JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Application of Local Second-Order Møller−Plesset Perturbation Theory to the Study of Structures in Solution

Johannes M. Dieterich,[†] João C. A. Oliveira,[†] and Ricardo A. Mata*[,†]

[†]Institut für Physikalische Chemie, Universität Göttingen, Tammannstrasse 6, D-37077 Göttingen, Germany

Supporting Information

ABSTRACT: In this work, we discuss the use of local second order Møller-Plesset perturbation theory (LMP2) in combination with the COSMO continuum solvation model for obtaining optimized geometries of molecules in solution. Density-fitting approximations, which reduce the computational cost relative to the basis set size, are also applied. We present results for small molecular systems, which show the same pattern observed in gas phase calculations. LMP2 results are found to be in excellent agreement with the canonical method. The only difference noticed is a slight increase in the average bond lengths, which is linked to the implicit reduction of basis set superposition effects (BSSE). Applications in the geometry optimization of an arginine model interacting with anions in solution as well as to the conformers of oligo-β-peptides are discussed.

1. INTRODUCTION

Local correlation approaches, as pioneered by Pulay and Saebø,[1−7] have become over the last years an established class of methods for the study of ground state properties, with extensions to excited states already developed.[8−10] Not only do they accomplish drastic savings in the computational cost, they also reduce the basis set superposition error by construction. Such properties make them state-of-the-art approaches to the study of weakly interacting systems. Although it would be of great interest to use such methods in the treatment of condensed systems, applications have been mostly restricted to gas phase molecules. QM/MM studies have been carried out,[11−14] but the cost of sampling an explicit solvent is still a strong limit to overcome. Continuum solvation models (CSMs)[15] offer an alternative, in that the solvent is represented by a dielectric continuum. This reduces the number of degrees of freedom to that of the original quantum molecular system in vacuo, in opposition to the case when one explictly describes the solvent molecules.

Among the various CSMs known to date, the 'COnductor-like Screening MOdel' (COSMO)[16] has been one of the most popular approaches. It holds a low computational cost and has proven to hold a good accuracy in the treatment of polar solvents. Alternative formulations, such as the polarizable continuum model (PCM),[17,18] may hold more developed approximations for the external potential or even include empirical corrections for dispersion.[19] These may prove important when dealing with apolar solvents, where the latter forces have a stronger weight in the overall interaction. However, the agreement between different CSMs is fairly good, and for most applications the computed results will not depend strongly on the choice made.

In this work, we explore the use of local second-order Møller−Plesset perturbation theory (LMP2)[20] in combination with the COSMO solvation model. In particular, we discuss an implementation of COSMO-LMP2 analytical gradients and its use in the optimization of molecular structures in solution. This combination makes for an efficient approach to the study of relatively large systems.

2. METHOD

We start by shortly reviewing the COSMO method, in particular with its use in Hartree−Fock (HF) calculations. The defining difference between COSMO and other CSMs is the use of a scaled conductor as an approximation to the dielectric continuum. A set of interlocking spheres, tagged to each individual center, is used do describe the solvation cavity around the solute. The size of each sphere is set according to the van der Waals radii of the atoms present (or optimized radii according to the model used). The polarization of the continuum, induced by the solute, is represented by the so-called screening charge density. A discretization of the surface leads to a set of surface segments, screening charges, and solute potentials which can be written in vector notation. If one considers Φ^X to be the vector containing the electrostatic potential generated by the solute X at each segment, the conductor boundary equation of vanishing total potential is given by

$$0 = \Phi = \Phi^X + \mathbf{A}\mathbf{q}^* \tag{1}$$

with $\mathbf{A}$ the Coulomb interaction matrix of the screening charges, and $\mathbf{q}^*$ the screening charges for a true conductor ($\varepsilon = \infty$). The screening charges have to be scaled in order to describe a finite dieletric. The relation between the solute electrostatic potential and the dielectric screening charges $\mathbf{q}$ is given by

$$\mathbf{A}\mathbf{q} = -\frac{\varepsilon - 1}{\varepsilon + x}\Phi^X = -f(\boldsymbol{\varepsilon})\Phi^X \tag{2}$$

The scaling factor $f(\varepsilon)$ describes the dependence on the permittivity, with x being set to 0.5 as a default. The solute electrostatic potential is defined by the charge density of the

Received: January 6, 2012

   dx.doi.org/10.1021/ct300009n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Selected Geometry Parameters (Degrees of Freedom - DOFs) for Different Optimized Structures at the COSMO-MP2 and COSMO-LMP2 Levels[a]

