032	0.032	
C−H	92	0.063	0.063	0.063	0.005	0.069	0.067	0.005
N−H (BL1)	58	0.248	0.047	0.047	0.015	0.056	0.056	0.029
N−H (BL2)	14	0.042	0.043	0.043		0.052	0.042	
O−H	2	0.011	0.016	0.016		0.011	0.011	

[a] With the SCI-PCM method, optimized geometries could only be obtained for Gly.

and between charged and noncharged compounds. For the 138 bond lengths present in polar amino acids, a MUE of 0.034 Å was obtained for the structures optimized in the gas phase. C-PCM and IEF-PCM gave MUEs of 0.029 Å in water, while in the protein environment model ($\epsilon = 4$) values of 0.034 and 0.032 Å were obtained. So once again, the accuracy of gas-phase optimizations and PCM optimizations in the protein environment is shown to be quite similar. The same general trends were obtained for the data set of 246 bond angles in polar amino acid residues, with the gas-phase optimization showing a MUE of 3.21°, against 2.03° in water with both C-PCM and IEF-PCM, and 2.47° (C-PCM) and 2.52° (IEF-PCM) in protein.

In nonpolar amino acids, a very similar general trend between the MUE calculated in vacuum, protein, and water was obtained. While the MUE values obtained for the bonds and angles present in this type of amino acids are higher than those calculated for the polar amino acids, the relative differences between the MUEs in the gas-phase optimizations and in continuum optimizations is now smaller. Very similar trends and values, albeit with a higher number of bonds and angles, were obtained for the noncharged amino acids (i.e., all with the exception of Arg, Asp, Glu, Lys).

For the group of bond lengths and angles present in charged amino acid residues (67 bonds and 130 angles), the relative difference between a gas-phase description and a continuum description was the highest in the test (0.007 Å difference in relation to a 0.013 IEF-PCM protein MUE). Nevertheless, the accuracy of gas-phase optimizations in terms of bonds lengths and angles was quite good, with MUEs of 0.020 Å and 2.96°. These results show that the use of gas-phase-optimized geometries of biological molecules as an approximation to geometries optimized in the presence of a continuum dielectric can work pretty well, not only for nonpolar and noncharged amino acids but also for the polar and charged molecules in the data set considered.

With Table 4 and Table 5 we try to see how this approximation works for different types of bonds. Table 4 and Table 5 present respectively the MSEs and MUEs calculated for the set of 20 natural amino acids considered in this study, by bond type. From Table 4 it is evident that with the exception of the C−N and O−H bonds in the gas phase, the bond lengths of all other bond types are typically overestimated. C−C bonds are typically overestimated by all methods tested by ca. 0.01 Å, while in N−H (BL2) this value is typically around 0.03 Å. For C−H all methods tend to overestimate bond lengths by something like 0.06 Å.

Table 5 shows that the gas-phase optimizations provided a very good approximation to a continuum description for C−C,

C−N, C−S, C−H, O−H, and N−H (for coherent protonation states) bond types. For C−O bonds however, more significant differences were obtained, with the gas-phase optimization resulting in a MUE of 0.056 Å, while the continuum optimization in water and protein resulted in MUEs of only 0.031 and 0.032 Å. A possible explanation for the increased bond lengths verified in the gas phase for the C−O bonds is the existence of intramolecular hydrogen bonds. In gas-phase optimizations, the importance and energetic impact of possible intramolecular hydrogen bonds is very large. The intramolecular hydrogen bonds formed with the carbonyl oxygen atom often decrease the C−O bond strength and lead to increased C−O bond lengths. However, continuum optimizations typically stabilize much more the groups that could participate in such intramolecular hydrogen bonds, decreasing the importance of these interactions in the final optimized geometry and therefore yielding shorter and stronger C−O bond lengths. In this regard, the continuum optimizations performed in water ($\epsilon = 80$) are typically more effective than the continuum optimizations performed while considering a model of the protein environment ($\epsilon = 4$), a feature that could explain the higher MUE obtained for the second case (0.025 Å vs 0.023 Å).

