THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Stacked and H-Bonded Cytosine Dimers. Analysis of the Intermolecular Interaction Energies by Parallel Quantum Chemistry and Polarizable Molecular Mechanics.

Nohad Gresh,[*,†,‡,§] Judit E. Sponer,[∥,⊥] Mike Devereux,[#] Konstantinos Gkionis,[∥] Benoit de Courcy,[†,‡,§] Jean-Philip Piquemal,[‡,§] and Jiri Sponer[∥,⊥]

[†]Chemistry & Biology, Nucleo(s)tides & Immunology for Therapy (CBNIT), CNRS UMR8601, Université Paris Descartes, PRES Sorbonne Paris Cité, UFR Biomédicale, 45 rue des Saints-Pères, 75270 Paris Cedex 06, France

[‡]Laboratoire de Chimie Théorique, Sorbonne Universités, UPMC, Paris 6, case courrier 137, 4, place Jussieu, Paris, F75252, France

[§]Laboratoire de Chimie Théorique, UMR 7616 CNRS, case courrier 137, 4, place Jussieu, Paris, F75252, France

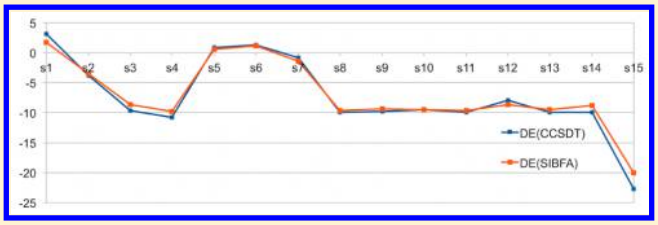
[∥]Institute of Biophysics, Academy of Sciences of the Czech Republic, Kralovopolska, 135, 612 65 Brno, Czech Republic

[⊥]CEITEC − Central European Institute of Technology, Campus Bohunice, Kamenice 5, 625 00 Brno, Czech Republic

[#]Department of Chemistry, University of Basel, Klingelbergstrasse 80, Basel CH 4056, Switzerland

Supporting Information

ABSTRACT: Until now, atomistic simulations of DNA and RNA and their complexes have been executed using well calibrated but conceptually simple pair-additive empirical potentials (force fields). Although such simulations provided many valuable results, it is well established that simple force fields also introduce errors into the description, underlying the need for development of alternative anisotropic, polarizable molecular mechanics (APMM) potentials. One of the most abundant forces in all kinds of nucleic acids topologies is base stacking. Intra- and interstrand stacking is assumed to be the most essential factor affecting local conformational variations of B-DNA. However, stacking also contributes to formation of all kinds of noncanonical nucleic acids structures, such as quadruplexes or folded RNAs. The present study focuses on 14 stacked cytosine (Cyt) dimers and the doubly H-bonded dimer. We evaluate the extent to which an APMM procedure, SIBFA, could account quantitatively for the results of high-level quantum chemistry (QC) on the total interaction energies, and the individual energy contributions and their nonisotropic behaviors. Good agreements are found at both uncorrelated HF and correlated DFT and CCSD(T) levels. Resorting in SIBFA to distributed QC multipoles and to an explicit representation of the lone pairs is essential to respectively account for the anisotropies of the Coulomb and of the exchange-repulsion QC contributions.



INTRODUCTION

Nucleic acids (DNA and RNA) belong to the most important classes of biomolecules. Therefore, nucleic acids and their molecular complexes with proteins and ligands have been frequently studied by diverse computational techniques, ranging from the highest accuracy quantum-chemical calculations[1] up to coarse-grain modeling.[2] Although all types of computations can provide useful results when applied appropriately and for wisely selected goals, perhaps the most biochemically compelling computational results for nucleic acids were so far obtained by approaches based on atomistic empirical potentials (force fields), as exemplified by atomistic explicit solvent molecular dynamics (MD) simulations.[3] This is probably because the atomistic potentials represent the best trade-off between the three contradicting requirements for realistic computations on nucleic acids, that is, accuracy of the primary description of the potential energy surface, realistic inclusion of their environment and conformational sampling. The quality of computations is usually determined by the least satisfactory component of the computations. Thus, with the advance of routine microsecond-scale simulations and robust enhanced sampling methods, the quality of the force fields is becoming the key weakness of the approaches based on atomistic force fields. Contemporary simulations are still based on relatively simple ∼20 years old "second-generation" pair additive molecular mechanics force fields[4,5] and their subsequent tuning, responding to urgent problems spotted in simulation studies.[6−11] Despite the numerous successes, fully accurate MD simulation description of even small nucleic acids systems is not always obtained. In addition, benchmark quantum chemical calculations of nucleic acids building blocks indicate that the accuracy that can be achieved by tuning of the currently dominant atom−atom pair-additive force fields based on the fixed point charge model may have principal physical

Received: February 19, 2015
Revised: June 25, 2015
Published: June 29, 2015

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

limits.[12,13] Therefore, development of more sophisticated force fields better reflecting the changes of the electronic structure upon conformational changes are highly desirable, despite their enormous conceptual, parametrization and tuning challenges. Recently, the first polarizable DNA simulation force field based on a Drude oscillator formalism has been parametrized.[14]

Nucleic acids and their complexes have lent themselves to numerous simulations in the context of classical molecular dynamics.[4] MD simulations have provided invaluable information regarding their sequence-dependent conformations,[15] response to stress,[16] and their binding to ligands[17,18] as well as to proteins.[19,20] Nevertheless several limitations were progressively unraveled.[3] These are due to the polyanionic character of the phosphodiester backbone,[21] and to the highly polar and polarizable nature of the bases.[22,23] Thus, both the backbone and the grooves of nucleic acids bind to, and nucleate, highly structured water molecules,[24,25] and both bind metal cations,[26] which can play either a structural role or act as enzyme cofactors. The interactions which involve metal cations are notoriously known to include sizable polarization effects[27] [and references therein] and would need polarizable potentials for an accurate description. This is also the case for water clusters for which polarization was shown to be a major contributor to the total interaction energy, with a weight increasing with the size of the cluster due to cooperativity.[28] Within the framework of pair-additive force fields, the polarization effects can be included only indirectly. The need for polarizable potentials could be particularly exacerbated in the vicinity of these clusters to polyanionic solutes. An additional issue relates to the anisotropy of the total intermolecular interaction energies. This can be exemplified by the stacking interactions of the bases.[29,30] We wish to monitor the evolutions of the quantum chemistry (QC) intermolecular interaction energy, ΔE, and of its individual contributions upon rotating and translating one base over the other, and evaluate how reliably these evolutions could be accounted for by APMM approaches. Such issues are a critical factor for the perspective of reliable, long-duration molecular dynamics with APMM potentials, to account for sequence-dependent conformations of DNA and RNA.

At this stage, it is appropriate to set the present approach in the general context of development of advanced intermolecular potentials. With such potentials, all energy contributions, not only E_{MTP}, should be proven to display the proper angular dependencies. Yet this was shown not to be the case for E_{rep} if simply formulated as a sum of atom−atom terms,[31,32] and this could impair the overall accuracy of potentials resorting to such a formulation. The effective fragment potential (EFP)[33] uses the same distributed dipole polarizabilities as SIBFA, expresses E_{rep} as a function of intermolecular overlap, and should thus be endowed with the correct angular dependencies. On the basis of the results published in ref 34 regarding the stacked and H-bonded G-C and A-T dimers, an accuracy comparable to that of the present study is expected. There is one reservation, however. These EFP results did not embody the charge-transfer contribution, even though it has been formulated explicitly in the method, because as stated in ref 34, it had been previously shown to be small for most dimers of neutral molecules. This is not the case, however, for the doubly H-bonded Cyt dimers, for which E_{ct} with the aug-cc-pVTZ basis set amounts to −4 kcal/mol and has a comparable magnitude in SIBFA. EFP conformational energy studies have not been reported so far to our knowledge, while the extension to the sugar−phosphate chain is one of the next steps toward a generalized application of SIBFA to nucleic acids. There are also promising emerging QC-grounded approaches, such as a quantum mechanical force field based on a divide-and-conquer framework (QMFF/mDC)[35] the fragment molecular orbital method (FMO),[36] the X-Pol potential,[37] and the quantum chemical topology force-field (QCTFF),[38,39] and the ab initio force-field (AIFF) method.[40] In the context of AIFF, interactions between neighbor fragments are computed by QC and those between farther ones are computed with distributed multipoles and polarizabilities for EC and E_{pol}, respectively, but there are no overlap-dependent contributions (repulsion, penetration and charge-transfer). This could certainly be justified on account of the larger intermolecular interaction distances. The fact that dimer interactions are computed by QC disables a comparison with the present dimer calculations, but it would be instructive to evaluate the merits of the two approaches upon passing to large oligomer complexes. Both QMFF/mDC and QCTFF resort to distributed multipoles on individual fragments. In QMFF/mDC, they are derived from the fragment density matrix resulting from QC computations in the presence of the potential exerted by the other fragments. In X-Pol and FMO, a large molecule is divided into fragments, and the fragment MO's are computed in the presence of the embedding potential due to the other fragments. In QCTFF, the amplitudes of the multipoles can be predicted as a function of geometry changes after a process of machine learning. In both approaches, polarization effects could be accounted for by the changes of the amplitudes of the multipoles. There is no emphasis on the short-range contributions, which are presently not included in QCTFF. They are in QMFF/mDC presently limited solely to the repulsive term, as a sum of interfragment atom−atom $1/R^{12}$ terms along with attractive $1/R^6$ ones, and this is also the case with X-Pol. Such a simplified expression for E_{rep} would preclude a comparison with the present calculations, which underline the need to account for anisotropy.

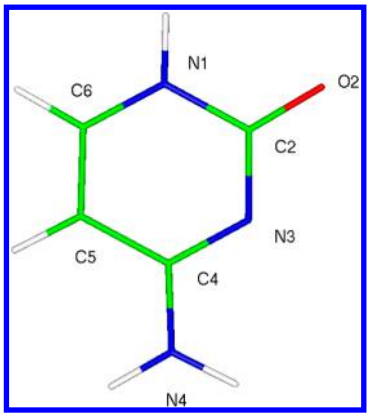
We have previously validated the advanced APMM SIBFA polarizable force field[31,41,42] for binding of the hydrated divalent cations Mg(II) and Zn(II) to guanine, adenine, G-C and A-T base pairs[43] and the guanosine monophosphate.[44] Those studies were motivated by the outstanding importance of such cations in the stabilization of nucleic acids and the known inability of standard biomolecular force fields to describe interactions between divalent ions and biopolymers, due to very significant charge-transfer and polarization contributions. The SIBFA computations neatly differentiated between the two cations and successfully described their distinct binding modes to the nucleotides.

In the present study, we tackle one of the dominant forces in nucleic acids, namely base stacking. In principle, force field description of base stacking is more satisfactory compared to binding of divalent cations. On the other hand, base stacking is so ubiquitous in nucleic acids that even rather small discrepancies in its description may be significant, especially when the description does not appropriately include polarization effects that can be associated with variable stacking patterns. As a benchmark, we will use the QC study by Jurecka et al. on 14 cytosine (Cyt) dimers.[45] This reference set includes structures derived from a starting structure with maximal overlap by rotations of the second monomer around the z axis and horizontal translations, yielding a total of 14 complexes covering diverse parts of the potential energy surface. The stacked structures were complemented by a doubly H-bonded

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

symmetrical Cyt-Cyt base pair with two N4−H···N3 H-bonds, where the extracyclic N4 of each Cyt donates a proton to the N3 atom of the other Cyt. The convention for atom numbering in cytosine is recalled in Scheme 1.

Scheme 1



In the present study, we will redo these computations using the aug-cc-pVTZ (-f) basis set,[46,47] with which a new library of SIBFA fragments is being constructed.[48] The intermolecular interaction energies will be decomposed in terms of their individual contributions. Such analyses will be done at both Hartree−Fock (HF) and correlated levels. We will accordingly perform the comparisons with SIBFA in two successive stages. In the first stage, comparisons with the HF results will be done with uncorrelated multipoles and polarizabilities derived for the cytosine molecular fragment. In the second stage, the comparisons with correlated DFT and CCSD(T) computations will accordingly resort to correlated multipoles and polarizabilities. The fact that the two Cyt monomers have an extensive overlap constitutes a demanding test for molecular mechanics force-fields in general. Indeed, as shown below, the two first-order QC contributions, EC (Coulomb), and EX (exchange) display marked directionalities even though the amplitude of overlap changes little upon the z rotations. Regarding the translations, their quantitative evolutions upon reduction of the overlap should be accounted for quantitatively as well by APMM. The pioneering studies by Langlet et al.[49] have considered several hydrogen-bonded and stacked bases and shown that multipole−multipole interaction terms, up to quadrupoles, had weights within the total electrostatic contribution comparable, if not larger, than the monopole−monopole (ie charge−charge) term. We had considered ourselves the variations of the stacking energies of two formamides upon rotation of the second monomer over the other around the z axis.[31] It was found that a multipolar formulation of the electrostatic contribution augmented with an explicit penetration term (E_{MTP*}) gave rise to a curve which superimposed virtually completely with that of the Coulomb contribution EC. We will here evaluate to what extent E_{MTP} is able to describe EC for the Cyt dimers, which exhibit a much larger mutual overlap than the formamide dimers and a richer variety of electron-rich and electron-deficient sites. If initial fitting is performed at the HF level with multipoles derived from the uncorrelated Cyt molecular orbitals (MO), it will be necessary to next evaluate how well the corresponding variations of EC upon passing to correlated energy-decompositions could be accounted for by E_{MTP*} using multipoles accordingly derived from correlated MO's. Another issue relates to the short-range repulsion EX, which in SIBFA is formulated as sums of bond−bond, bond−lone pair, and lone pair−lone pair interactions. This implies a proper localization and calibration of the π lone-pairs. This will have to be carried out beforehand, upon probing each lone-pair with two probes, namely Mg(II) or water approaching through one of its H atoms at the vertical of each of the seven Cyt atoms, and then validated on the Cyt dimers. The final issue relates to the dispersion contribution. In the context of SIBFA, it embodies an explicit exchange-dispersion component, which can be compared to its counterpart from SAPT energy-decompositions. Both the "genuine" dispersion term and the exchange-dispersion have dependencies upon the lone pairs. They can be both calibrated beforehand upon resorting to a model system, namely the water dimer. We will then evaluate how well such a calibration enables E_{disp} (SIBFA) to reproduce its SAPT counterpart in the 15, much larger, Cyt dimers. Last but not least, it will clearly be essential that the magnitude of the intermolecular interaction energy, ΔE, and of its components, be calculated in a balanced way for the stacked and the H-bonded complexes, where a more limited, yet much stronger, set of in-plane interactions are involved, that is, predominantly with the N4 and N3 nitrogen atoms.