		toluene		water				toluene		water	
	DOF	MP2	LMP2	MP2	LMP2		DOF	MP2	LMP2	MP2	LMP2
water	r(O−H)	0.960	0.960	0.961	0.962	cytosine	r(C−N)	1.371	1.374	1.362	1.364
	α(H−O−H)	103.5	103.4	103.1	103.0		r(C−O)	1.223	1.223	1.238	1.238
							r(N−H)	1.010	1.010	1.009	1.010
ammonia	r(N−H)	1.012	1.013	1.013	1.014		α(N−C−N)	117.2	117.2	117.6	117.6
	α(H−N−H)	105.8	105.7	105.2	105.2		α(N−C−O)	124.6	124.5	123.3	123.3
							α(C−N−C)	119.9	119.9	119.7	119.7
furan	r(O−H)	1.360	1.361	1.362	1.364		θ(H−N−C-N)	12.5	12.6	3.5	4.9
	r(C−C)	1.365	1.365	1.365	1.366						
	r(C−C)	1.426	1.428	1.428	1.429	acetyl alcohol	r(C−O)	1.402	1.406	1.409	1.413
	α(C−O−C)	106.8	106.8	106.9	106.9		r(O−H)	0.970	0.970	0.971	0.970
	α(O−C−H)	115.9	115.8	115.9	115.9		r(C−O)	1.221	1.221	1.224	1.224
	α(O−C−C)	110.5	110.5	110.4	110.4		r(H···O)	2.000	2.029	1.993	2.024
							α(O−C−C)	119.2	119.5	119.0	119.3
ethanol	r(O−H)	0.961	0.962	0.963	0.963		α(C−C−O)	111.4	111.7	111.2	111.5
	r(O−C)	1.426	1.428	1.430	1.433		α(H−O−C)	104.2	104.6	104.2	104.7
	r(C−C)	1.510	1.514	1.510	1.514		α(O···H−O)	120.6	119.9	120.2	119.8
	α(H−O−C)	107.7	107.5	107.3	107.1		θ(H−O−C-C)	3.2	3.2	3.2	3.4
	α(O−C−H)	110.7	110.6	110.3	110.3						
	α(O−C−C)	107.5	107.6	107.8	107.9	glycine	r(C−O)	1.210	1.210	1.215	1.216
							r(C−O)	1.350	1.352	1.340	1.342
diethyl ether	r(O−C)	1.416	1.420	1.421	1.424		r(O−H)	0.970	0.970	0.972	0.972
	r(C−H)	1.096	1.097	1.095	1.096		r(N−H)	1.012	1.013	1.013	1.014
	r(C−C)	1.510	1.514	1.510	1.514		r(H···O)	2.279	2.281	2.291	2.292
	α(C−O−C)	111.5	111.5	111.3	111.3		α(O−C−O)	123.0	123.0	123.1	123.1
	α(O−C−C)	108.3	108.4	108.4	108.5		α(C−O−H)	106.1	106.0	107.0	107.0
	α(O−C−H)	109.8	109.8	109.6	109.6		α(O···H−O)	76.0	76.1	75.5	75.2
							θ(C−C−N-H)	62.1	63.5	66.4	67.7

[a]The basis set used was cc-pVTZ throughout. The bond distances are given in Å, angles in degrees. Two solvents were considered, toluene (ε = 1.89) and water (ε = 80.4).

underlying quantum mechanical calculation. In the Hartree−Fock case, a self-consistent treatment of the solvent polarization can be carried out by adding to the gas phase Fock operator $\mathbf{F}^{g,n}$ at each SCF cycle n, the external potential $\mathbf{Q}$ of the screening charges determined by the charge density $\mathbf{P}^{(n-1)}$ of the $(n-1)$th cycle

$$\mathbf{F}^{(s,n)} = \mathbf{F}^{(g,n)} + \mathbf{Q}(\mathbf{P}^{(n-1)}) \tag{3}$$

The Fock operator used in the SCF cycles is $\mathbf{F}^{(s,n)}$, which will determine the Hartree−Fock solution. The s in the upper index stands for solution, and g stands for gas phase.

In the case of HF, the conditions are well-defined. However, under a perturbative treatment, as in the case of MP2, the question of how the CSM should be invoked is dubious. There are three main approaches to the problem, which are general to CSMs.[21−28] The first approach would be to perform the calculation with the converged COSMO-HF Fock operator, containing the HF boundary potential. This is usually termed as the noniterative energy only scheme (PTE). In this case, one performs the COSMO-HF calculation, the MP2 calculation being performed on top of the canonical orbitals and Fock operator thus obtained. The screening charges are strictly defined on the basis of the HF potential. Another alternative would be to run the calculation, recompute the screening charges on the basis of the MP2 density, generating a new effective potential $\mathbf{Q}(\mathbf{P}^{(2)})$, which would then be back-fed into a new HF calculation. This would then be followed by successive MP2 and HF cycles, until convergence is reached. This approach is termed as iterative energy and density scheme (PTED). In the last approach, a nonconverged MP2 density is used. This can be obtained by a single cycle of PTED or using the vacuum MP2 density. The potential obtained is kept frozen and used in a new HF and MP2 run. This latter alternative is termed as PTD, which can be viewed as a truncated PTED scheme. As pointed out by Ángyán,[21,29] the only scheme which provides an energy correction consistent with the $2n+1$ rule is the simplest PTE scheme. Including MP2 correlation effects in the screening charges mixes correlation corrections above second order. We will therefore focus on the PTE model, and the remaining of this work will reference exclusively this approach.

In the PTE framework, due to the uncoupling of the correlation calculation and the COSMO perturbation, the extra gradient terms needed are rather simple, and their derivation analogous to the PCM gradient.[30] The interested reader should refer to the original work, where the derivation of the working equations for the PTE model are given. The difference lies in the boundary condition, so that one only needs to replace the PCM solvent operator for the COSMO equivalent (in eq (37) of ref 30). For example, variations in the Coulomb interaction $J^{(2)}$ (between the MP2 density and the HF converged CSM field) with respect to nuclear movement (R_α) will be given by

$$J^{(2)}_{R_\alpha} = f(\varepsilon)\left[\frac{1}{2}\mathbf{Q}(\mathbf{P}^{\mathrm{HF}}) + \mathbf{Q}(\mathbf{P}^{(2)})\right]\nabla_{R_\alpha}\mathbf{A}\mathbf{Q}(\mathbf{P}^{\mathrm{HF}}) \tag{4}$$

dx.doi.org/10.1021/ct300009n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

In order to calculate the COSMO-LMP2 gradient, the boundary potential with the LMP2 density also has to be computed, as evidenced in eq 4. For this, one needs to compute the second-order correction to the density matrix, which is obtained by solving a set of coupled-perturbed Hartree−Fock equations (CPHF). In the case of LMP2, a further set of coupled-perturbed localization (CPL) equations are needed,[31] where the effect of nuclear displacements on the localized molecular orbital coefficients is included. Both the CPHF as well as the CPL cycles are necessary in a LMP2 gradient calculation for isolated molecules (not solvated), so that the only change needed is to perform updates to the boundary potential during the CPHF cycles. The CPL equations are kept unchanged, since the potential is defined by the HF density.

