JCTC Journal of Chemical Theory and Computation

Intramolecular Basis Set Superposition Error Effects on the Planarity of DNA and RNA Nucleobases

David Asturiol,[†] Miquel Duran,[‡,‡] and Pedro Salvador*[,‡,‡]

Institut de Química Computacional, Parc Científic i Tecnològic de la Universitat de Girona, Edifici Jaume Casademont, Pic de Peguera 15 (la Creueta), 17003 Girona, Spain, and Institut de Química Computacional and Departament de Química, Universitat de Girona, Campus de Montilivi, 17071 Girona, Spain

Received January 30, 2009

Abstract: Molecules of utmost importance like DNA and RNA nucleobases are predicted to be nonplanar by a typical ab initio method, such as second order Møller−Plesset perturbation theory (MP2) combined with standard Pople's basis sets. Similarly to the case of other planar aromatic systems, these pitfalls can be explained in terms of intramolecular basis set superposition error (BSSE) effects, induced by local basis set deficiencies. We demonstrate that conventional BSSE correction techniques such as the Counterpoise method can account for this wrong behavior and provide proper correction whenever spurious results occur, mainly in case of thymine, uracil and guanine but also to lower extent for adenine and cytosine. We also show that special care must be taken when assessing the BSSE by means of ghost-orbital calculations for strongly overlapping fragments. Often molecular orbitals in the isolated fragment calculation have a different orientation as in the ghost-orbital calculation. This can lead to bogus derivatives of the CP-correction term, essential to account for geometry and vibrational BSSE effects.

Introduction

In the last years a number of studies[1−10] have reported conventional ab initio calculations at the correlated level producing nonplanar minima for systems such as benzene, polycyclic aromatic hydrocarbons, and other nonrigid cyclic systems, such as the hexacoordinated carbon anion $(B_6C)^{2-}$. The planar stationary structures corresponded to saddle-points on the potential energy surface (PES) with one or more imaginary frequencies, associated to low-lying out-of-plane vibrational modes.

Until very recently the origin of such anomalies had not yet been clearly determined. Martin et al.[4] studied in detail the benzene molecule and suggested a basis set superposition error (BSSE) origin. Moran et al.[6] reported a wide number of correlated calculations combined with Pople's basis sets for aromatic systems that produced nonplanar minima. In these cases, their analysis revealed a strongly geometry-dependent two-electron basis set incompleteness error (BSIE) which increased for nonplanar geometries. However, no alternatives but the use of a different (more balanced) basis set or the careful extrapolation to the complete basis set limit were provided as a solution for these problematic cases.

Recently,[11] we have offered a solution to the problem based on the use of conventional BSSE correction techniques which have been successfully applied to correct for BSSE in ab initio descriptions of intermolecular complexes. We have confirmed that the origin of the reported pitfalls arise from local basis set deficiencies. This clearly indicates the fact that intra- and intermolecular BSSE have common origin and can be managed in a similar fashion. The same philosophy has been applied very recently by Balabin et al.[12] to obtain accurate energy differences for several conformations of normal alkanes.

The origin of the intermolecular BSSE in ab initio calculations is rooted on the use of truncated basis sets centered on the atomic positions. The interaction energy

* To whom correspondence should be addressed. E-mail: pedro.salvador@udg.edu.

[†] Institut de Química Computacional, Parc Científic i Tecnològic de la Universitat de Girona.

[‡] Institut de Química Computacional and Departament de Química, Universitat de Girona.

10.1021/ct900056u CCC: $40.75
Published on Web 07/31/2009

between two molecules is usually calculated subtracting the energy of the fragments from the energy of the complex. In such a calculation, the energy of the fragments is obtained using the basis set of each fragment, whereas the energy of the complex is calculated using the basis set of all fragments.

It is well-known that BSSE affects not only the interaction energy of the complex but also its potential energy surface. This translates into geometrical and vibrational effects associated with the change of the position of the stationary point and nearby curvature of the PES. Several examples of how BSSE can dramatically affect the geometry of intermolecular complexes can be found in the literature.[13−15]

For a single molecule, there is no a priori problem with the fact that atoms or groups of atoms make use of basis functions centered on other parts of the molecule, as it takes into account polarization and charge transfer effects. However, if the use of these external basis sets is the result of a lack of flexibility of the fragment's own basis set, these local basis set deficiencies may result into spurious stationary points and vibrational frequencies associated. One can refer to such phenomena as intramolecular BSSE.[16−20]

There are several strategies to correct for BSSE[21−24] but the Counterpoise method[21] (CP) is the most widely used due to its simplicity. The CP-correction to the energy for a system formally composed of N interacting fragments reads as

$$\delta^{CP}(\vec{R}) = \sum_{i}^{N} \varepsilon_i^i(\vec{R}) - \varepsilon_i^{\text{full}}(\vec{R}) \quad (1)$$

where $\varepsilon_i^i(\vec{R})$ and $\varepsilon_i^{\text{full}}(\vec{R})$ represent the energy of the i-th fragment of the system calculated with its own basis set and with the full basis set of the system (ghost-orbital calculation), respectively. Explicit dependence on the atomic positions has been included to stress that the CP correction is geometry-dependent. Also, it is worth to note that the electronic state of each fragment must be specified in terms of its charge and spin multiplicity. The CP-correction, applied as an additive correction term to the total energy,[25]

$$E^{CP}(\vec{R}) = E(\vec{R}) + \delta^{CP}(\vec{R}) \quad (2)$$

provides BSSE corrected energies, as well as any property that can be obtained from the total energy of the system or its derivatives, namely, stationary points, vibrational frequencies, dipole moments, etc.