PROCEDURE

Quantum-Chemistry Computations. *1. QC Computations. Uncorrelated Computations.* The energy decompositions were done with the Reduced Variational Space analysis (RVS) of Stevens and Fink.[50] This procedure separates the total interaction energy into four contributions: the first order (E_1) Coulomb (EC) and short-range exchange-repulsion (EX) and second order (E_2) polarization (E_{pol}), and charge-transfer (E_{ct}). The basis set superposition error[51] was evaluated within the virtual orbital space.[52] We used the GAMESS software[53] and the augcc-pVTZ (-f) basis set.[46,47] We will denote below by E_{pol}(VR) ("variational") the difference between ΔE(HF) and the sum of E_1 and E_{ct}, which is the polarization energy at the outcome of the SCF variational procedure, after each monomer has relaxed iteratively in the presence of the other monomer. Such a process should be the equivalent of the mutual correlations of induced dipoles in the context of APMM approaches yielding the corresponding polarization contribution.

Correlated Computations. The energy decomposition analyses were done at the DFT/B3LYP level[54] with the symmetry adapted perturbation theory[55] in which ΔE is decomposed into Coulomb and exchange in first-order, and induction, dispersion, and exchange-dispersion in second-order.[56−60] We used the MOLPRO code.[61]

Reference interaction energies were computed at CBS(T) level ($\Delta E^{CBS(T)}$); i.e., they were derived from the RIMP2-interaction energies extrapolated to complete basis set (abbreviated as RIMP2/CBS) by adding a correction term for the higher order electron correlation contributions according to the following formula:[62,63]

$$\Delta E^{CBS(T)} = \Delta E_{CBS}{}^{MP2} + \Delta E_{6\text{-}31G^{**}}{}^{CCSD(T)} - \Delta E_{6\text{-}31G^{**}}{}^{MP2} \quad (1)$$

where $\Delta E_{6\text{-}31G^{**}}{}^{CCSD(T)}$ and $\Delta E_{6\text{-}31G^{**}}{}^{MP2}$ are the CCSD(T) and MP2 interaction energies computed with the 6-31G** basis set and $\Delta E_{CBS}{}^{MP2}$ is the RIMP2/CBS interaction energy:

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

$$E_X^{HF} = E_{CBS}^{HF} + Pe^{-\alpha X} \quad (2)$$

$$E_X^{corr} = E_{CBS}^{corr} + QX^{-3} \quad (3)$$

where X = 3, 4 stand for the aug-cc-pVTZ, and aug-cc-pVQZ calculations, respectively; E_X^{HF} is the HF component of the total electronic energy obtained with a given basis set; α = 1.43; E_{CBS}^{HF} and E_{CBS}^{corr} are the HF and correlation components of the MP2 total electronic energy extrapolated to the complete basis set, respectively. Thus:

$$E_{CBS}^{MP2} = E_{CBS}^{HF} + E_{CBS}^{corr} \quad (4)$$

In eqs 2 and 3, P and Q are the coefficients of a linear equation and are to be determined for each system separately.

RIMP2/CBS interaction energies were computed with the Turbomole code.[64−66]

$\Delta E_{6\text{-}31G^{**}}^{CCSD(T)}$ and $\Delta E_{6\text{-}31G^{**}}^{MP2}$ contributions were computed with the MOLPRO code.[61]

2. SIBFA Computations. In the SIBFA procedure,[31,41,42] the intermolecular interaction energy is computed as the sum of five contributions: electrostatic multipolar (E_{MTP^*}), short-range repulsion (E_{rep}), polarization (E_{pol}), charge transfer (E_{ct}), and dispersion:

$$\Delta E_{TOT} = E_{MTP^*} + E_{rep} + E_{pol} + E_{ct} + E_{disp}$$

E_{MTP^*} was computed with distributed multipoles (up to quadrupoles) derived from the augcc-pVTZ (-f) QC molecular orbitals (MO's) of a given molecular fragment and augmented with an overlap-dependent penetration term.[67] The multipoles were derived from Stone's GDMA analysis[68,69] and distributed on the atoms and the bond barycenters using a procedure developed by Vigné-Maeder and Claverie.[70] This was done by a home-built routine (Devereux, M., Paris, 2010). The anisotropic polarizabilities were distributed on the centroids of the localized orbitals (heteroatom lone pairs and bond barycenters) using a procedure due to Garmer and Stevens.[71] E_{rep} and E_{ct}, the two short-range contributions, were computed using representations of the molecular orbitals localized on the chemical bonds and on localized lone pairs. The cytosine molecular fragment was recomputed at the B3LYP level to derive accordingly the correlated distributed multipoles. The distributed polarizabilities at this correlated level were computed with an in-house version of Hondo/98.[72,89]

The availability of both components of E_{disp}, exchange-free and dispersion-exchange, for the 15 Cyt dimers prompted us to further elaborate on this contribution in an attempt to match both separately in the context of the SIBFA procedure. These developments are presented in the appendix, along with the corresponding new parameters. This enables a complete formulation of E_{disp} in this context.

RESULTS AND DISCUSSION

Interaction with H_2O and Mg(II) probes. The only contribution other than E_{rep} which was fitted by the probe approach was E_{pol}. The parameters for the other contributions in conjunction with the aug-cc-pVTZ(-f) basis set were derived in a recent study.[48]

Calibration of E_{pol}. The calibration of E_{pol} was done upon adjusting the values of the multiplicative factor and of the exponential factor of the Gaussian function which screens the polarizing electrostatic field.[41] We have performed in-plane variations of the approach of Mg(II) to N3 along its external bisector and to O2 along the C═O bond, covering the range of distances 1.7−2.2 Å. The values of these two parameters enabling E_{pol}(SIBFA) to best match the radial behavior of E_{pol}(RVS) are 0.68 and 1.50, respectively, equal and close, respectively, to the default values of 0.68 and 1.40.

Calibration of E_{rep}. The carbonyl O (O2), pyrimidine-like (N1 and N4) and pyridine-like (N3) nitrogens, and N-connected C atoms (C4 and C2) have the effective van der Waals radii that were calibrated and reported in the course of our previous study concerning the formamide and imidazole ligands.[48] The effective radius of the benzene-like C atoms (C5 and C6) was calibrated using the Mg(II) and water probes. The values are reported in the Appendix. Each of the seven heavy atoms was individually probed by water or Mg(II) approaching it through the perpendicular to the Cyt plane. The water approach was through one H atom, the angle probed atom, −H−O being 180°. N3 and O2 were also probed in-plane (see above). This enables calibration of each of the lone-pairs of the seven Cyt heavy atoms individually.

Table 1 lists the definition of each lone-pair in terms of the following: (a) its internal coordinates with respect to its bearer atom, namely the θ and φ angles and its distance to it; (b) its occupation number; (c) the increase of effective radius of the bearer atom along the direction of the lone-pair. The last column gives the designation of the bearer atom along the sequence N1−C6−C5−C4−N3−C2−N4−O2. More specifically, denoting by A the bearer atom, by B, the atom preceding A, and by C, the one preceding B, θ is the L_α−A−B "valence" angle, and φ is the L_α−A−B−C "torsion" angle. Such coordinates locate the sp lone pair of N3 and the two sp^2 lone-pairs of O2 in the positions given by ELF analysis[73,74] in a recent study.[75] The π lone pairs are inclined on both sides of the C plane, rather than at the vertical to it passing through each of the seven heavy atoms. A representation of their location is given in Figure 1. This is also consistent with the

Table 1. Location of the σ and π Lone Pairs of Cytosine (See Text for Definition)[a]

θ	φ	R	occ.	d(vdW)	
110.000	131.000	0.97	0.75	0.55	N1
110.000	229.000	0.97	0.75	0.55	N1
45.000	90.000	0.86	0.75	−0.10	C6
45.000	−90.000	0.86	0.75	−0.10	C6
45.000	90.000	0.86	0.50	0.35	C5
45.000	−90.000	0.86	0.50	0.35	C5
45.000	90.000	1.05	0.50	−0.10	C4
45.000	−90.000	1.06	0.50	−0.10	C4
119.483	180.000	0.80	2.00	−0.05	N3
45.000	90.000	0.86	0.50	0.55	N3
45.000	−90.000	0.86	0.50	0.55	N3
45.000	90.000	1.06	0.50	0.90	C2
45.000	−90.000	1.06	0.50	0.90	C2
45.000	90.000	0.85	1.00	0.25	N4
45.000	−90.000	0.85	1.00	0.25	N4
113.000	0.000	0.80	2.00	−0.10	O2
113.000	180.000	0.80	2.00	−0.10	O2
45.000	90.000	0.85	0.50	0.25	O2
45.000	−90.000	0.86	0.50	0.25	O2

[a]The θ and φ internal coordinate angles are in degrees, while the "R" coordinates and the increments of effective radii d(vdW) are in Ångstrom. The last column specifies the atom bearing the considered lone pair.

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

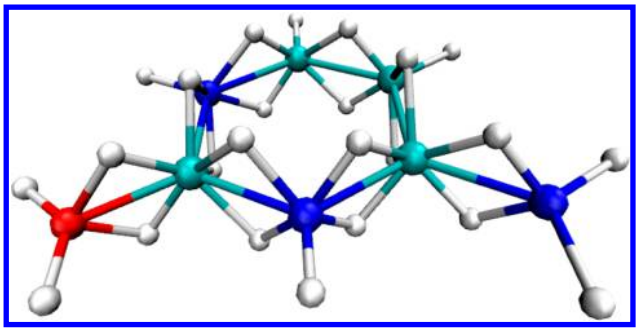


Figure 1. Representation of the lone-pair locations in the cytosine monomer.

ELF analysis. The values of the internal coordinates enable sharing of each π lone pair between the atoms at the origin and the extremity of the bond, above and below the plane (shown in Figure 1). There is a total of 16 lone-pair electrons. The two π lone-pairs of N1 and of C6 both have populations of 0.75. In an alternative representation where the π lone pairs are at the vertical of the bearer atoms, the corresponding populations on N1 and C6 were of 1 and 0.5 respectively. The averaged value of 0.75 is consistent with the adopted "sharing" of the lone-pair between these two atoms. For the present purposes, since the positions of the lone-pairs are set, and since the effective radii of the bearers were obtained from a recent study using the INoLLS interactive, nonlinear least-squares fitting program,[76] the sole remaining "free" parameters are the van der Waals increments along the lone pairs. We discuss shortly below the outcome of the calibration, prior to its application and evaluation on the 15 Cyt dimers. We will retain throughout the denomination "lone-pairs" to denote the location of nonbonded electrons, even though the actual occupation number could be less than two.

Results with the Water Probe. The results are reported in Table 2. We consider in succession: (a) in-plane approach to N3 and O2 in the 1.7—2.2 Å range of H−N/O distances; (b) perpendicular approach to the heteroatoms N1, N3, N4, and O2 in the same range of distances; (c) perpendicular approach to C6, C5, C4, and C2 in the 1.8—2.3 Å range of H−C distances.

a. In-Plane Approach. The calibration of N3 and O2 consisted of simply decreasing their effective radii along their sp^2 lone-pairs by 0.05 and 0.10 Å, respectively. A very close reproduction of ΔE(QC) and its contributions is observed in the whole range of distances (not shown).

b. Perpendicular Approach to the Heteroatoms. In order to ensure a satisfactory reproduction of EX(RVS), it was necessary to systematically increase the effective radii of the bearer atoms along the π lone-pairs, namely by 0.55 Å for N1 and N3, and 0.25 for N4 and O2. E_{rep}(SIBFA) reproduces EX throughout for water approach to N3, but underestimates it for the other atoms for the shorter (d < 2.2 Å) range of distances. This is, however, compensated by a corresponding underestimation of EC by E_{MTP*}, so that E_1(SIBFA) has a better agreement with E_1(RVS) than E_{rep}(SIBFA) with EX. In addition, with the exception of O2, the d < 2.2 Å range corresponds to repulsive values of ΔE(RVS), and is thus less relevant energetically. Both SIBFA E_{pol} and E_{ct} contributions reproduce closely their QC counterparts for distances of 2.2 Å and beyond. The agreement between ΔE(SIBFA) and ΔE(RVS) is satisfactory whenever the binding is attractive.

c. Perpendicular Approach to the C Atoms. The calibration that we found most satisfactory had on the major groove side an alternation of small decrements and average increments of effective radii along the π lone-pairs: −0.10 Å for C6 and C4, and +0.35 Å for C5. On the other hand, it was found necessary to significantly increase the effective radius of C2 on the minor groove side, namely by up to 0.90 Å. The best agreement concerning the short-range repulsion is with atom C2. For the three other C atoms, E_{rep}(SIBFA) is underestimated, but this is again, compensated by E_{MTP*} as compared to EC(RVS). As a consequence, with the exception of C5, the agreements in terms of E_1 are more satisfactory. Attempts to improve the agreements concerning the short-range repulsion were deleterious for the stacked Cyt dimers, and were not pursued. This could be understood owing to the maximized overlaps which take place between the two Cyt planes as contrasted to the limited overlap of the incoming water probe approaching vertically to the Cyt plane, even though the equilibrium distance could be shorter (2.2−2.4 Å as contrasted to 3.3 Å in the stacked Cyt dimers). It is noted that, with the exception of C2, the whole range of distances explored corresponds to overall repulsive values of ΔE, and could thus be of lesser energy relevance. In the range of distances at and past 2.2 Å, E_{MTP*}(SIBFA) can match EC(RVS) satisfactorily, except in the case of water binding to C6, in marked contrast with its close agreement with EC when water binds the neighboring atoms C5 and N1. The very flat and slightly repulsive behavior of both EC(RVS) and E_{MTP*}(SIBFA) for water binding to C4 can be noted. As for heteroatom binding, both E_{pol}(RVS) and E_{ct}(RVS) are well reproduced by their SIBFA counterparts, at 2.1 Å and beyond.