The structure optimizations in this work have been carried out with the cc-pVDZ and cc-pVTZ Dunning basis sets.[32] Density-fitting calculations have been used throughout,[20,33] so we will drop the 'DF-' prefix in the following text. The density fitting basis used were the JKFIT and MP2FIT sets for the corresponding basis.[34] The LMP2 calculations were carried out using Pipek-Mezey localized orbitals[35] and the NPA criteria for orbital domains,[36] with $T_{NPA} = 0.03$. One should note that for every molecule considered in this study, independent of the value of ε used, the domains were found to be constant. Also in the case of the β-peptide structures, where different conformers have been computed, the applied criteria led to the same domain definition. A further point which could be raised is about the smoothness of the potential when using local correlation methods. During an optimization, as the molecular geometry is updated, the domain definition can change. This could give rise to convergence issues. However, as discussed in previous publications,[31,37] the standard procedure is to freeze the domains once the optimization is close to convergence, similar to the case of density functional theory (DFT) calculations, where the integration grid should be kept unchanged in the last optimization steps. LMP2 and MP2 calculations have been carried out with a development version of Molpro2010.2.[38] The DFT calculations were performed with Orca.[39]

3. RESULTS AND DISCUSSION

3.1. Benchmark Results. In the first series of benchmark calculations, we have optimized a number of small molecules under different environments with both the MP2 and LMP2 methods. The molecules included are as follows: water, ammonia, furan, ethanol, diethyl ether, cytosine, acetyl alcohol, and glycine. In order to show a large spread in the dielectric values, we considered the solvents toluene and water. One should also take into account that the value of toluene is close to some of the values usually applied when approximating the effect of an enzyme pocket (with $\varepsilon = 3-6$). Selected geometry parameters are shown in Table 1 for optimizations with the COSMO-MP2 and COSMO-LMP2 methods, using the cc-pVTZ basis set.

As it is visible in Table 1, the differences between MP2 and LMP2 are rather small. The trend is, as in the gas phase,[31] that the LMP2 bonds are slightly larger. This is most likely due to reduced basis set superposition effects, a characteristic from local methods by construction.[40,41] However, these differences are usually around 0.001−0.004 Å. The angles also compare rather well between both methods. The effect of the solvent variation is very small, of the same order as the differences between MP2 and LMP2. One should note that the trends are kept, irrespective of the method.

As an exemplary case, we wish to discuss the acetyl alcohol in detail. The different solvents have the biggest influence on the internal hydrogen bond length r(H···O) and angle α(O···H−O). In toluene solution, the internal hydrogen bond length is 0.029 Å shorter with MP2 than with LMP2. As pointed out before, this should be linked to the reduced BSSE of local methods in comparison to canonical ones. Changing the solvent from toluene to water decreases the hydrogen bond length in the canonical case by 0.007 Å as compared to 0.005 Å for the local one, showing a similar solvent effect. The same holds true for the angle of the hydrogen bond which is calculated to be 120.6° with MP2 and 119.9° with LMP2 in toluene. Changing to aqueous solvation reduces the angle by 0.4° for MP2 and 0.1° for LMP2, again preserving the overall tendency.

As a further test set, the structure of a few molecular dimers (water, ammonia, and formic acid) were also optimized. Intermolecular degrees of freedom (DOFs) are more strongly affected by the use of CSMs and/or local correlation approaches than internal parameters. Selected intermolecular DOFs are given in Table 2. The structures are shown in the Supporting Information. Just as in the case of monomers, MP2 and LMP2 agree extremely well, with slightly larger intermolecular distances in the latter structures.

Table 2. Selected Geometry Parameters for the Water, Ammonia, and Formic Acid Dimer Structures at the COSMO-MP2 and COSMO-LMP2 Levels[a]

		toluene		water	
	DOF	MP2	LMP2	MP2	LMP2
$(H_2O)_2$	r(O^a···H)	1.911	1.958	1.844	1.890
	α(O^a···H−O^d)	176.0	175.4	180.0	179.4
	θ(H−O^a O^d−H)	124.3	124.5	123.0	123.3
$(NH_3)_2$	r(N^a···H)	2.244	2.297	2.202	2.257
	α(N^a···H−N^d)	172.9	169.9	177.8	179.4
	θ(H−N^a N^d−H)	68.8	72.6	65.2	68.5
$(HCOOH)_2$	r(O^a···H)	1.653	1.656	1.656	1.717
	α(O−C═O)	126.2	125.6	125.6	125.6
	α(O^a···H−O^d)	178.9	177.7	177.6	178.2

[a]The basis set used was cc-pVTZ throughout. The bond distances are given in Å, angles in degrees. The 'd' and 'a' superscripts are used to distinguish between hydrogen-bond donor and acceptor, respectively. Two solvents were considered, toluene (ε = 1.89) and water (ε = 80.4).