In the case of intermolecular complexes, the definition of the fragments is usually trivial; each interacting molecule is considered as a fragment forming the so-called supermolecule. This permits to obtain BSSE-corrected interaction or stabilization energies with respect to the corresponding fragments. Some ambiguities in the specification of the fragments may arise in the specific case of charged intermolecular complexes or interactions involving open-shell species, namely, which fragment bears the charge and which multiplicity is to be specified. Nevertheless, such cases have already been considered and satisfactory results have been obtained.[19]

The situation is rather different in the case of intramolecular BSSE correction. One must face the ambiguity of subdividing the molecular unit into subunits from which a BSSE-correction will be determined. This may involve the rupture of chemical bonds, and therefore, the specification of the electronic state of each individual fragment becomes another source of arbitrariness. An alternative to avoid such unsaturated fragments is to estimate BSSE by modeling the system with a proper intermolecular complex with the same geometry.[26−28] Such approach seems to provide reasonable results for single-point calculations but it is certainly nontrivial how it could be applied for CP-corrected geometry optimizations and frequency calculations.

Another possible general approach for a molecule could be to use atomic fragments. However, the CP philosophy might be difficult to accomplish in this case (vide infra). Ideally, atoms ought to be promoted to the hybridization state they appear in the molecule when determining the basis set extension effects, and this is not a trivial task. The use of atomic fragments also has dramatic implications with the computational cost associated to the CP-correction, which is roughly $N + 1$ times ($N =$ the number of fragments) that of the uncorrected calculation. For any midsize molecule such computational scheme can easily become unfeasible.

It is worth noting that in the case of a more involved methodology that has proven to be very successful in correcting for BSSE in the intermolecular case, namely the Chemical Hamiltonian Approach (CHA), the fragments are essentially defined by their associated basis functions so that charge and multiplicity do not need to be specified for each fragment. In fact, the CHA method does not rely on any extra fragment calculations since the total energy obtained in the calculations is already free from BSSE. The reason for not using it in this context is that the method does not seem to behave properly for strongly overlapping fragments, which is the case of intramolecular BSSE (a variant[16] of the method at the Hartree−Fock level was developed some years ago with promising results for intramolecular BSSE correction, too). Nevertheless, what one learns from an a priori BSSE correction method such as the CHA is that the CP method is essentially an ingenious balance error technique that, as pointed out by Mayer,[29] assumes that the energy lowering induced by the use of external basis functions for a given fragment within the supermolecule is the same as the one obtained by considering the single fragment together with the whole set of basis functions (ghost orbitals). The more similar the electronic state of the isolated fragment is to the *local state* of that fragment within the molecule, the more appropriate the CP-correction will be.

CP-corrected calculations are nowadays carried out quite routinely without paying much attention to the electronic state of the fragment's calculations. A notably exception are Alexander and co-workers, who for a simple system like $B\cdots H_2$ brought up the use of diabatic states carefully chosen in order to obtain accurate CP-corrections for different electronic states.[30] Fortunately, numerical experience so far has shown that the CP correction is, in this aspect, quite robust. For instance, in the case of charged intermolecular complexes such as the protonated water dimer, quite reasonable results were obtained using either two or three fragments.[19] For a genuine open-shell complex such as HF···NO, we described[31] difficulties in the selection of proper ghost-

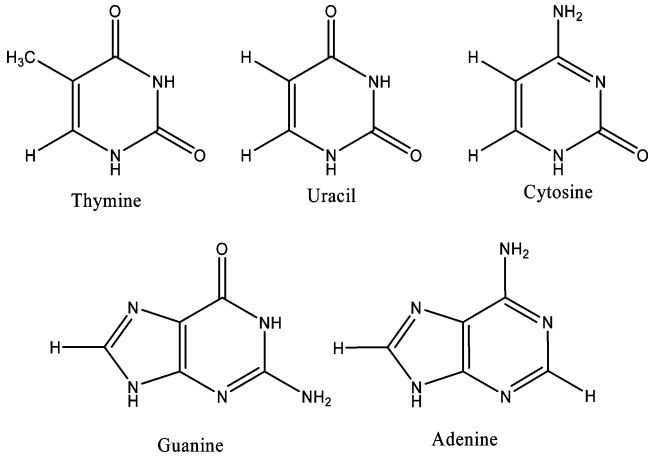

Figure 1. Nucleobases considered in this study.

orbital states for obtaining CP-correction for different electronic states, but the energetic differences between them were very small.