Results with the Mg(II) Probe. The results are reported in Table 3.

a. In-Plane Binding. We have considered the 1.7−2.2 Å range. For both N3 and O2 binding, a close agreement of ΔE(SIBFA) with ΔE(QC) and their contributions is obtained at equilibrium distance (2.0 and 1.8 Å for N3 and O2, respectively) and beyond.

b. Perpendicular Approach to the Heteroatoms. We have considered the 1.7−2.2 Å range of distances. The binding to O2 is the most favored energetically, and is also the one giving rise to the most satisfactory agreements of ΔE(SIBFA) with ΔE(QC). This results, however, from compensations between overestimations of E_1 and underestimations of E_{pol}. The agreements for N1 and N4 are acceptable with errors in total energies limited to 2.5 and 5 kcal/mol for N4 and N1 binding. With SIBFA the shortest distances are more stabilized than in QC, although the equilibrium distance is only 0.1 Å shorter than in QC, because E_{rep}(SIBFA) is underestimated with respect to EX(QC). It is noted that contrary to binding to O2, E_1 is repulsive for the whole range of investigated distances. The binding to N3 is the least satisfactory in the series of heteroatoms, with interaction energies overestimated by 20% in SIBFA. This is due to an underestimation of E_{rep}(SIBFA) which amplifies the trends found with the water probe approaching N3, the numerical values of E_{rep} and EX now being multiplied by a factor of app. 5. As with the water probe, attempts to increase E_{rep} by further increases of the increment of the effective radii of N3 and of its C2 neighbor, while enabling a modest increase of E_{rep} at shorter distances, downgraded the agreement in the case of the Cyt dimers, and had to be discarded.

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

Table 2. Complex of Cytosine with a Water Probe[a]

(b) Perpendicular Approach of Water to N1, N3, N4, and O2						
d	1.9	2	2.1	2.2	2.3	2.4
EX(RVS) N1	13.6	9.6	6.8	4.8	3.4	2.4
E_{rep}(SIBFA) N1	13	9.4	6.9	5	3.6	2.7
EX(RVS) N3	14.7	10.6	7.5	5.4	3.8	2.7
E_{rep}(SIBFA) N3	11.2	8.1	5.9	4.3	3.1	2.3
EX(RVS) N4	14	10	7.1	5	3.6	2.5
E_{rep}(SIBFA) N4	10.2	7.3	5.2	3.7	2.6	1.8
EX(RVS) O2	10.3	7.3	5.1	3.6	2.5	1.8
E_{rep}(SIBFA) O2	6.7	4.8	3.4	2.5	1.8	1.3
EC(RVS) N1	−5.6	−4.3	−3.4	−2.7	−2.2	−1.9
E_{MTP}(SIBFA) N1	−5.3	−4.2	−3.5	−2.9	−2.5	−2.1
EC(RVS) N3	−7.7	−6.1	−4.9	−4	−3.3	−2.8
E_{MTP}(SIBFA) N3	−6.3	−5.1	−4.2	−3.5	−3	−2.5
EC(RVS) N4	−6.6	−5.1	−3.9	−3.1	−2.5	−2
E_{MTP}(SIBFA) N4	−4.6	−3.6	−2.8	−2.3	−1.8	−1.5
EC(RVS) O2	−8.6	−7.1	−5.9	−5	−4.3	−3.7
E_{MTP}(SIBFA) O2	−7.7	−6.5	−5.5	−4.8	−4.1	−3.6
E_1(RVS) N1	8	5.3	3.4	2.1	1.1	0.5
E_1(SIBFA) N1	7.7	5.2	3.4	2.1	1.2	0.5
E_1(RVS) N3	7.1	4.4	2.6	1.4	0.5	−0.1
E_1(SIBFA) N3	4.8	3	1.7	0.8	0.2	−0.2
E_1(RVS) N4	7.5	4.9	3.2	1.9	1.1	0.5
E_1(SIBFA) N4	5.7	3.7	2.4	1.4	0.8	0.4
E_1(RVS) O2	1.7	0.1	−0.8	−1.4	−1.8	−2
E_1(SIBFA) O2	−1	−1.7	−2.1	−2.3	−2.4	−2.3
E_{pol}(RVS) N1	−1.9	−1.5	−1.2	−0.9	−0.7	−0.6
E_{pol}(SIBFA) N1	−1	−0.8	−0.7	−0.6	−0.5	−0.4
E_{pol}(RVS) N3	−2.2	−1.7	−1.3	−1	−0.8	−0.6
E_{pol}(SIBFA) N3	−1.2	−1	−0.8	−0.7	−0.5	−0.5
E_{pol}(RVS) N4	−1.8	−1.3	−1	−0.8	−0.6	−0.5
E_{pol}(SIBFA) N4	−0.9	−0.7	−0.6	−0.5	−0.4	−0.3
E_{pol}(RVS) O2	−2	−1.5	−1.2	−1	−0.8	−0.6
E_{pol}(SIBFA) O2	−1.4	−1.1	−0.9	−0.7	−0.6	−0.5
E_{ct}(RVS) N1	−1.8	−1.3	−0.9	−0.7	−0.5	−0.3
E_{ct}(SIBFA)N1	−1.2	−1	−0.8	−0.6	−0.5	−0.4
E_{ct}(RVS) N3	−2.2	−1.6	−1.2	−0.9	−0.6	−0.5
E_{ct}(SIBFA) N3	−1.1	−0.9	−0.7	−0.6	−0.5	−0.4
E_{ct}(RVS) N4	−1.9	−1.4	−1	−0.7	−0.5	−0.4
E_{ct}(SIBFA) N4	−0.9	−0.7	−0.6	−0.5	−0.4	−0.3
E_{ct}(RVS) O2	−1.6	−1.1	−0.8	−0.6	−0.4	−0.3
E_{ct}(SIBFA) O2	−0.6	−0.5	−0.4	−0.3	−0.3	−0.2
ΔE(RVS) N1	3.9	2.3	1.1	0.3	−0.2	−0.6
ΔE(SIBFA) N1	5.5	3.4	1.9	0.9	0.2	−0.3
ΔE(RVS) N3	2.4	0.7	0.1	−0.8	−1.2	−1.4
ΔE(SIBFA) N3	2.5	1.1	0.2	−0.5	−0.9	−1.1
ΔE(RVS) N4	3.4	1.9	0.9	0.3	−0.2	−0.5
ΔE(SIBFA) N4	3.9	2.3	1.2	0.5	0.1	−0.2
ΔE(RVS) O2	−2.2	−2.8	−3.1	−3.2	−3.2	−3.1
ΔE(SIBFA) O2	−3	−3.3	−3.4	−3.3	−3.2	−3.1

(c) Perpendicular Approach of Water to C6, C5, C4, and C2						
d	1.9	2	2.1	2.2	2.3	2.4
EX(RVS) C6	15.8	11.5	8.4	6.1	4.4	3.2
E_{rep}(SIBFA) C6	12.3	8.9	6.4	4.6	3.4	2.3
EX(RVS) C5	19.3	14.6	10.4	7.6	5.6	4.1
E_{rep}(SIBFA) C5	11.2	8.1	5.9	4.2	3.1	2.2
Ex(RVS) C4	14	10.1	7.3	5.2	3.8	2.7
E_{rep}(SIBFA) C4	10.9	7.9	5.8	4.2	3.1	2.2
EX(RVS) C2	11.7	8.6	5.9	4.2	3	2.1
E_{rep}(SIBFA) C2	13	9.4	6.8	5	3.6	2.7
EC((RVS)C6	−2.8	−2	−1.5	−1.1	−0.8	−0.5
E_{MTP}(SIBFA) C6	−0.6	−0.5	−0.4	−0.3	−0.2	−0.1
EC(RVS) C5	−6	−4.5	−3.2	−2.3	−1.5	−1
E_{MTP}(SIBFA) C5	−4.9	−3.8	−2.9	−2.1	−1.5	−1
EC((RVS)C4	0.2	0.4	0.6	0.6	0.7	0.7
E_{MTP}(SIBFA) C4	0.6	0.4	0.4	0.3	0.3	0.3
EC(RVS) C2	−2.7	−2.6	−2.5	−2.4	−2.3	−2.3
E_{MTP}(SIBFA) C2	−3.1	−3	−2.9	−2.8	−2.7	−2.6
E_1(RVS) C6	13	9.5	6.9	5	3.6	2.7
E_1(SIBFA) C6	11.7	8.3	6	4.3	3.2	2.3
E_1(RVS) C5	13.4	10.1	7.3	5.4	4	3.1
ΔE(SIBFA) C5	6.3	4.3	3	2.1	1.6	1.2
E_1(RVS) C4	14.2	10.5	7.8	5.9	4.4	3.4
E_1(SIBFA) C4	11.5	8.4	6.2	4.6	3.4	2.6
E_1(RVS) C2	9	6.1	3.5	1.8	0.7	−0.2
E_1(SIBFA) C2	9.9	6.4	3.9	2.2	0.9	0
E_{pol}(RVS) C6	−1.9	−1.4	−1.1	−0.9	−0.7	−0.6
E_{pol}(SIBFA) C6	−1.4	−1.1	−0.9	−0.7	−0.6	−0.5
E_{pol}(RVS) C5	−2.3	−1.8	−1.3	−1.1	−0.8	−0.7
E_{pol}(SIBFA) C5	−1.4	−1.1	−0.9	−0.7	−0.6	−0.5
E_{pol}(RVS) C4	−1.9	−1.5	−1.2	−0.9	−0.7	−0.6
E_{pol}(SIBFA) C4	−1.5	−1.2	−1	−0.8	−0.7	−0.6
E_{pol}(RVS) C2	−1.9	−1.5	−1.2	−0.9	−0.7	−0.6
E_{pol}(SIBFA) C2	−1.2	−1	−0.8	−0.7	−0.6	−0.5
E_{ct}(RVS) C6	−2.1	−1.5	−1.1	−0.8	−0.6	−0.5
E_{ct}(SIBFA) C6	−2	−1.6	−1.3	−1	−0.8	−0.6
E_{ct}(RVS) C5	−3	−2.4	−1.7	−1.3	−0.9	−0.7
E_{ct}(SIBFA) C5	−1.6	−1.3	−1.1	−0.9	−0.7	−0.6
E_{ct}(RVS) C4	−1.7	−1.3	−0.9	−0.7	−0.5	−0.4
E_{ct}(SIBFA) C4	−1.6	−1.2	−1	−0.8	−0.6	−0.5
E_{ct}(RVS) C2	−1.4	−1	−0.7	−0.5	−0.4	−0.3
E_{ct}(SIBFA) C2	−1.2	−1	−0.8	−0.6	−0.5	−0.4
ΔE(RVS) C6	8.7	6.2	4.4	3.1	2.2	1.5
ΔE(SIBFA) C6	8.2	5.6	3.8	2.6	1.8	1.2
ΔE(RVS) C5	7.7	5.6	3.9	2.8	2	1.5
ΔE(SIBFA) C5	3.3	1.9	1	0.6	0.3	0.2
ΔE(RVS) C4	10.2	7.5	5.5	4	3	2.2
ΔE(SIBFA) C4	8.4	5.9	4.2	2.9	2.1	1.5
ΔE(RVS) C2	5.4	3.2	1.3	0.1	−0.7	−1.2
ΔE(SIBFA) C2	7.4	4.4	2.3	0.8	−0.2	−0.9

[a]Compared RVS and SIBFA contributions: (a) in-plane binding to N3 and O2; (b) perpendicular binding to N1, N4, N3, and O2; (c) perpendicular binding to C6, C5, C4, and C2. Distances in Ångstrom, energies in kcal/mol.

c. Perpendicular Approach to C6, C5, C4, and C2. E_{rep}(SIBFA) reproduces closely EX(RVS) for the approach to C6 over the whole range of distances. The agreement is satisfactory for C4 at the equilibrium distance of 2.1 Å and beyond. Mg(II) binding to C2 is the sole situation in which, for perpendicular Mg(II) approaches, E_{rep}(SIBFA) is larger than EX(RVS). By contrast, it is significantly underestimated with respect to it for the binding to C5. ΔE(SIBFA) has a close agreement with ΔE(QC) for the C6 complexes and a satisfactory one for the C2 ones, which is also the most stable of the four C-bound complexes. It is noted that even E_{MTP^*}(SIBFA) has difficulty reproducing EC(RVS) for some

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

Table 3. Complex of Cytosine with a Mg(II) Probe[a]