The benchmark calculations are a mere indication of the correctness of the approach when comparing to the canonical case. It is somewhat tedious to examine the effect of a continuum solvent description on bond lengths, since these are already known to be rather small. The largest impact will be on the relative energies and structural parameters of interacting charged systems. To develop on this point, we take a look at the example of a protonated arginine interacting with halide anions in solution. As a second, more realistic application, we consider the conformation of a small β-peptide. These are known to adopt a large number of different secondary structures, more than their α counterparts, with variations even being observed upon changes in the solvent. These will be dominated by torsion potentials, which can be quite sensible to the level of theory. We also draw a comparison to DFT results.

dx.doi.org/10.1021/ct300009n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

3.2. Amino Acid Interactions. Some of the authors have recently presented results on the potential energy surface (PES) of interacting anion-amino acid systems.[42] In that work, only COSMO-LMP2 energies were used, with the amino acid structures being optimized at the COSMO-DFT level. The LMP2 energies were computed on the basis of a grid, placing the anion at discrete positions on the plane of the amino acid and computing the interaction energy. Although such an approach does bring some insight to the overall PES, one is usually interested in finding relevant minima. Plotting a full dimensional PES for this purpose is, of course, out of reach. It would also be of interest to compute the geometries and energies at an even footing. The interaction of heavier anions (such as Br^- and I^-) is strongly conditioned by dispersion forces and the overall surface can be rather featureless, so that different methods can give large variations in the structure optimization. The above cited difficulties were in fact the starting motivation for coupling the LMP2 method to a CSM. On the other hand, the treatment of anion-including complexes with continuum solvation models is a difficult task, and some caution is advised. Due to the diffuse nature of the anions, the cavity size can strongly influence the results. Also, the smaller anions have very strong interactions which are most accurately described with explicit solvation. One possibility is to add a first solvation shell around the interaction site. However, since in this work we are mostly concerned with the comparison between canonical and local correlation methods, we found the use of standard criteria to be sufficient. The accuracy of the results relative to the CSM parameters are discussed in more detail in another publication.[42]

We will take this opportunity to demonstrate the application of COSMO-LMP2 and analyze specific interactions between anions and water with a protonated arginine. The guanidinium cation has deserved considerable attention in the past,[43,44] as it has been shown to serve as an effective motif for binding anions.[45] This moiety is present in the protonated arginine side chain, which can be used by proteins to dock anionic species. Up until now, quantum chemical calculations have mostly been performed on the gas phase. Computed interaction energies between guanidinium and chloride, for example, are in the range of 450 kJ/mol. However, when studying systems in solution, one is instead interested in the solvated complex formation energy. This is the energy difference between the complex and the two separated species, all in solution. The latter should be significantly smaller when studying an ion-pair, since the solvation energy for the complex will be much lower than that for the separated constituents. The interaction itself can also be shielded by the polar solvent.

We have considered a series of minima for an arginine model interacting with water as well as chloride and bromide anions. The N-terminus was acetylated and the C-terminus methylated, thereby removing the acidic and basic moieties in the backbone. In this way, we focus on the interactions of the guanidinium group. Afterward, we performed full geometry optimizations, placing either a chloride anion, bromide anion, or a water molecule at an identified minimum location. We considered three docking sites, which are displayed in Figure 1. Calculations were carried out both in the gas phase and with COSMO in order to include a water solvation effect (ε = 80.4). We measured the distance from the amino acid to each interacting species as well as the angle formed according to the scheme shown in Figure 1. The contact acidic proton is taken as a reference. The results are given in Table 3.

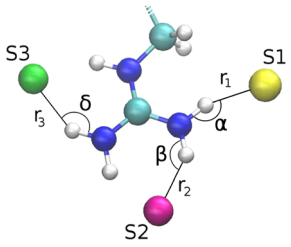

Figure 1. Schematic representation of the three interaction sites considered for building the protonated arginine minima. At sites S2 and S3, the anion/water species interacts with two protons.

Table 3. Geometric Parameters (Distances in Å, Angles in Degrees) for the Interacting Anion/Molecule and the Acidic Proton in the Optimized Arginine Systems[a]

		water			gas phase		
		distance	angle	ΔE	distance	angle	ΔE
S1	Cl^-	2.132	166.4	−26.3	−−	−−	−−
	Br^-	2.382	163.4	−19.8	1.964	171.1	−430.1
	H_2O	1.864	168.3	−11.6	1.844	171.3	−51.9
S2	Cl^-	2.346	153.0	−35.9	1.917	156.4	−476.0
	Br^-	2.496	156.0	−28.5	2.130	156.5	−443.6
	H_2O	1.971	149.6	−21.9	2.003	149.1	−65.7
S3	Cl^-	2.175	160.0	−34.2	1.865	162.0	−483.5
	Br^-	2.370	162.0	−30.7	2.075	164.2	−481.7
	H_2O	1.938	151.8	−20.9	1.973	149.9	−100.4

[a]The calculations in the gas phase were carried out with LMP2/cc-pVDZ, calculations in water with COSMO-LMP2/cc-pVDZ (ε = 80.4). The energy difference (in kJ/mol) between the complex and the isolated species has been computed with the aug-cc-pVTZ basis set.