As mentioned above, in a recent work[11] we have tackled the correction of intramolecular BSSE effects in the case of benzene and several arenes, including charged systems such as cyclopentadienyl and indenyl anions at the MP2 and Configuration Interaction with Singlet and Doublet excitations (CISD) levels of theory combined with standard Pople's basis sets. In all cases, the problems were associated with out-of-plane bending low-lying modes for which one or more imaginary frequencies can be found (as large as 1181*i* at the MP2/6-311++G** level of theory). Simple inspection showed that the intramolecular BSSE did not affect bonding distances or angles (otherwise stretching and other bending modes would have been affected). Taking this in consideration, we showed that such specific intramolecular basis set deficiencies could be solved by taking as fragments diatomic C−H units constituting the arenes. To maintain the molecule's symmetry all C−H fragments must be equivalent. In the case of charged systems, careful choice of the CP correction combining charged and neutral C−H moieties had to be designed.

The aim of this work is to show that the anomalous behavior observed for benzene and other arenes seems to be in fact quite common for cyclic planar molecules with π-systems. Here we describe a series of spurious imaginary frequencies associated with out-of-plane bending modes for planar stationary points of adenine, cytosine, thymine, guanine, and uracil nucleobases (see Figure 1) obtained at the MP2 level of theory and Pople's standard basis sets. Luckily, what they seem to be the most widely used basis sets in the literature for these systems, namely 6-31G* and 6-311G*, do not present any spurious imaginary frequency in any case. However, if a better description is needed and diffuse functions are added to those bases, thymine, uracil and guanine optimized planar structures can present one or more imaginary frequencies.

These molecules, apart from being essential building blocks of life, are especially important in the photophysics field as they are some of the hotter molecules of the last and present decade. The fact that these molecules can present such pitfalls at the MP2 level is indeed more relevant as the use of MP2 is very common in photophysics studies.[13,32−35] A normal procedure in such works[36−38] is to optimize ground state structures including dynamic correlation, therefore, the use of MP2 is rather general. In addition, Franck−Condon vertical excitations are carried out at the ground state minimum structure, and all subsequent studies on the excited states are started from that point. Thus, it is essential to get a proper starting point (ground state minimum) to perform an accurate study as minor geometrical changes can imply a change in the order of the states.

Another point of interest is to determine to what extent such basis set deficiencies are localized in a region of the molecule and whether it would be sufficient to correct for intramolecular BSSE only locally, that is, using a specific Counterpoise function that would take into account only a subset of atoms of the system. This would likely be the case of an intermolecular hydrogen bond formation or for instance the interaction between the two ends of a long chain-like molecule. In such cases, a local treatment may be of use not just because BSSE correction would be irrelevant in most parts of the molecule but also to avoid spurious effects of the CP-correction itself. In this respect, with this paper we also aim to show that when the overlap between fragments is strong (i.e., when breaking a chemical bond) the electronic state of the fragment and that of the ghost-orbital calculation might differ, causing spurious CP corrections. Our results indicate that in the intramolecular case it is not of utmost importance which is the electronic state of each fragment, but that the isolated fragment calculation and the corresponding ghost orbital calculation must correspond to the *same* state to obtain a proper BSSE removal.

Computational Details

All ab initio calculations have been carried out with Gaussian 03[39] program. Standard CP-corrected geometry optimizations and vibrational frequency calculations have been performed using the automatic procedure as implemented in Gaussian 03 at the MP2 level of theory (frozen core). For special Counterpoise function definitions we have also used our own code, which allows us to exploit symmetry if any and also permits the use of different specific Gaussian keywords for each fragment calculation (with the Counterpoise keyword the process is automatized in Gaussian 03 but all fragment calculations share the same options).

Thymine and uracil nucleobases were optimized within *Cs* symmetry. No symmetry constraint other than planar ring was used for cytosine, guanine and adenine. An active space including all π orbitals (10 electrons in 8 orbitals for thymine) was used for the Complete Active Space Self Consistent Field (CASSCF) calculations.

Results and Discussion

Let us focus first in the particular case of thymine. We have performed geometry optimizations and frequency calculations at the MP2 and CASSCF levels of theory for the same group of basis sets used by Moran et al. in the benzene case (over 24 basis sets featuring Pople's 3-21G, 6-31G, and 6-311G families and Dunning's cc-pVXZ basis). At the MP2 level,