(a) In-Plane Approach to N3 and O2						
d	1.7	1.8	1.9	2	2.1	2.2
EX(RVS) N3	104.8	73	50.9	35.6	24.9	17.4
E_{rep}(SIBFA) N3	132.8	84.7	54.7	35.7	23.5	15.6
EX(RVS) O2	53.2	34.6	22.4	14.5	9.4	6
E_{rep}(SIFA) O2	54.5	35.4	23	15.1	9.9	6.5
EC(RVS) N3	−136.4	−121.5	−109.1	−98.8	−90	−82.5
E_{MTP}(SIBFA) N3	−140.1	−124.4	−111.3	−100.3	−90.9	−82.6
EC(RVS) O2	−128.9	−118.3	−109.3	−101.4	−94.6	−88.5
E_{MTP}(SIBFA) O2	−124.6	−114.3	−105.3	−97.4	−90.4	−84.2
E_1(RVS) N3	−31.6	−48.5	−58.2	−63.2	−65.1	−65.1
E_1(SIBFA) N3	−7.2	−39.7	−56.6	−64.5	−67.3	−67.2
E_1(RVS) O2	−75.7	−83.7	−86.8	−86.9	−85.2	−82.5
E_1(SIBFA) O2	−70.2	−79	−82.3	−82.3	−80.5	−77.7
E_{pol}(RVS) N3	−90.5	−84.8	−79.8	−75.3	−71.1	−67
E_{pol}(SIBFA) N3	−85.9	−81.1	−76	−70.7	−65.4	−60.2
E_{pol}(RVS) O2	−80.4	−73.6	−67.3	−61.4	−55.8	−50.6
E_{pol}(SIBFA) O2	−80.6	−75.6	−70	−64.1	−58.2	−52.6
E_{ct}(RVS) N3	−3	−2.8	−2.6	−2.3	−2	−1.8
E_{ct}(SIBFA) N3	−4.7	−4.3	−3.8	−3.4	−3	−2.6
E_{ct}(RVS) O2	−2.2	−2	−1.8	−1.7	−1.5	−1.4
E_{ct}(SIBFA) O2	−3.1	−2.8	−2.5	−2.2	−2	−1.7
ΔE(RVS) N3	−125.6	−136.6	−141	−141.2	−138.6	−134.3
ΔE(SIBFA) N3	−121.3	−125	−136.4	−138.6	−135.8	−130.1
ΔE(RVS) O2	−158.8	−159.8	−156.3	−150.3	−142.8	−134.7
ΔE(SIBFA) O2	−153.9	−157.4	−154.8	−148.6	−140.7	−132
(b) Perpendicular approach to N1, N3, N4, and O2						
d	1.7	1.8	1.9	2	2.1	2.2
EX(RVS) N1	97.3	66.5	45.4	31	21.2	14.4
E_{rep}(SIBFA) N1	83.3	58.1	40.7	28.6	20.1	14.2
EX(RVS) N3	106.4	74.3	52	36.3	25.3	17.9
E_{rep}(SIBFA) N3	70.7	49.2	34.3	24	16.9	11.7
EX(RVS) N4	102.1	70.5	48.7	33.7	23.3	16.1
E_{rep}(SIBFA) N4	74.4	50.7	34.7	23.8	16.4	11.3
EX(RVS) O2	79.7	54.1	36.7	24.9	16.9	11.4
E_{rep}(SIBFA) O2	49.4	33.3	22.5	15.3	10.5	7.2
E_1(RVS) N1	82.3	56.7	39	26.8	18.3	12.4
E_1(SIBFA) N1	58.1	38.1	24.6	15.3	9.01	4.7
E_1(RVS) N3	40.4	17.4	2.2	−7.7	−14	−17.9
E_1(SIBFA) N3	4.4	−9	−17.2	−22	−24.6	−26
E_1(RVS) N4	77.2	53.1	36.8	25.8	18.4	13.3
E_1(SIBFA) N4	54.8	36.6	24.7	17	12	8.7
E_1(RVS) O2	−24.2	−38.5	−46.6	−50.7	−52.3	−52.2
E_1(SIBFA) O2	−51.7	−57.5	−59.5	−59.2	−57.7	−55.5
E_{pol}(RVS) N1	−112	−103.4	−95.7	−88.5	−81.8	−75.4
E_{pol}(SIBFA) N1	−109.3	−100.9	−92.7	−84.7	−77.2	−70.1
E_{pol}(RVS) N3	−108.5	−100.5	−93.4	−86.9	−80.8	−75.1
E_{pol}(SIBFA) N3	−117	−108.2	−99.4	−90.9	−82.7	−74.9
E_{pol}(RVS) N4	−101.3	−94.3	−88.1	−82.4	−76.9	−71.6
E_{pol}(SIBFA) N4	−94.5	−87.8	−80.9	−74.1	−67.5	−61.3
E_{pol}(RVS) O2	−93.1	−86	−79.6	−73.6	−68	−62.7
E_{pol}(SIBFA) O2	−76.8	−71.3	−65.7	−60.1	−54.8	−49.8
E_{ct}(RVS) N1	*b*	*b*	−2.1	−1.9	−1.8	−1.7
E_{ct}(SIBFA) N1	−1.5	−1.4	−1.3	−1.2	−1.1	−1
E_{ct}(RVS) N3	−3	−2.8	−2.6	−2.4	−2.2	−2.1
E_{ct}(SIBFA) N3	−2.4	−2.3	−2.1	−1.9	−1.7	−1.6
E_{ct}(RVS) N4	−2.3	−2.1	−2	−1.9	−1.8	−1.8
E_{ct}(SIBFA) N4	−1	−0.9	−0.9	−0.8	−0.7	−0.7
E_{ct}(RVS) O2	*b*	−2.8	−2.6	−2.4	−2.4	−2.3
E_{ct}(SIBFA) O2	−2.8	−2.6	−2.4	−2.2	−2	−1.6

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

Table 3. continued

(b) Perpendicular approach to N1, N3, N4, and O2						
ΔE(RVS) N1	−32.6	−49.4	−59.1	−64	−65.6	−65
ΔE(SIBFA) N1	−52.7	−64.2	−69.4	−70.7	−69.3	−66.4
ΔE(RVS) N3	−71.7	−86.4	−94.2	−97.3	−97.4	−95.4
ΔE(SIBFA) N3	−115	−119.4	−118.7	−109	−102.3	−102.4
ΔE(RVS) N4	−26.7	−43.7	−53.6	−58.7	−60.6	−60.4
ΔE(SIBFA) N4	−40.8	−52.2	−57.1	−57.9	−56.3	−53.3
ΔE(RVS) O2	−120.8	−127.7	−129.2	−127.1	−122.9	−117.5
ΔE(SIBFA) O2	−131.1	−131.2	−127.3	−121.3	−114.2	−106.8
(c) Perpendicular approach to C6, C5, C4, and C2						
d	1.8	1.9	2	2.1	2.2	2.3
EX(RVS) C6	66	46.9	33.4	23.8	17	12.13
E_{rep}(SIBFA) C6	65.7	45.7	31.9	22.3	15.6	11
EX(RVS) C5	89.8	64.9	47	34.1	24.7	17.9
E_{rep}(SIBFA) C5	61.7	43	30	20.9	14.7	10.3
EX(RVS) C4	56.4	39.4	27.5	19.2	13.4	9.4
E_{rep}(SIBFA) C4	49.3	34.5	24.3	17.1	12	8.5
EX(RVS) C2	48.5	33.1	22.6	15.4	10.5	7.1
E_{rep}(SIBFA) C2	57.2	39.9	28	19.6	13.8	9.7
E_1(RVS) C6	77.8	58.4	44.5	34.5	27.3	22
E_1(SIBFA) C6	79.3	57.2	42	31.5	24.2	19
E_1(RVS) C5	73.6	53.5	39.4	29.4	22.3	17.3
E_1(SIBFA) C5	40.3	25.8	16.7	10.9	7.4	5
E_1(RVS) C4	61.9	42.9	29.5	19.9	13.1	8.3
E_1(SIBFA) C4	44.4	27.9	16.8	9.2	4.2	0.8
E_1(RVS) C2	30.7	13.4	1.5	−6.7	−12.2	−15.9
E_1(SIBFA) C2	25.6	8	−3.7	−11.5	−16.6	−19.8
E_{pol}(RVS) C6	−107	−100.2	−93.7	−87.6	−81.8	−76.2
E_{pol}(SIBFA) C6	−103.9	−96.8	−89.6	−82.5	−75.5	−68.8
E_{pol}(RVS) C5	−105.9	−99.6	−94	−88.8	−83.9	−79.3
E_{pol}(SIBFA) C5	−101.3	−93.8	−86.3	−79	−71.9	−65.2
E_{pol}(RVS) C4	−116.3	−108.2	−100.6	−93.2	−86.2	−79.5
E_{pol}(SIBFA) C4	−133.9	−122.4	−111.3	−100.7	−90.8	−81.6
E_{pol}(RVS) C2	−107.7	−99.5	−91.8	−84.4	−77.4	−70.9
E_{pol}(SIBFA) C2	−103.8	−93.9	−84.7	−76.2	−68.4	−61.3
E_{ct}(RVS)C6	*b*	*b*	−2.2	*b*	−2.1	*b*
E_{ct}(SIBFA) C6	−1.6	−1.5	−1.4	−1.3	−1.1	−1
E_{ct}(RVS) C5	−2.6	−2.5	−2.4	−2.4	−2.4	−2.5
E_{ct}(SIBFA) C5	−1.4	−1.4	−1.3	−1.2	−1.1	−1
E_{ct}(RVS) C4	−2.3	−2.2	−2.1	−2	−1.9	−1.9
E_{ct}(SIBFA) C4	−1.7	−1.6	−1.5	−1.3	−1.2	−1.1
E_{ct}(RVS) C2	−2.3	−2.3	−2.2	−2	−1.9	−1.8
E_{ct}(SIBFA) C2	−2.1	−1.9	−1.7	−1.6	−1.4	−1.2
ΔE(RVS) C6	−31.9	−44.4	−51.8	−55.6	−57	−56.7
ΔE(SIBFA) C6	−26.2	−41.1	−48.9	−52.2	−52.5	−50.9
ΔE(RVS) C6	−31.9	−44.4	−51.8	−55.6	−57	−56.7
ΔE(SIBFA) C6	−26.2	−41.1	−48.9	−52.2	−52.5	−50.9
ΔE(RVS) C5	−35.4	−49	−57.5	−62.2	−64.4	−64.8
ΔE(SIBFA) C5	−69.7	−76.8	−78.4	−76.5	−72.6	−67.5
ΔE(RVS) C4	−57.1	−67.9	−73.5	−75.6	−75.3	−73.3
ΔE(SIBFA) C4	−79.1	−91.3	−96	−96	−92.8	−87.8
ΔE(RVS) C2	−79.9	−88.8	−92.8	−93.4	−91.8	−88.7
ΔE(SIBFA) C2	−91.1	−97.9	−99.6	−98	−94.3	−89.6

[a]Compared RVS and SIBFA contributions: (a) in-plane binding to N3 and O2; (b) perpendicular binding to N1, N4, N3, and O2; (c) perpendicular binding to C6, C5, C4, and C2. Distances in Ångstrom; energies in kcal/mol. [b]Unconverged.

complexes at shorter distances than <2.1 Å for the C5, C4, and C2 complexes. As a consequence, ΔE(SIBFA) overestimates ΔE(QC) by 20% for the C5 and C4 complexes. The fact that the QC energy minima occur at short Mg−C distances (2.1−2.2 Å) is due to the very large polarization energy of the Cyt ring by the bare dipositive charge of Mg(II).

Thus, overall, considering the perpendicular Mg(II) approach to the Cyt heteroatoms, ΔE(QC) could be well reproduced at, and past equilibrium distances, for N1, N4, and

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

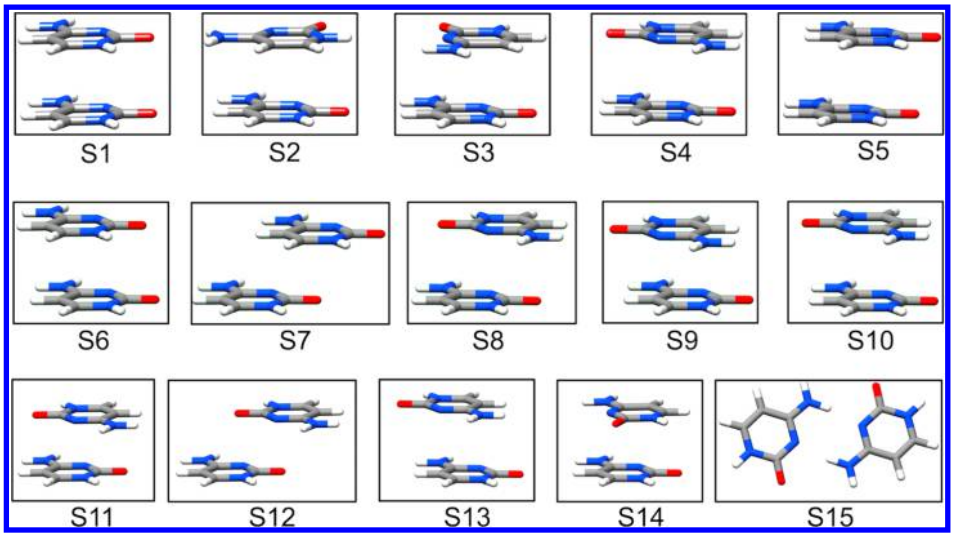


Figure 2. Representation of the structures of the 15 Cyt dimeric complexes.