A first comment has to be made on why the S1 results for Cl^- in the gas phase are not given. In this case, the chloride anion actually deprotonates the gaunidinium moiety, contrary to the COSMO optimization. The two structures are shown in Figure 2. This is naturally linked to the role of the solvent in

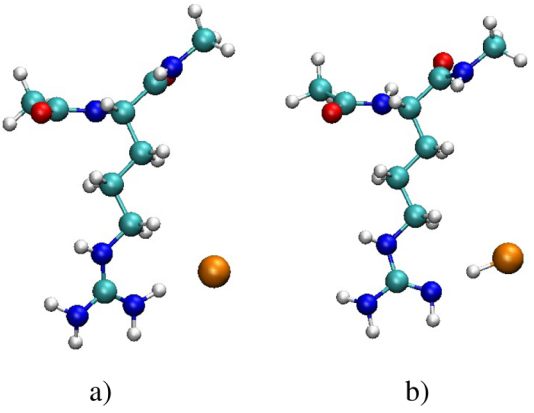

Figure 2. Optimized structures of the protonated arginine and chloride system (S1): a) in water (ε = 80.4) and b) gas phase.

stabilizing charge separation. This result clearly shows the impact of solvent effects in the structure, as a study using a gas phase optimization would be impossible. The effect is also visible in the case of bromide, although much weaker. The Br^- distance to the proton is much shorter in the gas phase (a difference of about 0.4 Å). The proton is still connected to the arginine, but the gas phase shows a much stronger attraction,

dx.doi.org/10.1021/ct300009n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

which is in line with the discussion above. The same trend is observed for the other interaction sites, both with Cl^- and Br^-. The distances in the gas phase are significantly shorter (0.3−0.5 Å). At S2 and S3 we did not observe deprotonation of the amino acid.

Comparing the distances obtained in solution to the minima found in our previous work,[42] where we used a frozen amino acid structure and a point grid for the anion, we find a good agreement. For example, at the S1 site, the Cl^- is 2.132 Å away from the proton (compared to 2.192 Å), and the Br^- 2.382 Å (compared to 2.392 Å). One should take into account that the basis sets were different, the amino acid structure is relaxed, and that the anion is not constrained to a plane. Differences in such a range are perfectly acceptable. Finally, we also have results for a single water molecule interacting with the arginine. Since the molecule is not charged, the solvent effect is much smaller, leading to a change of only 0.02−0.04 Å for the proton to water distances. Nevertheless, we see again the obvious trend for gas phase optimizations to overestimate the interaction.

In Table 3 we have also included the energy difference between the complex and the isolated species (amino acid and ion/water), either in solution or in the gas phase. These values were computed through single point calculations at the LMP2/aug-cc-pVTZ level of theory. The internal geometries for the amino acid and water molecules were kept from the complex optimization. As previously discussed, the two values are quite different, since the value in solution also takes into account the difference in solvation energy for each of the species. The stabilization energy in water solution follows some obvious trends. First of all, the interactions at site S1 are somewhat weaker, which is connected to the fact that only an interacting proton is present in this case. The difference to the other sites is rather constant, around 10 kJ/mol. Comparing the different interacting species, one observes, as expected, that the complexes with chloride are the most stable. Most importantly, both chloride and bromide have larger stabilization energies than a water molecule, which confirms the ability for arginine to bind anions (both species should be able to compete with surrounding water molecules for docking). The difference is between 7 and 15 kJ/mol, depending on the anion and the docking site. This is, of course, just a rough estimate since other effects would come into play. One of them would be the relaxation of the amino acid structure, but this would require extensive sampling which is out of the scope of this work. The gas phase values are comparable to previous calculations using only the guanidinium cation. Rozas et al.[44] reported an interaction energy between −437.7 and −449.9 kJ/mol at site S2 for the chloride anion. We computed an interaction of −476.0 kJ/mol, somewhat closer to their CBS-QB3[46] value of −473.8 kJ/mol. The anion interaction energies show the largest difference to the values in solution. This is simply connected to the fact that a neutral complex is formed, with a low solvation energy compared to what one obtains for the separate ions. In the case of water, the complex formed is positively charged. Sites S2 and S3 also seem to be favored, but other trends are difficult to observe, one of the reasons being that the S1 Cl^- result is not available. One rather surprising result is the difference in energy between chloride and bromide at site S3. They are practically identical, when one would expect chloride to have a stronger interaction. This could be connected to poor sampling in optimizing the complexes or to shortcomings in the use of CSM for this particular application. As noted at the beginning of this Section, a more accurate description of the system could require the use of explicit solvent molecules.

We have also reoptimized the structures in solution with canonical MP2. A comparison is made between the MP2 and LMP2 distances in Figure 3. The solid line represents the case

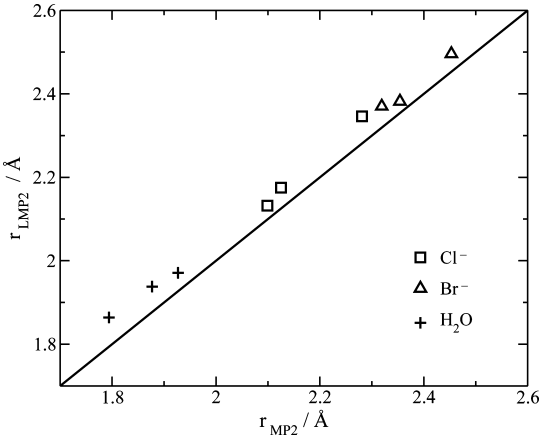

Figure 3. Shortest H−X (X = Cl^-, Br^-, H_2O) distances (in Å) for optimized structures using COSMO-MP2/cc-pVDZ and COSMO-LMP2/cc-pVDZ.

where both results would strictly agree. As expected, the local correlation method optimized structures show slightly longer H−X (X = Cl^-, Br^-, H_2O) distances. The deviations are in the range of 0.03−0.07 Å, which is reasonable taking into account that a double-ζ basis was used. Basis set effects can have a strong influence for such intermolecular distances. However, the trends are in general well reproduced.