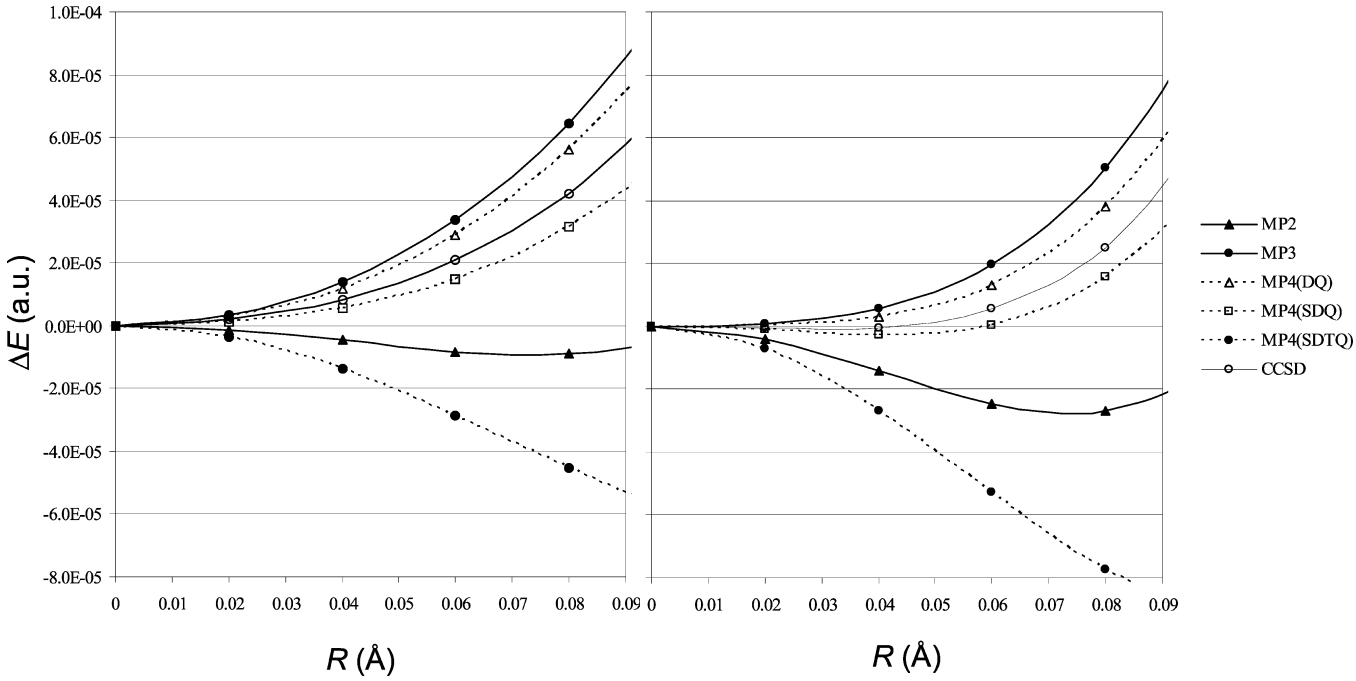

Figure 2. CCSD and MPn energies along the vibrational mode associated to the imaginary frequency for thymine at the MP2 level with the 6-31+G* (left) and 6-311+G*(right) basis sets.

we have obtained imaginary frequencies for the planar optimized structures for 12 of the basis sets used (see Table S1 in the Supporting Information). For particularly unbalanced basis sets such as 6-311++G and 6-311+G, up to three imaginary frequencies have been found. The results are slightly worse than in the case of benzene, as for thymine also the MP2/6-31+G* and MP2/6-31+G** lead to spurious results. Again, these spurious imaginary frequencies were found in correspondence to out of plane bending low-lying modes.

At the CASSCF level, no imaginary frequencies have been found in any case showing that the problems could be associated to two-electron excitations to high energy virtual orbitals with large diffuse character. We have also performed single-point calculations at higher levels of theory with the 6-31+G* and 6-311+G* basis sets along the (uncorrected) vibrational mode associated to the imaginary frequency at the MP2 level of theory (see Figure 2). In both cases, MP3 and MP4(DQ) energies produce the correct profile. The inclusion of the triples in the MP4 energy expression leads to a wrong profile and the inclusion of singles leads to wrong description only for the large basis set. This is also the case for the CCSD method, for which the energy profile is extremely flat in the case of the 6-311+G* basis. This simple analysis shows that these problems are not restricted to the MP2 level and may occur even at the CCSD level. It is also difficult to infer that the problem at the MP2 level might be due to a convergence problem of the MPn series. To answer these questions a much more systematic study would be required, which is beyond the scope of the present work.

Motivated by these similarities with the benzene case, we have attempted an analogous approach for the BSSE removal based upon (mainly) diatomic fragments. Due to the heteroatomic character of the nucleobases (See Scheme 1) we encounter, aside of C−H moieties, other units such as C═O, N−H or CCH_3 fragments. Lewis structures suggested the use of doublet and singlet multiplicities for C−H and C═O fragments, and numerical evidence recommended to use triplet and quadruplet for N−H and CCH_3 fragments, respectively. The reasons behind this choice will be made clear later on.

Table 1. CP-Corrected and Uncorrected Frequencies of Optimized Planar Structures of Pyrimidine Nucleobases[a]

	thymine		uracil		cytosine	
	MP2	CP-correct.	MP2	CP-correct.	MP2	CP-correct.
6-31G*	107	106	134	135	128	128
	138	139	159	156	203	202
	148	153	371	369	357	358
6-31+G*	*80i*	71	*20i*	86	96	105
	100	109	132	138	190	193
	140	153	315	335	337	349
6-311G*	103	104	137	136	123	124
	139	141	151	152	203	201
	147	151	370	370	360	359
6-311+G*	*160i*	75	*113i*	93	67	108
	84	108	103	137	189	193
	140	151	280	342	309	359

[a] Imaginary frequencies are displayed in italics.