Table 4. Cytosine Dimers in 14 Stacked Arrangements and in the Doubly H-Bonded Complex[a]

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
Ec/RVS	7	−1.3	−7	−7.8	4	5.1	2.7	−6.3	−8.8	−7	−8	−4.9	−7.1	−7.7	−28.6
E_{MTP}/SIBFA	5.1	−2.9	−9	−10.7	2.9	2.6	2.2	−7.4	−10.2	−8.5	−11	−4.5	−7.6	−9.6	−27.6
	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
E_{exch}/RVS	9.3	10.5	9.8	8.8	9.4	10.1	5.2	6.2	10.4	8.8	9.2	3	6.5	9.6	22
E_{rep}/SIBFA	8.8	10.8	10.9	10.6	10	10.4	4.9	8.5	10.5	9.7	10	4.6	7.6	11.1	23.9
	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
E_1/RVS	16.3	9.2	2.8	1	14.4	15.2	8	−0.1	1.7	1.8	1.2	−1.9	−0.5	1.9	−6.6
E_1/SIBFA	14	7.9	1.9	0	13	13	7.2	1.1	0.3	1.22	−1	0.1	0	1.5	−3.7
	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
E_{pol}(VL)		−1.8	−2.4	−2.4	−1.7	−1.7	−1.3	−2.4	−1.9	−2.1	−2.5	−1.8	−1.3	−2.3	−8.8
E_{pol}/SIBFA	−0.9	−1	−1.5	−1.4	−1	−1	−1	−1.7	−1	−1.2	−1.6	−1.7	−1	−1.5	−8.4
	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
E_{ct}(QC)		−0.7	−0.8	−0.7	−0.6	−0.7	−0.4	−0.5	−0.7	−0.5	−0.6	−0.3	−0.5	−0.8	−3.5
E_{ct}(SIBFA)	0	0	0	0	0	0	−0.1	0	0	−0.1	−0.1	0	−0.1	0	−4.3
BSSE(QC)		−0.5	−0.5	−0.5								−0.3	−0.4		−0.5
	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
ΔE(RVS)	14.4	6.7	−0.4	−2.1	12.1	12.8	6.3	−3	−0.9	−0.9	−1.9	−4	−2.8	−1.2	−18.6
ΔE(SIBFA)	13	6.8	0.4	−1.4	11.9	12	6.3	−0.5	−0.8	0.13	−2.7	−1.6	−0.9	0.12	−16.4

[a]Values of the RVS and SIBFA energy contributions and values of ΔE(RVS) and ΔE(SIBFA). Energies in kcal/mol.

O2 binding. Binding to O2 is the most favorable and also the one giving rise to the best agreement. Binding to N3 gave rise to the poorest agreement. This was unexpected, considering the proximity of O2 and N3. The agreements between ΔE(SIBFA) and ΔE(QC) were also nonuniform concerning the C atoms. Fair agreements at, and past equilibrium distances, were obtained for C6 and C2, but not for C4 and C5. The outcome is unlikely to be modified upon inclusion of correlation effects. Thus, it was earlier shown[77] that the interaction energies of the alkaline-earth cations Mg(II) and Ca(II) underwent little or virtually no variations upon passing from the HF to the MP2 level. It has to be stressed in any case that the present calculations in which a bare dipositive cation probes the heavy atoms of a conjugated and polarizable ring are an extreme test for molecular mechanics. Such binding positions are unlikely to be encountered with such cations. Even if they did, the dipositive charge of the cations would be screened by the field of the other ligands, which, along with the mutual ligand−ligand repulsion, would increase the equilibrium distances, and thus lead to an improved agreement of ΔE(SIBFA) with ΔE(QC). This could be evaluated upon comparing the results for a perpendicular approach of a monovalent cation, Na(I). The results are given in Supporting Information, section S3 and are only shortly commented on here regarding the total energies. A good agreement is retained for the binding over O2, N4, N1, C6, and C2. The errors in total ΔE are reduced at least 3-fold for binding over N3, C5, and C4. The equilibrium distances are increased by 0.3−0.4 Å. Some limitations of present calibration with sole manual fitting could be overcome as a next step upon resorting to the INoLLS automated fitting procedure. Additional parameters to be fit could include the electron population on the individual π lone pairs, with the constraint of a total π electron population of 10, as well as possibly the rise parameters of each π lone pair with respect to its bearer atom. For pragmatic reasons, and as close proximity of Mg(II) cations perpendicular to the cytosine ring is unlikely to occur, we considered the present agreements with the water probe as sufficient to retain the present set of parameters and

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

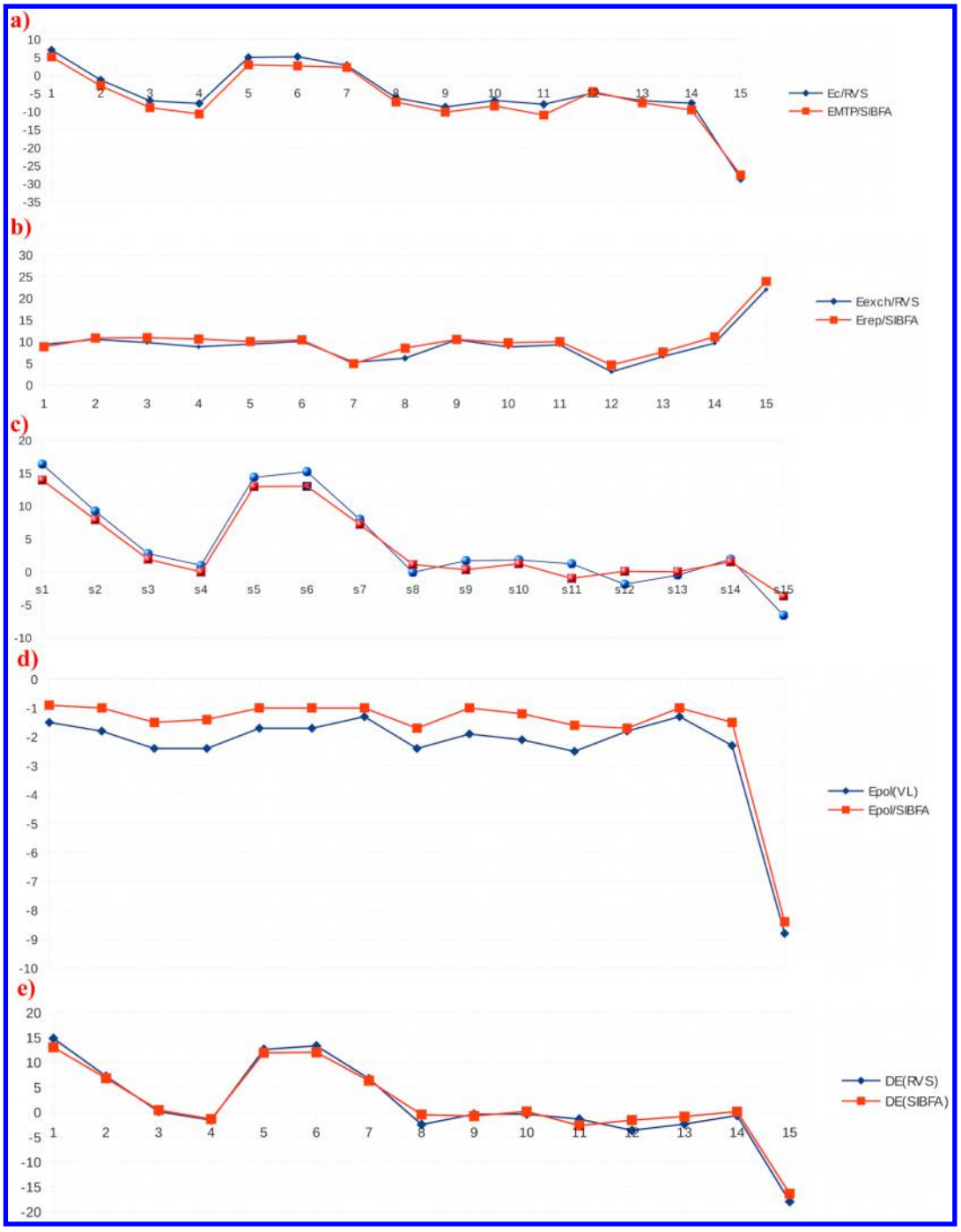


Figure 3. Stacked Cyt dimers. Uncorrelated calculations. Compared evolutions as a function of the conformer number: (a) E_{MTP*} and EC; (b) E_{rep} and EX; (c) E_1(SIBFA) and E_1(RVS) (red, SIBFA; blue, E_1); (d) E_{pol}(SIBFA) and E_{pol}(VR/HF); (e) ΔE(SIBFA) and ΔE(RVS).

evaluate in the next step how it enables ΔE(SIBFA) to perform compared to ΔE(QC). This will be carried out in succession at the uncorrelated (HF/RVS) level, then at the correlated DFT/SAPT and CCSD(T) levels.

HF/RVS Studies of the 15 Cyt Dimers. A representation of the structures of the 15 Cyt dimeric complexes is given in Figure 2. The results are reported in Table 4, comparing the RVS and SIBFA interaction energies and their contributions. Parts a−e of Figure 3 compare their corresponding evolutions along the series of 15 dimers. E_{MTP*} and E_{rep} can match well the evolutions of EC and EX, respectively, for all 15 complexes. Regarding the first-order contributions, some compensations of errors are noted, E_{MTP*} being invariably more attractive or less repulsive than EC for all 14 stacked complexes, while E_{rep} is invariably more repulsive than EX for all dimers. No obvious imbalance is seen for these two contributions compared to the RVS values upon passing from the stacked to the hydrogen-bonded complex. Preserving such a balance could be essential for the prospect of long-duration molecular dynamics of nucleic acids. E_{pol}(SIBFA) can also satisfactorily reproduce the evolution of E_{pol}(VR), but is invariably smaller in magnitude than it. The very large increase of magnitude of E_{pol}(VR) upon passing from the stacked to the hydrogen-bonded complex, namely by a factor of about four, is correctly accounted for by

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

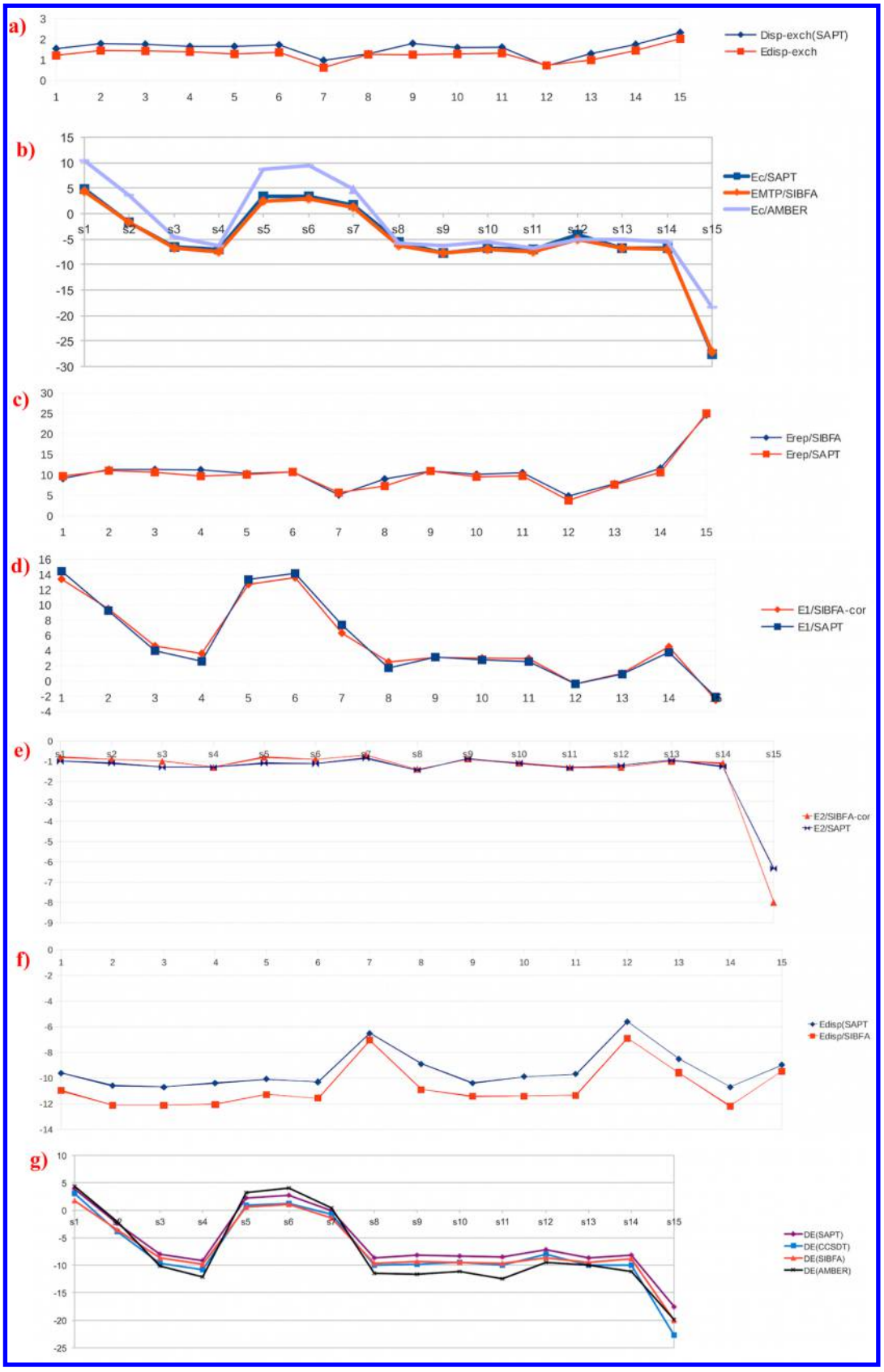


Figure 4. Stacked Cyt dimers. Correlated calculations. Compared evolutions as a function of the conformer number: (a) $E_{\text{disp-exch}}$(SIBFA) and $E_{\text{disp-exch}}$ (SAPT); (b) $E_{\text{MTP*}}$ and EC(SAPT); (c) E_{rep} and EX(SAPT); (d) E_1(SIBFA) and E_1(SAPT); (e) E_2(SIBFA) and E_2(SAPT); (f) E_{disp}(SIBFA) and E_{disp} (SAPT); (g) ΔE(CCSD(T)), ΔE(SAPT), ΔE_{tot}(SIBFA), and ΔE(AMBER).