Conclusions

This study has shown that the gas-phase optimization of single biological molecules can be a very reasonable alternative to the optimization of such molecules using the more computational intensive PCM optimizations, in which the effect of the environment is partially accounted for through the use of a continuum dielectric constant. For cases in which the protonation state of the biological molecule can be reasonably described in vacuum, gas-phase optimizations of standard biological molecules, as the 20 amino acids considered in this study, yield average mean unsigned errors that are at the same level of the ones obtained in optimizations performed with different PCM methods for the dielectric constants of water and of a generic protein environment. In particular, for bond lengths these differences in MUEs are as low as 0.005 Å in comparison with the optimizations performed in water, and as 0.002 Å in comparison with the ones performed in the protein environment. Also for angles, differences between the two general types of geometry optimizations yield average differences of less than 1°.

These general trends were maintained between different types of amino acids, including polar and nonpolar, and charged and uncharged molecules. Naturally, the difference in accuracy between gas-phase optimizations and PCM optimizations was higher in charged amino acids. However, even for these molecules average MUEs of only 0.020 Å and 2.96° were obtained. In general, the good performance of gas-phase optimizations was observed for different bond types particularly C−C, C−N, and O−H bonds. C−O bonds were the only ones to reflect a systematic accuracy difference between gas-phase and PCM optimizations, with the former resulting in average MUEs three times higher than the latter.

These observations suggest the use of gas-phase-optimized geometries in first principles computational enzymology, in which single biological molecules or small active-site biological models are employed to model enzymatic reactions, with the effect of the environment being included only through PCM single point energy calculations. Some caution should however be taken into consideration when studying systems with transition state structures for which very significant charge separation takes place in comparison with the two energy minima, or which are significantly more polar or electronegative. For such systems the use of gas-phase-optimized geometries with PCM single point energy calculations could result in doubtful activation barriers. The reader is also advised that although such an approach could be very reasonable for systems and molecules with dimensions and properties very close to the ones included in the data set considered in this study, for enzyme active sites models comprised by several nonconstrained amino acid residues, and for biological systems that incorporate two or more charged molecules, or whose structure depends on a great part of the formation of specific hydrogen bonds, these differences may be much larger, and PCM optimizations might need to be employed.

Supporting Information Available: Gas-phase equilibrium structures of all amino acids in the gas phase and CSD reference structures. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
(2) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.
(3) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.
(4) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Acc. Chem. Res.* **2000**, *33*, 889.
(5) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760.
(6) Tomasi, J. *Theor. Chem. Acc.* **2004**, *112*, 184.
(7) Orozco, M.; Luque, F. J. *Chem. Rev.* **2000**, *100*, 4187.
(8) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
(9) Himo, F. *Theor. Chem. Acc.* **2006**, *116*, 232.
(10) Leopoldini, M.; Marino, T.; Michelini, M. D.; Rivalta, I.; Russo, N.; Sicilia, E.; Toscano, M. *Theor. Chem. Acc.* **2007**, *117*, 765.
(11) Ramos, M. J.; Fernandes, P. A. *Acc. Chem. Res.* **2008**, *41*, 689.
(12) Leopoldini, M.; Chiodo, S. G.; Toscano, M.; Russo, N. *Chemistry* **2008**, *14*, 8674.
(13) Hopmann, K. H.; Himo, F. *J. Chem. Theor. Comput.* **2008**, *4*, 1129.
(14) Siegbahn, P. E. M.; Tye, J. W.; Hall, M. B. *Chem. Rev.* **2007**, *107*, 4414.
(15) Borowski, T.; Blgomberg, M. R. A.; Siegbahn, P. E. M. *Chemistry* **2008**, *14*, 2264.
(16) Calvaresi, M.; Garavefli, M.; Bottoni, A. *Proteins* **2008**, *73*, 527.
(17) Sharma, S.; Rauk, A.; Juffer, A. H. *J. Am. Chem. Soc.* **2008**, *130*, 9708.
(18) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Comput. Chem.* **2007**, *28*, 1160.
(19) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Proteins* **2007**, *66*, 205.
(20) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J *Biophys. J.* **2005**, *88*, 483.
(21) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Mol. Struct.: THEOCHEM* **2005**, *729*, 125.
(22) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Chemistry* **2009**, *15*, 4243.
(23) Liao, R. Z.; Yu, J. G.; Raushel, F. M.; Himo, F. *Chemistry* **2008**, *14*, 4287.
(24) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Am. Chem. Soc.* **2007**, *129*, 1378.
(25) Tamames, B.; Sousa, S. F.; Tamames, J.; Fernandes, P. A.; Ramos, M. J. *Proteins* **2007**, *69*, 466.
(26) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; 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. ; Hratchin, H. P.; Cross, J. B.; 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.; Malik, 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-Lahan, 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.: Wallingford, CT, 2004.