The structures were reoptimized according to the total CP-corrected energy and CP-corrected frequency calculations were performed on the CP-optimized planar stationary structures. In Table 1, we present the three lowest vibrational frequencies for thymine, uracil and guanine obtained for four selected basis set cases. With the above-mentioned fragment definition the CP procedure provided excellent results in all cases. The imaginary frequencies were removed in the problematic cases and no significant effect was observed for those which showed proper behavior.

Nucleobases present less symmetry constraints than benzene and other arenes considered in our previous work. Thus, in the present case, one has more freedom to choose proper

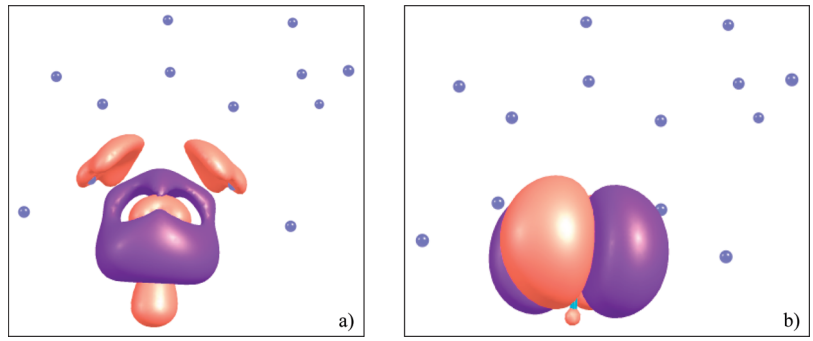

Figure 3. Density difference plot between ghost-orbital and isolated calculation for a N−H fragment in thymine for (a) triplet and (b) singlet electronic states. The position of the ghost-atoms is shown with semitransparent blue spheres. See text for isosurface values.

fragments. In fact, if the basis set deficiencies would be rather localized, one could use a Counterpoise function including only those fragments that would be needed to correct for such deficiencies. Accordingly, we have explored several fragment definitions and Counterpoise functions for this system. Some of our findings are described next.

First of all, the use of different multiplicity on the N−H fragments has a dramatic effect on the low lying out-of-plane mode. The reason is that, in the singlet case, conventional ghost orbital calculations lead to a qualitatively different state than that of the isolated fragment calculation. The explanation is simple in terms of molecular orbital occupations. In the singlet case the HOMO corresponds to one of the p_x,p_y degenerate orbitals in the isolated fragment calculation. However, in the ghost-orbital calculation, this degeneracy is broken and the in-plane p orbital is stabilized by the presence of ghost-orbitals (mainly of s symmetry) of the neighboring atoms. Energetically speaking there is no apparent problem in the energy difference between the isolated and ghost-orbital calculations. However, if the HOMO in the isolated fragment calculation does not happen to have the same orientation as in the ghost-orbital calculation artificial effects appear beyond energy correction, namely, first- and second derivatives of the energy. This can be visualized by comparing the difference between the two densities obtained with and without ghost orbitals at the Hartree−Fock level, as shown in Figure 3. The position of the ghost-atoms is represented by semitransparent blue spheres. The isosurface value in the triplet case is set to 0.0005. Thus, the differences are very small and partially localized in the closest carbon ghost-atoms. Because of BSSE-like basis set extensions, the density is redistributed in the ghost-orbital calculation, increasing in the vicinity of the closest atoms from which the basis functions are used and slightly decreasing in the region close to the nuclei. However, in the case of the singlet calculation, the density difference between the ghost-orbital calculation and the isolated fragment is much larger (isosurface value is set to 0.005 for clarity) and localized in the N−H unit. The typical polarized picture suggests that the two densities correspond to two rotated electron distributions. The inclusion of such energy (and specially energy derivative) differences in eq 1 leads to an essentially wrong CP-correction, which has no correcting effect on the out-of-plane molecular distortions and introduces spurious effects on the stretching modes associated to the N−H moieties. A similar effect has been observed for the rather unchemical C−CH_3 fragment arising from the methyl substituent in the heterocycle. The fragment in the doublet state exhibits a double bond between the carbon atoms whose orientation is again strongly affected by the presence of ghost-orbitals. Such a fragment definition leads to meaningless CP-corrected frequencies. Of course, these problems could be solved simply by proper rotation of the orbitals of isolated fragment calculation, but this might not be easily achieved in automatized procedures such as the Counterpoise keyword in Gaussian 03, for instance. This just shows that one must be very careful in these cases when carrying out routine ghost-orbital calculations to quantify basis set extension effects.

We also explored the effect of using multiplicity specification for the C−H and C═O fragments and no noticeable differences were observed. In the case of the C−H fragment one might foresee similar problems associated with the partial occupation of degenerate p orbitals in the low spin case. It seems rather fortunate that the conventional ghost-orbital calculation lead to a similar orientation of the SOMO orbital.