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

Table 5. Cytosine Dimers in 14 Stacked Arrangements and in the Doubly H-Bonded Complex[a]

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
Ec/SAPT	4.8	−1.8	−6.6	−7.1	3.3	3.4	1.7	−5.6	−7.8	−6.8	−7.1	−4.1	−6.7	−6.8	−27.5
E_{MTP}/SIBFA	4.4	−1.7	−6.7	−7.6	2.4	2.9	1.2	−6.4	−7.8	−7.1	−7.5	−5.2	−6.7	−7	−27.1
	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
E_{rep}/SAPT	9.7	11	10.6	9.7	10.1	10.7	5.7	7.3	10.9	9.5	9.7	3.7	7.6	10.6	25
E_{rep}/SIBFA	9.1	11.3	11.3	11.2	10.3	10.7	5.1	8.97	10.9	10.1	10.5	4.8	7.8	11.6	24.6
	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
E_1/SAPT	14.4	9.3	4	2.6	13.4	14.1	7.4	1.7	3.1	2.8	2.6	−0.4	0.9	3.7	−2.1
E_1/SIBFA-cor	13.4	9.5	4.6	3.6	12.7	13.6	6.3	2.5	3.1	3	3	−0.4	1	4.5	−2.5
	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
E2/SAPT	−1	−1.1	−1.3	−1.3	−1.1	−1.1	−0.9	−1.5	−0.9	−1.1	−1.3	−1.2	−1	−1.3	−6.3
E2/SIBFA-cor	−0.8	−0.9	−1	−1.3	−0.8	−0.9	−0.7	−1.4	−0.9	−1.1	−1.3	−1.3	−1	−1.1	−8
	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
Disp-exch(SAPT)	1.5	1.8	1.8	1.7	1.7	1.7	1	1.3	1.8	1.6	1.6	0.7	1.3	1.7	2.3
Edisp-exch	1.2	1.5	1.4	1.4	1.3	1.4	0.6	1.3	1.3	1.3	1.3	0.7	1	1.5	2
	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
ΔE(CCSDT)	3.1	−3.8	−9.6	−10.8	0.9	1.3	−0.8	−10	−9.8	−9.5	−9.9	−8	−9.9	−9.9	−22.7
ΔE(SIBFA)	1.7	−3.5	−8.6	−9.8	0.6	1.12	−1.42	−9.7	−9.3	−9.5	−9.7	−8.6	−9.5	−8.8	−20
ΔE(AMBER)	4.4	−2.0	−10.1	−12.1	3.2	4	0.5	−11.5	−11.6	−11.2	−12.4	−9.5	−9.9	−11.2	−19.9

[a]Values of the SAPT and correlated SIBFA energy contributions and values of ΔE(SAPT), ΔE(CCSD(T)), and ΔE_{TOT}(SIBFA). Energies in kcal/mol.

E_{pol}(SIBFA). E_{ct}(RVS) has for the stacked complexes values in the range −0.8 to −0.4 kcal/mol, but these are to a large extent compensated by the BSSE correction, which amounts to −0.3 to −0.5 kcal/mol. E_{ct} undergoes a large increase in magnitude upon passing to the doubly H-bonded complex, in which it amounts to −3.5 and −3.0 kcal/mol without and with the BSSE correction, respectively. E_{ct}(SIBFA) has also a relatively large magnitude (−4.3 kcal/mol) in the doubly H-bonded complex, but has negligible values for the stacked complexes. The last feature is because in its formulation[78,79] the only electron-acceptor sites on one Cyt monomer are its X−H bonds, with X denoting a heavy atom, and these are at large distances from the electron-donating lone-pairs of the other Cyt monomer. E_{ct}(SIBFA) embodies dependencies upon the cosines of the two angles formed between the bond connecting the lone-pair donor and its bearer on the one hand, and on the other hand, the bond connecting the bearer to either the X or the H atom of the electron acceptor bond. Such angles are suboptimal in the stacked complexes, but near to optimal in the optimized H-bonded complex as well as for the perpendicular complexes of the probes. In terms of total energies, the match of ΔE(SIBFA) to ΔE(RVS) reflects the one found for the summed first- and second-order contributions. ΔE(SIBFA) can reproduce ΔE(RVS) to within 1.5 kcal/mol, except for complexes s8, s12, s13, and s15, the maximal error being 2.5 kcal/mol for s8.

Correlated Computations of the 15 Cyt Dimers. In order to pass on to the correlated level, the most consistent procedure is to resort to correlated multipoles and polarizabilities to compute E_{MTP*} and E_{pol}. An important issue is then whether any refitting to QC results is necessary. We have for the present study retained the same calibration of the adjustable parameters as at the uncorrelated level regarding these two contributions as well as E_{rep} and E_{ct}. The only contribution that needed calibration was E_{disp}. We have simply retained the value of effective van der Waals (vdW) value of 0.90 Å for oxygen sp^3 lone-pairs, for the sp^2 lone-pairs of O, and the sp lone-pair of N3. All π lone-pairs were given one and the same effective vdW value, namely 1.05 Å. This is reasonable, on account of the larger spatial extension of the π lone-pairs than of the sigma-like ones. This can be seen in Figure 4a, comparing the evolutions of the SIBFA exchange-dispersion term to that of its SAPT counterpart, showing an overall match to within 0.4 kcal/mol.

The values of the SAPT and SIBFA contributions are reported in Table 5, which also lists the values of the total CCSD(T), DFT/SAPT, and SIBFA intermolecular interactions. The evolutions of the energy contributions are reported in Figure 4, parts b−g. Both E_{MTP*} and E_{rep} have improved agreements with respect to their SAPT counterparts as compared to the situation at the uncorrelated level. EC(SAPT/B3LYP) is by 1.5 kcal/mol out of 27 less attractive than EC(RVS/HF) for the doubly H-bonded dimer. There are no clear-cut trends in the stacked dimers, EC(SAPT) being less repulsive than EC(RVS) for dimers s1, s5, s6, and s7 while for the others, it can be either more, or less, attractive than it but by differences less than 1 kcal/mol. E_{MTP*}(SIBFA) can match EC(SAPT) to within 0.8−1.1 kcal/mol in the three least favorable cases, s5, s8, and s12, the agreement being to within 0.5 kcal/mol with the 12 other complexes. EX(SAPT) is invariably larger in magnitude than EX(RVS). This trend is followed by E_{rep}(SIBFA) as well, even though as stressed above, no recalibration had to take place. It is recalled that the only dependency of E_{rep} to electrostatics is, given a pair of interacting atoms I and J, through the prefactors $\mathrm{sqrt}[(1 - Q_I/N_{val}(I))\,(1 - Q_J/N_{val}(J))]$. In this expression, similar to the formulation of E_{disp} given above, Q_I and Q_J denote the electronic populations of I and J, and $N_{val}(I)$ and $N_{val}(J)$ the number of their valence electrons.[80] Namely, the more electron-rich an atom, the more negative its atomic charge, and thus the larger the prefactor, and, conversely, the poorer electronically an atom, the smaller the prefactor. Such an expression was introduced by Claverie[81] in his first formulation of an atom−atom repulsive potential to embody a dependency of the short-range repulsion upon the electron populations of the interacting atoms. The reduction of the error of E_{rep}(SIBFA) with respect to EX(SAPT) is visible in particular for s3 (0.7 kcal/mol out of 10 as compared to 1.2), s4 (1.5 kcal/mol out of 10 as compared to 2.2), s8 (1.6 kcal/mol

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

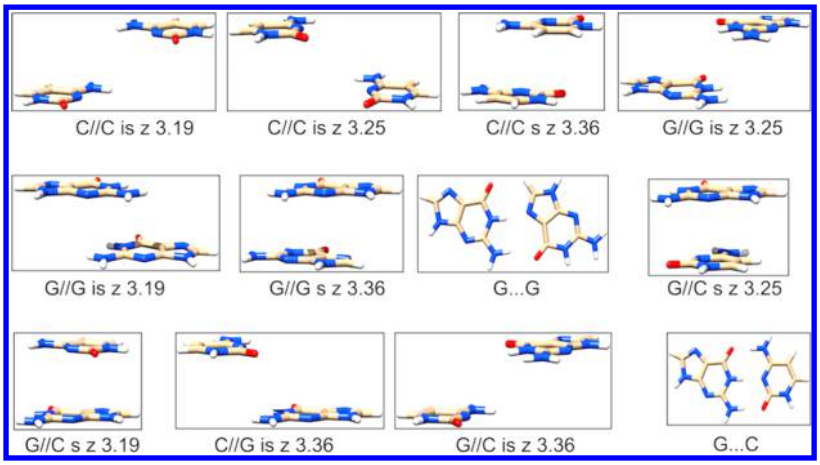


Figure 5. Representative stacked and hydrogen-bonded C-C, G-G, and G-C dimers.

Table 6. Representative Stacked and Hydrogen-Bonded Cytosine and Guanine Dimers[a]

	C//C is z 3.19	C//C is z 3.25	C//C is z 3.36	G//G is z 3.25	G//G is z 3.19	G//G is z 3.36	G··G	G//C is z 3.25	G//C is z 3.19	C//G is z 3.36	G//C is z 3.36	G---C
EC/RVS	2.6	4.9	2.46	5.33	−1.08	1.75	−23.7	−9	−5	−1.76	−4.25	−43.2
E_{MTP}/SIBFA	2.04	3.85	0.98	5.01	−1.58	0.1	−24.9	−9.73	−6.12	−1.81	−4.19	−41.47
Eexch/RVS)	0.18	0.01	5.78	1.03	5.44	7.76	18.16	11.4	6.8	0.65	0.17	40
E_{rep}/SIBFA	0.16	0.01	6.58	0.74	5.52	8.38	18.98	10.89	7.44	0.55	0.11	39.24
E_1/RVS	2.78	4.9	8.24	6.36	4.35	9.51	−5.56	2.38	1.79	−1.11	−4.09	−3.3
E_1/SIBFA	2.19	3.86	7.57	5.76	3.94	8.46	−5.91	1.15	1.32	−1.26	−4.07	−2.23
E_{pol}/VR	−0.21	−0.32	−1.08	−1.24	−0.76	−1.56	−6.39	−1.84	−2.17	−0.81	−0.49	−13.48
E_{pol}/SIBFA	−0.12	−0.23	−0.71	−0.68	−0.62	−0.73	−4.79	−1.03	−1.81	−0.55	−0.38	−11.95
E_{ct}/VR	−0.03	−0.01	−0.38	−0.12	−0.36	−0.56	−2.5	−0.7	−0.48	−0.08	−0.04	−5.79
E_{ct}/SIBFA	−0.02	−0.01	−0.01	−0.1	−0.19	−0.1	−2	−0.13	−0.05	−0.03	−0.02	−4.36
ΔE/RVS	2.58	4.57	6.46	5.36	3.19	8.06	−13.91	0.14	−0.38	−1.83	−4.59	−21.8
ΔE/SIBFA	2.96	3.62	6.78	4.98	3.12	7.63	−12.7	−0.01	−0.54	−1.83	−4.48	−18.55

[a]Values of the RVS and SIBFA energy contributions and values of ΔE(RVS) and ΔE(SIBFA). Distances in Angstrom, energies in kcal/mol.

out of 8 instead of 2.3), s13 (0.3 kcal/mol as compared to 1.1) and the doubly H-bonded complex s15 (0.4 kcal/mol out of 22 instead of 1.9). As a consequence of the improved reproductions by both E_{MTP*} and E_{rep}(SIBFA) of their SAPT counterparts, E_1(SIBFA) can now match E_1(SAPT) with a maximal error of 1.1 kcal/mol (s7), while this error could reach 2.9 kcal/mol at the uncorrelated level for s15.

The sum of E_{pol} and E_{ct}(SIBFA) has for all 14 stacked dimers a relatively flat behavior paralleling well that of E_2(SAPT). It has a larger magnitude (−8 versus −6 kcal/mol) for s15, yet this enables ΔE(SIBFA) to have a larger magnitude than ΔE(SAPT), namely −20.0 as compared to −17.5 kcal/mol, which is closer to the ΔE(CCSD(T)) value of −22.7 kcal/mol (see below). E_{disp}(SIBFA) embodying its exchange-dispersion component has a behavior paralleling closely that of E_{disp}(SAPT), but is invariably larger in magnitude, by 0.5−1.5 kcal/mol.

Finally, in terms of total energies, ΔE(SIBFA) is seen to closely parallel ΔE(CCSD(T)). It has values generally intermediate between those of ΔE(CCSD(T)) and ΔE(SAPT) (Figure 4g).

For completeness, we have also performed AMBER computations on the dimers. The evolution of ΔE(AMBER) compares well with that of ΔE(SAPT), although it can differ from it in some stacked complexes by up to 2 kcal/mol, as with complexes s5, s6, and s11. It is also noted that a lesser agreement is found at the level of the individual AMBER Coulomb contribution (Table 5 and Figure 4b) as compared to EC(SAPT). Thus, the resulting total agreement could stem from a partial compensation of errors at the level of the AMBER van der Waals contributions compared to the short-range and second-order SAPT contributions.

Toward an Extension to Other Base-Pairs. Following the request of a reviewer, we have evaluated whether the present approach could be generalized to other dimers of the nucleic acid bases. As discussed above, a stable calibration of all five bases (including uracil), should resort to the INoLLS software to fine-tune the base-specific parameters. These concern principally the π lone-pair locations to which E_{rep}, E_{ct}, and E_{disp} can be very sensitive. A preliminary extension to guanine is presented below. It is performed on representative stacked and hydrogen-bonded guanine−guanine and guanine-cytosine complexes, previously reported by one of our laboratories.[82] For E_{pol}, we proceeded in a manner analogous to cytosine. This was done by approaching an Mg(II) probe to O6 along the CO bond and to N7 along its external bisector. The guanine-specific screening parameters (multiplicative constant and exponent) of the Gaussian which shields the polarizing field were manually fit to match the values of E_{pol}(RVS) upon varying the Mg(II)−O and Mg(II)−N distances. These values are 0.78 and 1.95, respectively. The relevant data for the π lone-pairs were fit upon considering a fully overlapped stacked guanine dimer, with a rise of 3.4 Å. The second monomer was rotated around the z axis by stepwise variations of 60 degrees.