3.3. β-Peptide Conformers. As a further test system, we take a look at the conformation of oligo-β-peptides.[47] These are known to adopt a wide variety of different structures, mostly depending on the constitution of the backbone. There are several works focused on rationalizing the observed trends, with the use of quantum chemical methods playing an important role. These provide information at the atomistic level and identify specific interactions which in turn help rationalize the observed trends (see ref 48. and references contained therein). However, an overwhelming majority of the first principle calculations done in the field to date has been on the basis of DFT. The latter class of methods lacks the description of dispersion forces. Recently, empirical corrections have been suggested to overcome this problem.[49] Nonetheless, these energy terms are defined independently of the surrounding environment so that the gas phase dispersion corrections will be the same as the ones found in solution. In order to obtain a consistent treatment of interactions within a CSM calculation, a dependence on the external potential should be included.

We have chosen as a test system tetrapeptide models of the *cis*-2-aminocyclopentane carboxylic acid (*cis*-ACPC). The oligopeptides have been capped with an azide group at the N-terminus and a methyl ester at the C-terminus. The conformational properties of these oligopeptides have been thoroughly studied, through both theoretical as well as experimental methods.[50−53] We will use them in this context as a mean to test the applicability of the presented model for larger applications and obtain some information on the timings involved when comparing to gas phase calculations. The total system size for the model molecule used is 74 atoms. Starting

dx.doi.org/10.1021/ct300009n | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

from hand-generated structures, we optimized two conformers which have been readily identified, a sheet and a 14-helix. The sheet conformer is close to an unfolded structure, with very little strain on the backbone. The 14-helix has a repeating coil pattern, characterized by a hydrogen bond between an amide proton at residue (i) and a backbone carbonyl at residue (i+2). The smallest possible model for such a conformer would be the tetrapeptide, the reason for our choice. Since methanol is a commonly used solvent in experimental work with these structures, we also chose to use it (ε = 32.63). The geometries are shown in Figure 4 and have been optimized at the COSMO-LMP2/cc-pVDZ level of theory.

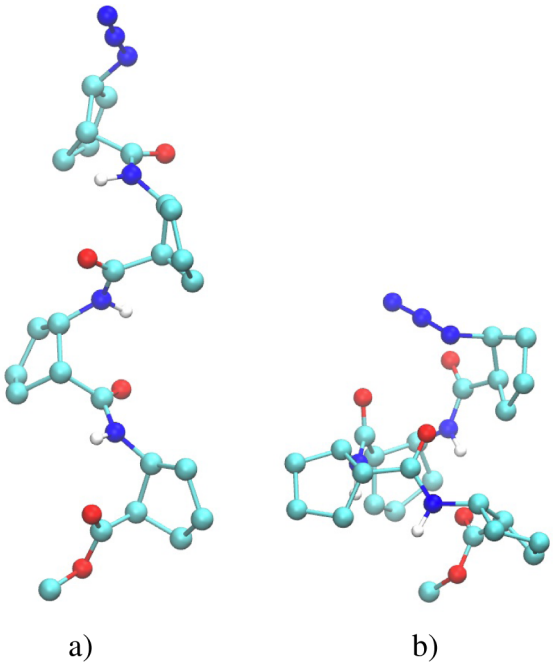

Figure 4. Optimized structures of the *cis*-ACPC tetrapetide conformers with COSMO-LMP2/cc-pVDZ (ε = 32.63): a) β-sheet and b) 14-helix. The allylic hydrogens were removed for clarity.

The structure found for the folded conformation does not correspond to a true 14-helix. From the two turns found in the tetrapepide only one is a 14 atom loop contact. The other shows a 10 atom pattern. The fact that we did not obtain a pure 14-helix is related to the small system size considered where the end groups significantly influence the overall conformation, allowing for a more compact fold. The sheet structure shows a regular pattern in agreement with the structures found in larger peptide chains.[51] In order to highlight the possible benefits of local correlation methods for the study of such structures, density functional calculations have also been carried out in the conformers. In a series of calculations, BP86[54,55] and B3LYP[56] single points with the def2-TZVP basis set[57] have been carried out on the LMP2 optimized structures. In order to evaluate the effect on the structures, the geometries were reoptimized at the BP86/def2-SVP level of theory. Another set of calculations was carried out, including dispersion corrections (DFT-D3) as proposed by Grimme.[49]

The energy differences between the sheet and the helix structures are displayed in Table 4. For the system type chosen (peptides with no aromatic side chains), LMP2/aug-cc-pVTZ should deliver close to converged results for the conformational energy. The accuracy of the method was recently evaluated for

Table 4. Energy Difference (ΔE = E(sheet)−E(helix)) for the Two Conformers of the *cis*-ACPC Tetrapeptide (in kJ/mol), Computed at Different Levels of Theory for Optimization and Single Point Calculations

structure	energy	ΔE
LMP2/cc-pVDZ	LMP2/aug-cc-pVTZ	24.6
	BP86/def2-TZVP	−41.4
	B3LYP/def2-TZVP	−37.6
	BP86-D3/def2-TZVP	12.9
	B3LYP-D3/def2-TZVP	9.2
BP86/def2-SVP	BP86/def2-TZVP	−36.4
BP86-D3/def2-SVP	BP86-D3/def2-TZVP	60.3

a large test set of α-peptides, which included tri- to octapeptides.[58] For the two geometries considered, LMP2 favors the helix conformer by 24.6 kJ/mol. This is in agreement with the DFT-D3 results. Using the same geometry, BP86-D3 and B3LYP-D3 give a 12.9 and 9.2 kJ/mol difference, respectively. The dispersion correction, in fact, is a determining factor. The pure DFT values strongly favor the sheet conformation. Since the helix is a much more compact structure, this effect is expected, as dispersion forces will play a much more dominant role in the latter.