Another point of interest is to determine to which extent the intramolecular BSSE exhibited by this system is a local effect or not. For this we have considered separately each fragment contribution to the Counterpoise correction and obtained the corresponding CP-corrected frequencies. It is worth to mention that the CP-optimization does not lead to meaningful geometry changes with respect to the conventional MP2 calculation. For instance, the largest deviation on the internal coordinates of thymine at the MP2/6-31+G* level induced upon CP-optimization were just 0.009 Å and 0.3° in bond distance and angles, respectively. That means that one can reasonably obtain frequency corrections with partial Counterpoise functions on the same CP-optimized geometry obtained with the full Counterpoise correction, which largely simplifies the following analysis.

At the MP2/6-31+G* level, the lowest lying out-of-plane vibration for planar thymine shows an imaginary frequency of 80*i*. The use of a CP-correction including contributions from just one of the six fragments depicted in Figure 4 does not lead in any case to a change in the topology of the stationary point. The value of the imaginary frequency decreases in all cases, reaching a highest value of 40*i* in the best case, for the N−H fragment in ortho position with respect to the C−H group (number 4 in Figure 4). Already

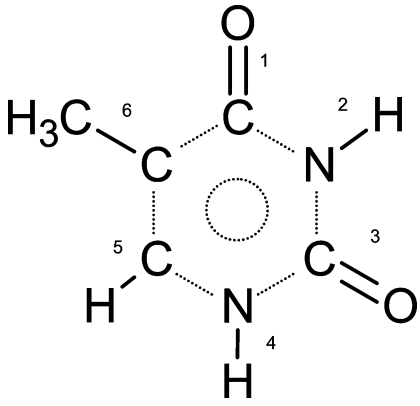

Figure 4. Intramolecular fragments used for CP-correction in thymine.

when considering two adjacent fragment's contributions at a time one can observe a change in the topology. Using fragments 4−5 and 3−4 in the Counterpoise function the lowest frequency assumes values of 39 and 37 cm^{-1}, respectively. Nevertheless, a similar value (35 cm^{-1}) is obtained including distant fragments in para position like 1 and 4. Other combinations involving the C−H fragment also provide corrections in the proper direction. With the progressive inclusion of more fragment contributions to the CP-correction the value of the lowest frequency increases monotonically up to the reported value of 71 cm^{-1} when using all six fragments. For instance, using the three adjacent fragments 3,4,5 the value is 51 cm^{-1} and including all contributions except that of the $C-CH_3$ fragment, an almost converged value of 63 cm^{-1} is obtained. With this analysis we can conclude that the intramolecular BSSE effects were to some extent localized around the N−H moiety in ortho position with respect to the C−H group. However, its removal is not enough to produce a change of topology and contributions from adjacent C−H and C═O groups must also be taken into account. We have also seen that contributions from distant fragments with little chemical significance such as the $C-CH_3$ could be safely ignored from the CP function if necessary. It arises from our results that BSSE effects seem to be quite delocalized on the heterocycle and accordingly to get a proper BSSE removal all fragment contributions should be taken into account. Nevertheless, we do not expect this to be a general trend for intramolecular BSSE problems. Further studies on the determination of the local character of BSSE effects in intramolecular hydrogen bonding situations are currently under work.

After considering the thymine case in deep detail, in the following, we will describe the results obtained for the rest of nucleobases we have considered.

The three lowest vibrational frequencies obtained for all DNA and RNA nucleobases at the MP2 level with four representative basis sets are given in Tables 1 and 2. The CP-corrected values obtained after proper intramolecular BSSE removal are also provided. Thymine and uracil structures were optimized with *Cs* symmetry, whereas guanine, adenine, and cytosine present a planar ring with a NH_2 group slightly out of plane due to pyramidalization that breaks the symmetry. CP-corrected results obtained suggest that geometrical reoptimizations might not be necessary as only very minor changes are observed between the uncorrected and CP-corrected geometries. Nevertheless, all CP-corrected frequency calculations were carried out upon CP-optimized geometries.

Table 2. CP-Corrected and Uncorrected Frequencies of Optimized Planar Structures of Purine Nucleobases[a]

	adenine		guanine	
	MP2	CP-correct.	MP2	CP-correct.
6-31G*	159	158	129	127
	207	205	151	152
	272	272	192	194
6-31+G*	126	151	*55i*	116
	185	192	128	133
	275	262	165	178
6-311G*	160	161	131	133
	213	208	156	155
	273	274	195	194
6-311+G*	139	150	8	113
	192	195	127	131
	274	276	164	175

[a] Imaginary frequencies are displayed in italics.

Fortunately, all optimized structures using the common 6-31G* and 6-311G* basis sets were characterized as true minima. The CP-correction (see below) did not change this situation and in fact, the values of the three lowest frequencies were obtained within a deviation of 3% with respect to the corresponding conventional calculation.

When using diffuse functions the intramolecular BSSE effects can be very important. Concerning the CP-correction, pyrimidine derivatives present no special difficulties for a proper fragment definition. They are characterized by a six member ring and six substituents, (except for cytosine, which has one unsubstituted position in the ring). Uracil is very similar to thymine as they only differ by a methyl group, which in the case of uracil is a C−H unit. Accordingly, the results obtained for uracil follow the same tendency as those observed for thymine. Imaginary frequencies associated to the low lying out-of-plane mode are found for the 6-31+G* and 6-311+G* basis sets. When no diffuse functions are included in the basis set, the planar structures correspond to true minima. Again, the CP-correction using the analogous fragment definitions as in the case of thymine is able to account for this pitfall in both cases, with very little effect on the already correct descriptions.