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

Table 7. Representative Stacked and Hydrogen-Bonded Cytosine and Guanine Dimers. Energies Computed Including Correlation/Dispersion[a]

	C//C is z 3.25	C//C is z 3.19	C//C is z 3.36	G//G is z 3.25	G//G is z 3.19	G//G is z 3.36	G··G	G//C is z 3.25	G//C is z 3.19	C//G is z 3.36	G//C is z 3.36	G---C
ΔE(CBS(T))	3.31	1.41	−1.65	1.42	−3.87	−4.12	−18.71	−11.25	−8.08	−3.29	−4.94	−28.69
ΔE(SIBFA)	3.11	1.24	−1.9	1.43	−4.9	−5.49	−18.36	−11.76	−8.78	−3.57	−4.88	−28
ΔE(AMBER)	4.2	1.99	−0.38	3.76	−4.18	−4.11	−19.11	−12.11	−8.6	−3.6	−5.47	−28.1

[a]Values of ΔE(CCSD(T)), ΔE_{TOT}(SIBFA), and ΔE(AMBER) . Distances in Ångstrom; energies in kcal/mol.

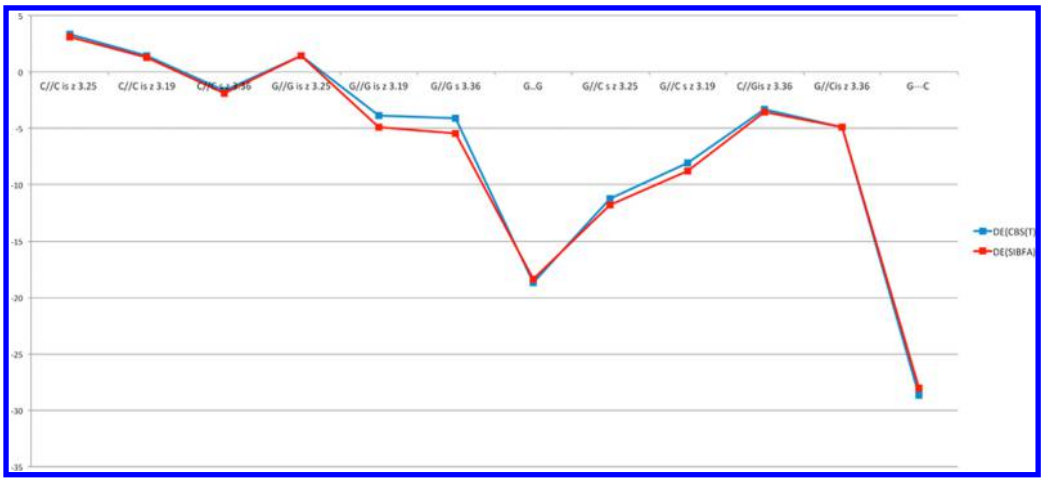


Figure 6. Compared evolutions of ΔE(CCSD(T)) and ΔE_{tot} (SIBFA) in representative stacked and hydrogen-bonded C-C, G-G, and G-C dimers.

The elevation angle θ of each π lone-pair and the increment of van der Waals radius of its individual bearer along its direction were fit in order for E_1(SIBFA) to reproduce E_1(RVS). The results of the fit are reported as Supporting Information, section S4, along with the parameters used for the guanine lone-pairs. The tests subsequently performed bore on three stacked guanine dimers and on the doubly hydrogen-bonded one, and on four stacked G-C dimers and on the Watson−Crick hydrogen-bonded one. The intramolecular geometries of the bases are standard ones and were unrelaxed in the process of energy-minimization in the dimeric complexes. For completeness, three additional stacked C dimers were considered as well. We use the same notations as in ref 82. The double-bars are to denote the stacked complexes. Complexes with bases on the same and on opposite strands are given the "s" and "is" subscripts, respectively, z denoting the rise. The bases are listed following the 5′ → 3′ sequence. The 12 complexes are represented in Figure 5. As for the Cyt dimers, we have first compared the values of ΔE(SIBFA) with HF multipoles and polarizabilities to those of ΔE(RVS). The comparisons bore on the individual contributions as well. The results are reported in Table 6. The overall agreements between ΔE(SIBFA) and ΔE(RVS), as well as between their individual contributions, are comparable to those previously found with the Cyt dimer at the HF level. It is noteworthy to observe that, as with the Cyt dimer, both H-bonded complexes are dominated by the second-order contributions, E_{pol} and E_{ct}. This is the most accented in the triply H-bonded Watson−Crick G-C base-pair. In it, due to the very strong repulsion contribution E_{exch}/E_{rep} opposing the first-order electrostatics E_C/E_{MTP}, E_1 has a modest contribution to ΔE, limited to 12−15% of ΔE. Such dominant weights of E_2 over E_1 were noted in the above-mentioned EFP study on the nucleic acid bases.[34] They were also found in SIBFA/RVS studies performed in ordered, "ice-like" arrangements of water oligomers (n = 12−20 waters),[31] and, more recently, in the complexes of halobenzene derivatives with a G-C base pair of the HIV-1 intasome.[83]

The results at the correlated level are given in Table 7. For conciseness, the QC results are limited to the CCSD(T) extrapolated at the complete basis set limit. A close agreement of ΔE_{tot}(SIBFA) with ΔE(CBS(T)) is observed, with the exception of one interstrand stacked guanine dimer (G//G is, z = 3.19 Å) and one intrastrand complex (G//G s, z = 3.36 Å) for which SIBFA overestimates the QC calculations by 1 and 1.4 kcal/mol, respectively. We report in Figure 6 the compared evolutions of the QC and SIBFA energies, showing the overall parallelism in their evolutions. We can finally note that ΔE(AMBER) has a good overall agreement with the QC results. The relative errors are more accented with the three C//C stacked bases and the first interstrand G//G dimer, for which it amounts to 2.3 kcal/mol. This does not prejudge on the outcome with larger oligomeric complexes, such as G quadruplexes and their cation complexes, where shortcomings of nonpolarizable force-fields were previously noted[82]

CONCLUSIONS AND PERSPECTIVES

APMM procedures have a considerable, yet uncharted, potential, for large-scale simulations of DNA and RNA. Extremely stimulating perspectives for long-duration dynamics and the handling of very large entities will be enabled by the very recent possibility of massively parallelizing all the individual contributions of the energy, including polarization.[84] The calibration and validation of each contribution is essential for safe large-scale prospective simulations. An essential asset to perform these is the availability of QC energy-decomposition procedures at both HF and correlated levels. We have applied these to address a demanding test of APMM. It concerns the interactions between two cytosine bases, namely stacked in 14 structures having extensive overlap, as well as in the doubly hydrogen-bonded complex. Such complexes enable evaluation

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

of how well an APMM procedure could account for the anisotropy of each of its individual contributions upon rotating and translating one base over the other and whether the balance of the individual contributions is preserved upon passing from the stacked complexes to the doubly H-bonded complex. The anisotropy of E_{MTP*} has been previously illustrated and validated by QC analyses for the stacked formamide dimer[31] but had to be further confirmed regarding the Cyt dimers, along with its trends upon passing from the uncorrelated to the correlated level. This was addressed in the present study. Somewhat unexpectedly, an even better agreement of E_{MTP*} with EC was obtained at the correlated level than at the uncorrelated one. A more uncertain issue related to the anisotropy of the other first-order contribution, E_{rep} as compared to EX. It was previously investigated and validated in the case of H-bonded[31] and cation-ligand[41] but was little considered concerning stacked complexes [see, e.g., ref 31]. For such complexes, we have in this study performed a deliberately limited recalibration of E_{rep}. It bore only, for each individual π lone-pair, on the increment of the effective radius of the bearer atom along its direction. This was done by approaching the π lone pair of each of the seven Cyt heavy atoms by an incoming water or Mg(II) probe along the vertical to the Cyt plane passing through the atom. With the water probe satisfactory results were obtained for E_{rep} compared to EX, as well as for ΔE(SIBFA) compared to ΔE(RVS), for all probed atoms except one, namely C5. The probed C5 position remained repulsive, however, in the range of distances explored, 1.9−2.4 Å. The results with the Mg(II) probe were more unequal, with errors of 20% for binding to N3, C4 and C5. Such positions represent nevertheless unrealistic positions of binding of a bare divalent cation and do not correspond to the lowest energy minimum perpendicular to the ring, which is above O2. These shortcomings indicate in any case where improvements could be sought. They would consist, starting from the present calibration, of using INoLLS to relax in a controlled manner some additional parameters, which could involve the π lone pair elevation, and its occupation number with a constraint that the total occupation number remains equal to ten for the π lone pairs. Nevertheless, upon passing to the 15 Cyt dimers, the present calibration enabled a very satisfactory reproduction of the rotational and translational dependencies of EX in the stacked complexes, and, at the correlated level, of E_{disp} as well, since in SIBFA this contribution embodies dependencies upon the lone pairs as well. In fact, the agreement between SIBFA and QC calculations bore on all the energy contributions. For unanticipated reasons, it was even better at the correlated level than at the uncorrelated one. While the reported calibration phase involving two distinct probes (Mg(II) and water) over all cytosine atomic sites can appear cumbersome, it has clearly to be performed once and for all for each of the five NA bases. Parameter optimization prior to validation could further be automated and refined with the help of dedicated procedures such as INoLLS. If indeed confirmed, the transferability of the potential should then enable consideration of all possible homo- and heterodimers and oligomers.

The present study shows that, already in the present state of calibration, a close reproduction of high-level CCSD(T) results could be reached. It stresses again the need to account for the anisotropy of each contribution, which could be illustrated here in the case of EC, EX, and E_{disp}, and the need for a separability of the potential into five contributions, each of which is in turn decomposed into separate terms. Extensions of the present work will bear on dimer complexes between the other bases, extensions to tetramers and beyond, which could also involve metal cations[27,85−87] and the calibration of the conformation parameters of the sugar−phosphate backbone. We hope to report on these in due time.

APPENDIX

Dispersion Contribution

Let I and J denote the pair of interacting atoms/lone pair centroids, W_I and W_J their effective radii, and R_{IJ} their distance of separation. For n having values 6, 8, and 10, A_{dampn} and B_{dampn} are two general, atom-type independent parameters. $L(I, J)$ are pretabulated pair-wise parameters, which are a function of the atomic numbers of I and J. They are given in Supporting Information, Section 1. If I or J is a lone pair centroid, an atomic number of one is adopted for it. Following refs 80 and 88, we first define a damping function:

Given

$$D_{amp(n)} = A_{dampn}*\{[B_{dampn}*(W_I + W_J)/R_{IJ}] - 1\}$$

we can derive:

$$E_{damp(n)} = \exp(-\mathrm{abs}|D_{amp(n)}|)$$

Such an exponential function decreases smoothly when R_{IJ} decreases since $D_{amp}(n)$ increases accordingly (and would tend to infinity for $R_{IJ} \rightarrow 0$). It thus prevents unrealistically large increases of the magnitude of E_{disp}.

We resort to the index n (n = 6, 8, 10) under parentheses to denote the three E_{disp} components

$$E_{disp(n)} = -\sum_{I,J} L(I, J) \times E_{damp(n)}/({Z_{IJ}}^n)$$

with

$$Z_{IJ} = R_{IJ}/2\sqrt{W_I \times W_J}$$

This expression is extended to lone-pairs, which are represented as fictitious atoms with reduced effective vdW radii.

Exchange-Dispersion

We define Q_I and Q_J as the electronic, Mulliken-like charges, of atoms I and J and by $N_{val}(I)$ and $N_{val}(J)$ as the number of their valence electrons. Q_I is the sum of the monopolar charge of I and of half the sums of the monopoles on all bonds emanating from I. We also denote by A_{exchn} and C_{exchn} two other general parameters. Then:

$$K(I, J) = L(I, J) \times [(1 - Q_I/N_{val}(I)) \times (1 - Q_J/N_{val}(J))]$$

$$E_{disp-exch} = -\sum_{I,J} C_{exchn} \times K(I, J) \times \exp(-A_{exchn} \times Z_{IJ})$$

The formulation of exchange-dispersion differs from the previous one[80] by explicitly accounting for the lone pairs, while it was previously limited to atom−atom exchange-dispersion.

Regarding the introduction of the lone pairs under the form of fictitious atoms, there are two important points. Firstly, the increments of the effective radii of the lone pairs, used in the expression of E_{rep}, are now also introduced in the expression of the W effective lone pair radii. We can index by α and β two

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

lone pairs belonging to atoms A and B. Then secondly, the contribution to E_{disp} and $E_{disp\text{-}exch}$ of a lone-pair such as L_α is modulated by a factor 0.5 * occ(L_α), namely half its occupation number. Thus, if occ(L_α) equals 2, as in doubly occupied lone pairs with sp^3 or sp^2 hybridization, this factor amounts to one, that is, as for a "standard" atom. If it is equal to one, as for π sp lone pairs belonging to pyrimidine-like nitrogens, it amounts to 0.5, etc.

The introduction of this factor enables to modulate the contribution of a given lone-pair by its actual occupation number. When E_{disp} and $E_{disp\text{-}exch}$ involve two lone-pairs, such as Lα and Lβ, the multiplicative factor becomes 0.5 × occ(L_α) × 0.5 × occ(L_β).