(27) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995.

(28) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.

(29) Cances, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032.

(30) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151.

(31) Tomasi, J.; Mennucci, B.; Cances, E. *J. Mol. Struct.: THEOCHEM* **1999**, *464*, 211.

(32) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098.

(33) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007**, *111*, 10439.

(34) Blomberg, M. R. A.; Siegbahn, P. E. M.; Babcock, G. T. *J. Am. Chem. Soc.* **1998**, *120*, 8812.

(35) Siegbahn, P. E. M. *J. Am. Chem. Soc.* **1998**, *120*, 8417.

(36) Siegbahn, P. E. M.; Eriksson, L. A.; Himo, F.; Pavlov, M. *J. Phys. Chem. B* **1998**, *102*, 10622.

(37) Allen, F. H. *Acta Crystallogr., Sect. B* **2002**, *58*, 380.

(38) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.

(39) Destro, R.; Marsh, R. E.; Bianchi, R. *J. Phys. Chem.* **1988**, *92*, 966.

(40) Petrosyan, A. M.; Karapetyan, H. A.; Sukiasyan, R. P.; Aghajanyan, A. E.; Morgunov, V. G.; Kravchenko, E. A.; Bush, A. A. *J. Mol. Struct.* **2005**, *752*, 144.

(41) Pasternak, R. A.; Katz, L.; Corey, R. B. *Acta Crystallogr.* **1954**, *7*, 225.

(42) Flaig, R.; Koritsanszky, T.; Zobel, D.; Luger, P. *J. Am. Chem. Soc.* **1998**, *120*, 2227.

(43) Kerr, K. A.; Ashmore, J. P. *Acta Crystallogr., Sect. B* **1973**, *29*, 2124.

(44) Lehmann, M. S.; Nunes, A. C. *Acta Crystallogr., Sect. B* **1980**, *36*, 1621.

(45) Wagner, A.; Luger, P. *J. Mol. Struct.* **2001**, *595*, 39.

(46) Destro, R.; Roversi, P.; Barzaghi, M.; Marsh, R. E. *J. Phys. Chem. A* **2000**, *104*, 1047.

(47) Coppens, P.; Abramov, Y.; Carducci, M.; Korjov, B.; Novozhilova, I.; Alhambra, C.; Pressprich, M. R. *J. Am. Chem. Soc.* **1999**, *121*, 2585.

(48) Gorbitz, C. H. *New J. Chem.* **2003**, *27*, 1789.

(49) Coll, M.; Solans, X.; Fontaltaba, M.; Subirana, J. A. *Acta Crystallogr., Sect. C* **1986**, *42*, 599.

(50) Flaig, R.; Koritsanszky, T.; Dittrich, B.; Wagner, A.; Luger, P. *J. Am. Chem. Soc.* **2002**, *124*, 3407.

(51) Chen, C.; Parthasarathy, R. *Acta Crystallogr., Sect. B* **1977**, *33*, 3332.

(52) Alkaraghouli, A. R.; Koetzle, T. F. *Acta Crystallogr., Sect. B* **1975**, *31*, 2461.

(53) Koritsanszky, T.; Flaig, R.; Zobel, D.; Krane, H. G.; Morgenroth, W.; Luger, P. *Science* **1998**, *279*, 356.

(54) Frey, M. N.; Lehmann, M. S.; Koetzle, T. F.; Hamilton, W. C. *Acta Crystallogr., Sect. B* **1973**, *29*, 876.

(55) Yadava, V. S.; Padmanab, V. M. *Acta Crystallogr., Sect. B* **1973**, *29*, 854.

(56) Takigawa, T.; Ashida, T.; Sasada, Y.; Kakudo, M. *Bull. Chem. Soc. Jpn.* **1966**, *39*, 2369.

(57) Frey, M. N.; Koetzle, T. F.; Lehmann, M. S.; Hamilton, W. C. *J. Chem. Phys.* **1973**, *58*, 2547.

(58) Dalhus, B.; Gorbitz, C. H. *Acta Chem. Scand.* **1996**, *50*, 544.

JP902213T