The structures were also reoptimized using BP86 and BP86-D3 with the def2-SVP basis set. These results show that the use of DFT-D3 leads to a strong overestimation of the helix stability. Comparing the BP86/def2-TZVP values using either the LMP2 or the BP86 optimized conformers, one finds only a slight difference (about 5 kJ/mol). The structures also remain rather similar. The BP86-D3 geometry is slightly more contracted, and this is reflected in the 56 kJ/mol change, when comparing to the result obtained using the LMP2 geometries. The formation of a 14-fold is strongly linked to the hydrogen bond between the carbonyl oxygen of the first residue and the amide proton of the third. The O···H−N distances (between the proton and the oxygen) for BP86 and LMP2 are 2.289 and 2.305 Å, respectively. They are found to be in good agreement. The helix structure obtained with BP86-D3, however, reveals a hydrogen bond distance of 2.045 Å, much shorter than the LMP2 result. The reason behind these discrepancies is unclear but would clearly warrant further investigations. There is little information on the use of DFT-D3 to explore the energetics of β-peptides as well on how the results with CSMs may compare to higher-level methods.

The COSMO-LMP2 optimizations were carried out with the cc-pVDZ basis set, a total of 694 basis functions. The timings for a full HF calculation plus LMP2 and gradient calculation (with density fitting approximations used throughout) amount to 4.3 h (the LMP2 part plus gradient takes 2.5 h). These are wall-clock timings for a single-processor run on an Intel X5550 2.67 Ghz dual quad-core machine. The use of COSMO leads to only small changes in the timings. The HF calculation is the section where the greatest impact is observed, with 0.4 h added time. One should note that this is due to the inclusion a COSMO step in each SCF cycle. Since this increases with the number of cycles, and the timed calculations were performed with standard initial guesses, this is an upper estimate. In a regular optimization, the wave function from the previous step is reused, reducing the number of cycles. The additional LMP2 related gradient terms take 7.2 min to compute (matrix multiplications involving the second-order correction to the density). Memory requirements are not affected by the use of

COSMO since the extra calculations are performed outside the HF and LMP2 cycles, which establish the bottlenecks in memory usage. The results show that such an approach could be effectively applied to small oligopeptide chains, investigating at a high level of theory their conformational properties.

4. CONCLUSIONS

We have presented LMP2 calculations in combination with the COSMO continuum solvation model, making use of analytical energy gradients. The implementation is based on the noniterative energy only scheme. Comparison has been made to COSMO-MP2 results, showing a very small effect through the use of local approximations when examining internal degrees of freedom. We have also examined the structures of a model for a protonated arginine, interacting with chloride, bromide, and water. A strong effect is observed, as expected, when comparing gas phase optimized structures to those with a COSMO description for solution. In the case of chloride, a proton transfer in the gas phase calculation is observed which would otherwise not occur. These observations highlight the importance of including solvent effects (even if through an implicit solvation model) when studying such interactions.

The COSMO-LMP2 model presents itself as a cost-effective method for the study of molecular systems in solution. The reduced scaling of the computational cost together with the intrinsic neglect of basis set superposition effects in the correlated part of the calculation make it particularly suitable for application in systems with over 50 atoms. The use of local correlation methods to the study of β-peptides holds good promise, as one should be able to benchmark the conformational properties of small chains, just as previously done in the case of α-peptides. The former are particularly challenging, due to the large number of energetically close lying conformers which have to be analyzed.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Table 1 and Figure 1. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rmata@gwdg.de.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank the helpful comments from Dr. Michael Diedenhofen in the course of this work. The COSMO-LMP2 gradients code is based on a trial COSMO-MP2 implementation of Dr. Andreas Klamt and his co-workers in the Molpro program package. Financial support from the German Excellence Initiative, through the Free Floater Research Group program of the University of Göttingen, is also gratefully acknowledged.