Interestingly, no imaginary frequencies have been found in the case of cytosine, despite his similarity with thymine and especially with uracil (see Figure 1). Yet, again for the 6-31+G* and 6-311+G* basis sets the values of the lowest vibrational frequencies are somewhat too small compared with the results obtained with more balanced basis sets. Hence, there are some intramolecular BSSE effects but not to the extent of changing the topology of the planar stationary points. In fact, cytosine shares with the other two nucleobases the same three substituents that were observed to contribute more to the BSSE effects in thymine. This shows again that for such heterocyclic systems the BSSE effects are rather subtle and delocalized. For completeness we have performed also CP-corrected optimizations and frequency calculations for this system. The absence of substituent in ortho position with respect to the C═O group leads to the difficulty of

dealing with a single-atom fragment; highly symmetric in its isolated state but not in the presence of ghost-orbitals. To avoid the problems described before we have considered the N atom as a fragment in high spin state. As alternative, we have explored also the use of a larger fragment involving two adjacent positions of the ring, namely NC-NH_2. Both fragment definitions lead to very similar results (maximum deviation of 8 cm^{-1} in the third frequency at the MP2/6-31+G* level), and only the results from the first option are reported in Table 1. Even though no imaginary frequencies were observed, the CP-corrected frequencies are more similar among the four basis sets used than the uncorrected ones. For instance, the somewhat too low values of 67 and 309 cm^{-1} obtained with the 6-311+G* basis set are blue-shifted by 40 and 50 cm^{-1} to a value in much better agreement with the one obtained with the 6-311G* basis set.

Finally, the results obtained for adenine and guanine basis sets are collected in Table 2. The molecule of adenine seems to be the less prone to intramolecular BSSE effects. No imaginary frequencies have been observed for the planar optimized structures and only a slight drop of about 20−25 cm^{-1} in the value of the lowest vibrational frequency when including diffuse functions has been observed. The situation in the case of guanine is different as again difficulties are observed specially with the 6-31+G* basis set, for which an imaginary frequency of 55*i* is obtained. The value of 8 cm^{-1} obtained with the 6-311+G* basis set can also be considered as spurious.

Purine bases are characterized by a heterocyclic six-membered ring fused to an imidazole ring. Therefore, the definition of fragments to account for intramolecular BSSE must be somewhat different to that of the pyrimidine bases discussed above. The Lewis structures for each molecule show the presence of a double bond involving a C−C pair in the edge of the two fused rings which probably should not be broken. Following this premise, we ended up with C═O, N−H, and C═C diatomic fragments and larger N═C-NH_2 and N═CH fragments involving the unsaturated N atoms, which were considered in high spin state. Once again, the results obtained are very satisfactory. For guanine, the two wrong vibrational frequencies obtained with the 6-31+G* and 6-311+G* basis are efficiently removed with the CP-correction. Even in the case of adenine, where the BSSE effects were less pronounced, the CP-correction induced a slight blue-shift in the lowest vibrational frequencies to final values much closer to those obtained with more balanced basis sets. Several other fragment definitions were also tested, for instance involving a central NC═CN fragment. The CP-corrected results were proved to be very similar, provided situations like those described in detail in the case of thymine were not present.

Conclusions

We have shown that MP2 calculations combined with conventional basis sets including diffuse functions such as the 6-31+G* or 6-311+G* can incorrectly predict imaginary frequencies associated to out-of-plane vibrational modes of planar optimized structures of molecules of utmost importance such as the DNA and RNA nucleobases. Other basis sets like cc-pVXZ and aug-cc-pVXZ seem to be less prone to basis set deficiency problems and are more recommended for vibrational frequencies analysis.

The origin of such pitfalls has been demonstrated to be rooted in intramolecular basis set deficiencies, which eventually lead to intramolecular BSSE effects, similarly to the case of benzene and other planar arenes for which such problems have already been detected and analyzed in detail.

The application of conventional BSSE-correction techniques, such as the Counterpoise method, provide once again proper assessment and correction whenever spurious results occur and do not produce meaningful effects in those cases already correctly described. However, special care must be taken when dealing with strongly overlapping fragments (i.e., when breaking a chemical bond). Even though our results indicate that it is not of utmost importance which is the electronic state of each fragment, it is very important to make sure that isolated fragment and the associated ghost orbital calculations must correspond to the same state with the same orientation of singly occupied degenerate orbitals to obtain a proper BSSE removal.

Acknowledgment. Support for this work from the Spanish Ministerio de Ciencia y Tecnología (project No. CTQ2005-02698/BQU) and from the FPU program (Grant No. AP2004-4774). We are also grateful to Dr. Lluís Blancafort for useful discussions.

Supporting Information Available: Table showing the lowest harmonic vibrational frequencies. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Dkhissi, A.; Adamowicz, L.; Maes, G. *J. Phys. Chem. A* **2000**, *104*, 2112–2119.

(2) Goodman, L.; Ozkabak, A. G.; Thakur, S. N. *J. Phys. Chem.* **1991**, *95*, 9044–9058.

(3) Lampert, H.; Mikenda, W.; Karpfen, A. *J. Phys. Chem. A* **1997**, *101*, 2254–2263.

(4) Martin, J. M. L.; Taylor, P. R.; Lee, T. J. *Chem. Phys. Lett.* **1997**, *275*, 414–422.

(5) Michalska, D.; Zierkiewicz, W.; Bienko, D. C.; Wojciechowski, W.; Zeegers-Huyskens, T. *J. Phys. Chem. A* **2001**, *105*, 8734–8739.

(6) Moran, D.; Simmonett, A. C.; Leach, F. E.; Allen, W. D.; Schleyer, P. V.; Schaefer, H. F. *J. Am. Chem. Soc.* **2006**, *128*, 9342–9343.

(7) Saeki, M.; Akagi, H.; Fujii, M. *J. Chem. Theory Comput.* **2006**, *2*, 1176–1183.

(8) Shahbazian, S. *Chem. Phys. Lett.* **2007**, *443*, 147–151.

(9) Simandiras, E. D.; Rice, J. E.; Lee, T. J.; Amos, R. D.; Handy, N. C. *J. Chem. Phys.* **1988**, *88*, 3187–3195.

(10) Torii, H.; Ishikawa, A.; Takashima, R.; Tasumi, M. *J. Mol. Struct. (TheoChem)* **2000**, *500*, 311–321.

(11) Asturiol, D.; Duran, M.; Salvador, P. *J. Chem. Phys.* **2008**, *128*, 144108.

(12) Balabin, R. M. *J. Chem. Phys.* **2008**, *129*, 164101.

(13) Hobza, P.; Havlas, Z. *Theor. Chem. Acc.* **1998**, *99*, 372–377.

(14) Salvador, P.; Paizs, B.; Duran, M.; Suhai, S. *J. Comput. Chem.* **2001**, *22*, 765–786.

(15) Salvador, P.; Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **2000**, *113*, 5666–5674.

(16) Halasz, G. J.; Vibok, A.; Suhai, S.; Mayer, I. *Int. J. Quantum Chem.* **2002**, *89*, 190–197.

(17) Jensen, F. *Chem. Phys. Lett.* **1996**, *261*, 633–636.

(18) Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2001**, *105*, 1944–1950.

(19) Salvador, P.; Duran, M.; Dannenberg, J. J. *J. Phys. Chem. A* **2002**, *106*, 6883–6889.

(20) Sellers, H.; Almlof, J. *J. Phys. Chem.* **1989**, *93*, 5136–5139.

(21) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.

(22) Iwata, S.; Nagata, T. *Theor. Chem. Acc.* **2007**, *117*, 137–144.

(23) Mayer, I. *Int. J. Quantum Chem.* **1983**, *23*, 341–363.

(24) Nagata, T.; Iwata, S. *J. Chem. Phys.* **2004**, *120*, 3555–3562.

(25) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024–11031.

(26) van Mourik, T.; Karamertzanis, P. G.; Price, S. L. *J. Phys. Chem. A* **2006**, *110*, 8–12.

(27) Holroyd, L. F.; van Mourik, T. *Chem. Phys. Lett.* **2007**, *442*, 42–46.

(28) Shields, A. E.; van Mourik, T. *J. Phys. Chem. A* **2007**, *111*, 13272–13277.

(29) Mayer, I.; Hamza, A. *Int. J. Quantum Chem.* **2003**, *92*, 174–180.

(30) Alexander, M. H. *J. Chem. Phys.* **1993**, *99*, 6014–6026.

(31) Salvador, P.; Mayer, I. *J. Chem. Phys.* **2004**, *120*, 5882–5889.

(32) Broo, A.; Holmen, A. *J. Phys. Chem. A* **1997**, *101*, 3589–3600.

(33) Cai, Z. L.; Sha, G. H.; Zhang, C. H.; Huang, M. B. *Chem. Phys. Lett.* **1991**, *178*, 273–278.

(34) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1996**, *118*, 6811–6821.

(35) Leszczynski, J. *Int. J. Quantum Chem.* **1992**, *19*, 43–55.

(36) Hudock, H. R.; Levine, B. G.; Thompson, A. L.; Satzger, H.; Townsend, D.; Gador, N.; Ullrich, S.; Stolow, A.; Martinez, T. J. *J. Phys. Chem. A* **2007**, *111*, 8500–8508.

(37) Kobayashi, R. *J. Phys. Chem. A* **1998**, *102*, 10813–10817.

(38) Perun, S.; Sobolewski, A. L.; Domcke, W. *J. Phys. Chem. A* **2006**, *110*, 13238–13244.

(39) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery Jr., 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.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revison C.01; Gaussian, Inc.: Pittsburgh, PA, 2003.

CT900056U