■ REFERENCES

(1) Pulay, P. *Chem. Phys. Lett.* **1983**, *100*, 151−154.
(2) Saebø, S.; Pulay, P. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213−236.
(3) Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1996**, *104*, 6286−6297.
(4) Schütz, M.; Hetzer, G.; Werner, H.-J. *J. Chem. Phys.* **1999**, *111*, 5691−5705.
(5) Schütz, M.; Werner, H.-J. *Chem. Phys. Lett.* **2000**, *318*, 370−378.
(6) Schütz, M. *J. Chem. Phys.* **2000**, *113*, 9986−10001.
(7) Schütz, M.; Werner, H.-J. *J. Chem. Phys.* **2001**, *114*, 661−681.
(8) Korona, T.; Werner, H.-J. *J. Chem. Phys.* **2003**, *118*, 3006−3019.
(9) Kats, D.; Korona, T.; Schütz, M. *J. Chem. Phys.* **2006**, *125*, 104106.
(10) Kats, D.; Korona, T.; Schütz, M. *J. Chem. Phys.* **2007**, *127*, 064107.
(11) Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J. *Angew. Chem., Int. Ed.* **2006**, *45*, 6856−6859.
(12) Mata, R. A.; Werner, H.-J.; Thiel, S.; Thiel, W. *J. Chem. Phys.* **2008**, *128*, 025104.
(13) Mata, R. A. *Phys. Chem. Chem. Phys.* **2010**, *12*, 5041−5052.
(14) Dieterich, J. M.; Werner, H.-J.; Mata, R. A.; Metz, S.; Thiel, W. *J. Chem. Phys.* **2010**, *132*, 035101.
(15) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.
(16) Klamt, A.; Schüürman, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.
(17) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117−129.
(18) Cammi, R.; Tomasi, J. *J. Comput. Chem.* **1995**, *16*, 1449−1458.
(19) Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616−627.
(20) Werner, H.-J.; Manby, F. R.; Knowles, P. *J. Chem. Phys.* **2003**, *118*, 8149−8160.
(21) Ángyán, J. G. *Chem. Phys. Lett.* **1995**, *241*, 51−56.
(22) del Valle, F. J. O.; Tomasi, J. *Chem. Phys.* **1991**, *150*, 139−150.
(23) Aguilar, M. A.; del Valle, F. J. O.; Tomasi, J. *Chem. Phys.* **1991**, *150*, 151−161.
(24) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−2200.
(25) Christiansen, O.; Mikkelsen, K. V. *J. Chem. Phys.* **1999**, *110*, 1365−1375.
(26) Lipparini, F.; Scalmani, G.; Mennucci, B. *Phys. Chem. Chem. Phys.* **2009**, *11*, 11617−11623.
(27) Cammi, R. *J. Chem. Phys.* **2009**, *131*, 164104.
(28) Caricato, M.; Scalmani, G.; Trucks, G. W.; Frisch, M. J. *J. Phys. Chem. Lett.* **2010**, *1*, 2369−2373.
(29) Ángyán, J. G. *Int. J. Quantum Chem.* **1993**, *47*, 469−483.
(30) Cammi, R.; Mennucci, B.; Tomasi, J. *J. Phys. Chem. A* **1999**, *103*, 9100−9108.
(31) ElAzhary, A.; Rauhut, G.; Pulay, P.; Werner, H.-J. *J. Chem. Phys.* **1998**, *108*, 5185−5193.
(32) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(33) Polly, R.; Werner, H.-J.; Manby, F. R.; Knowles, P. J. *Mol. Phys.* **2004**, *102*, 2311−2321.
(34) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116*, 3175−3183.
(35) Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*, 4916−4926.
(36) Mata, R. A.; Werner, H.-J. *Mol. Phys.* **2007**, *105*, 2753−2761.
(37) Hrenar, T.; Rauhut, G.; Werner, H.-J. *J. Phys. Chem. A* **2006**, *110*, 2060−2064.
(38) Werner, H.-J.; Knowles, P. J.; Knizia, R.; Manby, F. R.; Schütz, M. and others, *MOLPRO, version 2010.2, a package of ab initio programs*, 2010. See http://www.molpro.net (accessed month day, year).
(39) *ORCA, an ab initio, DFT and semiempirical electronic structure package, version 2.8*, 2010.
(40) Schütz, M.; Rauhut, G.; Werner, H.-J. *J. Phys. Chem. A* **1998**, *102*, 5997−6003.
(41) Mata, R. A.; Werner, H.-J. *J. Chem. Phys.* **2006**, *125*, 184110.
(42) Oliveira, J. C. A.; Feldt, J.; Galamba, N.; Mata, R. A. *J. Phys. Chem. A* **2012**, *116*, 5464−5471.
(43) Rozas, I.; Kruger, P. E. *J. Chem. Theory Comput.* **2005**, *1*, 1055−1062.
(44) Rozas, I.; Alkorta, I.; Elguero, J. *Struct. Chem.* **2008**, *19*, 923−933.
(45) Best, M. D.; Tobey, S. L.; Anslyn, E. V. *Coord. Chem. Rev.* **2003**, *240*, 3−15.
(46) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532−6542.

(47) Martinek, T. A.; Fülöp, F. *Eur. J. Biochem.* **2003**, *270*, 3657−3666.
(48) Wu, Y.-D.; Han, W.; Wang, D.-P.; Gao, Y.; Zhao, Y.-L. *Acc. Chem. Res.* **2008**, *41*, 1418−1427.
(49) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
(50) Martinek, T. A.; Tóth, G. K.; Vass, E.; Hollósi, M.; Fülop, F. *Angew. Chem., Int. Ed.* **2002**, *41*, 1718−1721.
(51) Martinek, T. A.; Mádity, I. M.; Fülöp, L.; Tóth, G. K.; Vass, E.; Hollósi, M.; Forró, E.; Fülop, F. *J. Am. Chem. Soc.* **2006**, *128*, 13539−13544.
(52) Beke, T.; Somlai, C.; Perczel, A. *J. Comput. Chem.* **2006**, *27*, 20−38.
(53) Pandey, S. K.; Jogdand, G. F.; Oliveira, J. C. A.; Mata, R. A.; Rajamohanan, P. R.; Ramana, C. V. *Chem.—Eur. J.* **2011**, *17*, 12946−12954.
(54) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(55) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824.
(56) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(57) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(58) Kaminsky, J.; Mata, R. A.; Werner, H.-J.; Jensen, F. *Mol. Phys.* **2008**, *106*, 1899−1906.