Apart from these points, the energy formulas are the same for atom−atom, atom−lone pair, and lone pair−lone pair interactions. This implies that effective radii be allocated to the lone pairs along with those on the atom bearers. In the present work, we wished to keep the number of parameters to be varied to a minimum. Thus, the general parameters A_{dampn}, B_{dampn}, A_{exchn}, and C_{exchn} were fit on the linear water dimer, so that $E_{disp\text{-}exch}$ matches its counterpart from SAPT analyses upon performing distance variations over the d(O−H) 1.6-2.2 Å interval. The effective W radii on O and H were as previously, namely 1.438 and 1.20 Å, respectively.[48] A value of W for O_{sp^3} lone pairs of 0.90 Å was found to enable a proper reproduction of the radial decays of SAPT exchange-dispersion and total dispersion plus dispersion-exchange by their SIBFA counterparts, along with the following values for A_{dampn}, B_{dampn}, A_{exchn}, and C_{exchn}: 1.23, 1.37, 9.10, and 120.0. The two values for the damping are those of the previous calibration. While the previous value of A_{exchn} was 9.40, not significantly different from the present one, that of the multiplicative factor C_{exchn} was 408.0, significantly larger than the present value. Its reduction to 120.0 is consistent with the additional inclusion of atom-lone pair and lone pair-lone pair exchange-dispersion terms, which have faster distance variations than the atom−atom ones. With these new parameters, a final multiplicative factor for all regrouped E_{disp} contributions was fit on the linear water dimer. A value of 0.85 instead of the default value of 1.0 was adopted. The radial behaviors of the SIBFA and SAPT dispersion terms for this dimer are compared in Supporting Information, Section 2. We note that the summed E_{disp} in SIBFA is slightly lower than E_{disp} (SAPT), −1.2 kcal/mol at equilibrium distance (d_{O-H} = 1.95 Å) as compared to −1.5 in SAPT. This was to prevent ΔE_{TOT}(SIBFA) from being overestimated at equilibrium. This is because the sum of the other contributions, calibrated to match the QC ones at the HF/augcc-pVTZ(-f) is larger by about 0.3 kcal/mol than the corresponding SAPT sum. With the present calibration, ΔE amounts to −5.2 kcal/mol.

An innovative feature of the AIFF force-field is the inclusion of a three-body non-additive term in E_{disp}, based on the Axilrod−Teller formula.[40] Its AIFF inclusion concerned atomic centers belonging to three distinct molecular fragments. It was found to enable a very important improvement in the accuracy of crystal lattice energies. This is an incentive to include it as well in the SIBFA potential and test it on oligomeric complexes. It would not, however, impact the present results which bear on dimers.

■ ASSOCIATED CONTENT

Supporting Information

Section S1, values of the pretabulated pairwise parameters $L(I,J)$, with I and J denoting atomic numbers; section S2, plot of linear water dimer evolutions, section S3, data for complex of cytosine with a Na(I) probe; section S4; data for a stacked complex of two guanine bases with full overlap. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b01695.

■ AUTHOR INFORMATION

Corresponding Author

*nohad.gresh@parisdescartes.fr (N.G.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We wish to thank the Grand Equipement National de Calcul Intensif (GENCI): Institut du Développement et des Ressources en Informatique Scientifique (IDRIS), Centre Informatique de l'Enseignement Supérieur (CINES), France, Project No. x2009-075009), and the Centre de Ressources Informatiques de Haute Normandie (CRIHAN, Rouen, France), Project 1998053. J.S. and J.E.S. acknowledge support by Grant P208/11/1822 from the Grant Agency of the Czech Republic and institutional funding by project "CEITEC - Central European Institute of Technology" (CZ.1.05/1.1.00/02.0068) from the European Regional Development Fund. J.S. also acknowledges support from Praemium Academiae.

■ REFERENCES

(1) Sponer, J.; Sponer, J. E.; Petrov, A. I.; Leontis, N. B. Quantum Chemical Studies of Nucleic Acids Can We Construct a Bridge to the RNA Structural Biology and Bioinformatics Communities? *J. Phys. Chem. B* **2010**, *114*, 15723−15741.

(2) Drsata, T.; Lankas, F. Theoretical Models of DNA Stability. *WIREs Comput. Mol. Sci.* **2013**, *3*, 355−363.

(3) Sponer, J.; Banas, P.; Jurecka, P.; Zgarbová, M.; Kührová, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M. Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. *J. Phys. Chem. Lett.* **2014**, *5*, 1771−1782.

(4) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second-generation Force-field for the Simulation of Proteins, Nucleic acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(5) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. An All-atom Empirical Energy Function for the Simulation of Nucleic Acids. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.

(6) Foloppe, N.; McKerell, A. D. All-atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21*, 86−104.

(7) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. A., III; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids. Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007**, *92*, 3817−3829.

(8) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. A., III; Jurecka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886−2902.

(9) Zgarbova, M.; Javier Luque, F.; Sponer, J.; Cheatham, T. A., III; Otyepka, M.; Jurecka, P. Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013**, *9*, 2339−2354.

(10) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; MacKerell, A. D., Jr. Impact of 2′-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929−1943.

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

(11) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D., Jr. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348−362.
(12) Mladek, A.; Banas, P.; Jurecka, P.; Otyepka, M.; Zgarbova, M.; Sponer, J. Energies and 2′-Hydroxyl Group Orientations of RNA Backbone Conformations. Benchmark CCSD(T)/CBS Database, Electronic Analysis, and Assessment of DFT Methods and MD Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 463−480.
(13) Sponer, J.; Mladek, A.; Spackova, N.; Cang, X.; Cheatham, T. E., III; Grimme, S. Relative Stability of Different DNA Guanine Quadruplex Stem Topologies Derived Using Large-Scale Quantum-Chemical Computations. *J. Am. Chem. Soc.* **2013**, *135*, 9785−9796.
(14) Savelyev, A.; Mackerell, A. D., Jr. An All- Atom Polarizable Force Field for DNA Based on the Classical Drude Oscillator Model. *J. Comput. Chem.* **2014**, *35*, 1219−1239.
(15) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T., III; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; et al. Systematic Molecular Dynamics Study of Nearest-neighbor Effects on Base Pair and Base Pair Step Conformations and Fluctuations in B-DNA. *Nucleic Acids Res.* **2010**, *38*, 299−313.
(16) Curuksu, J.; Zacharias, M.; Lavery, R.; Zakrzewska, K. Local and Global Effects of Strong DNA Bending Induced During Molecular Dynamics Simulations. *Nucleic Acids Res.* **2009**, *37*, 3766−3773.
(17) Spackova, N.; Cheatham, T. E.; Ryjacek, F. Molecular Dynamics Simulations and Thermodynamics Analysis of DNA-drug Complexes. Minor Groove Binding Between 4′,6-Diamidino-2-phenylindole and DNA Duplexes in Solution. *J. Am. Chem. Soc.* **2003**, *125*, 1759−1769.
(18) Fulle, S.; Gohlke, H. Molecular Recognition of RNA: Challenges for Modeling Interactions and Plasticity. *J. Mol. Recognit.* **2010**, *S23*, 220−231.
(19) Mackerell, A. D., Jr.; Nilsson, L. Molecular Dynamics Simulations of Nucleic Acid-Protein Complexes. *Curr. Opin. Struct. Biol.* **2008**, *18*, 194−199.
(20) Bouvier, B.; Lavery, R. A Free Energy Pathway for the Interaction of the SRY Protein with Its Binding Site on DNA from Atomistic Simulations. *J. Am. Chem. Soc.* **2009**, *131*, 9864−9865.
(21) Sponer, J.; Mladek, A.; Sponer, J. E.; et al. DNA and RNA Sugar-Phosphate Backbone Emerges as the Key Player. An Overview of Quantum-Chemical, Structural Biology and Simulation Studies. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15257−15277.
(22) Baker, C. M.; Anisimov, V. M.; mackerell, A. D., Jr. Development of CHARMM Polarizable Force Field for Nucleic Acid Bases Based on the Classical Drude Oscillator Model. *J. Phys. Chem. B* **2011**, *115*, 580−596.
(23) Sponer, J.; Riley, K. E.; Hobza, P. Nature and Magnitude of Aromatic Stacking of Nucleic Acid Bases 2008. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595−2610.
(24) Reblova, K.; Spackova, N.; Stefl, R.; et al. Non-Watson-Crick Base-pairing and Hydration in RNA Motifs: Molecular Dynamics of 5S RNA Loop E. *Biophys. J.* **2003**, *84*, 3564−3582.
(25) Auffinger, P.; Hashem, Y. Nucleic Acid Solvation: From Outside to Insight. *Curr. Opin. Struct. Biol.* **2007**, *17*, 325−333.
(26) Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. Cations in Charge: Magnesium Ions in RNA Folding and Catalysis. *Curr. Opin. Struct. Biol.* **2012**, *22*, 262−272.
(27) Gkionis, K.; Kruse, H.; Platts, J.; Mladek, A.; Koca, J.; Sponer, J. Ion Binding to Quadruplex DNA Stems. Comparison of MM and QM Descriptions Reveals Sizable Polarization Effects Not Included in Contemporary Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 1326−1340.
(28) Gresh, N. Model, Multiply Hydrogen-Bonded Water Oligomers (N = 3−20). How Closely Can a Separable, ab Initio-Grounded Molecular Mechanics Procedure Reproduce the Results of Supermolecule Quantum Chemistry Computations? *J. Phys. Chem. A* **1997**, *101*, 8680−94.
(29) Sponer, J.; Sponer, J. E.; Mladek, A.; Jurecka, M.; Banas, P.; Otyepka, M. Nature and Magnitude of Aromatic Base Stacking in DNA and RNA: Quantum Chemistry, Molecular Mechanics, and Experiment. *Biopolymers* **2013**, *99*, 978−988.
(30) Banas, P.; Mladek, A.; Otyepka, M.; Zgarbova, M.; Jurecka, M.; Svozil, D.; Lankas, F.; Sponer, J. Can We Accurately Describe the Structure of Adenine Tracts in B-DNA? Reference Quantum-Chemical Computations Reveal Overstabilization of Stacking by Molecular Mechanics. *J. Chem. Theory Comput.* **2012**, *8*, 2448−2460.
(31) Piquemal, J.-P.; Chevreau, H.; Gresh, N. Toward a Separate Reproduction of the Contributions to the Hartree- Fock and DFT Intermolecular Interaction energies by Polarizable Molecular Mechanics with the SIBFA Potential. *J. Chem. Theory Comput.* **2007**, *3*, 824−837.
(32) Lu, Z.; Zhu, N.; Wu, Q.; Zhang, Y. J. Directional Dependence of Hydrogen Bonds. A Density-Functional Based Energy Decomposition and its Implication on Force-Field Development. *J. Chem. Theory Comput.* **2011**, *7*, 4038−4049.
(33) Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. V. Accurate First Principles Model Potentials for intermolecular Interactions. *Annu. Rev. Phys. Chem.* **2013**, *64*, 553−578.
(34) Smith, Q. A.; Gordon, M. S.; Slipchenko, L. V. Effective Fragment Potential Study of the Interaction of the DNA Bases. *J. Phys. Chem. A* **2011**, *115*, 11269−11276.
(35) Giese, T. J.; Panteva, M. T.; Chen, H.; York, D. M. Multipolar Ewald Methods. Applications Using a Quantum Mechanical Force-Field. *J. Chem. Theory Comput.* **2015**, *11*, 451−461.
(36) Fedorov, D. G.; Nagata, T.; Kitaura, K. Exploring Chemistry with the FMO Method. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562−7577.
(37) Gao, J. Design of a Next-Generation Force-Field: the X-Pol Potential. *J. Chem. Theory Comput.* **2007**, *3*, 1890−1900.
(38) Fletcher, T. L.; Davie, S. J.; Popelier, P. L. A. Prediction of Intramolecular Polarization of Aromatic Amino Acids using Kriging Machine Learning. *J. Chem. Theory Comput.* **2014**, *10*, 3708−3719.
(39) Mills, M. J. L.; Popelier, P. L. A. Electrostatic Forces. Formulas for the First Derivatives of a Polarizable, Anisotropic Electrostatic Potential Energy Function Based on Machine Learning. *J. Chem. Theory Comput.* **2014**, *10*, 3840−3856.
(40) Wen, S.; Beran, G. J. O. Accurate Molecular Crystal Lattice Lattice Energies from a Fragment QM/MM Approach with On-the-Fly Ab initio Force Field Parametrization. *J. Chem. Theory Comput.* **2011**, *7*, 3733−3742.
(41) Gresh, N. Energetics of Zn2+ Binding to a Series of Biologically-relevant Ligands. A Molecular Mechanics Investigation Grounded on ab initio SCF Supermolecular Computations. *J. Comput. Chem.* **1995**, *16*, 856−882.
(42) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. J. Chem. Theory Comput. Anisotropic, Polarizable Molecular Mechanics Studies of Inter-, Intramolecular Interactions, and Ligand-macromolecule Complexes. A Bottom-up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960−1986.
(43) Gresh, N.; Sponer, J. E.; Spackova, N.; Leszczynski, J.; Sponer, J. Theoretical Study of Binding of Hydrated Cations Zn(II) and Mg(II) to Guanosine 5′Monophosphate. Towards Polarizable Molecular Mechanics for DNA and RNA. *J. Phys. Chem. B* **2003**, *107*, 8669−8681.
(44) Gresh, N.; Sponer, J. Complexes of Pentahydrated Zn2+ with Guanine, Adenine, and the Guanine-Cytosine and Adenine-Thymine Base-pairs. Structures and Energies Characterized Guanine-Cytosine and Adenine-Thymine Base-Pairs by Polarizable Molecular Mechanics and ab initio Calculations. *J. Phys. Chem. B* **1999**, *103*, 11415−11427.
(45) Jurecka, P.; Sponer, J.; Hobza, P. Potential Energy Surface of the Cytosine Dimer: MP2 Complete Basis Set Limit Interaction Energies, CCSD(T) Correction Term, and Comparison with the AMBER Force Field. *J. Phys. Chem. B* **2004**, *108*, 5466−5471.
(46) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(47) Feller, D. The Role of Databases in Support of Computational Chemistry Calculations. *J. Comput. Chem.* **1996**, *17*, 1571−85.

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

(48) Devereux, M.; Gresh, N.; Piquemal, J.-P.; Meuwly, M. A Supervised Fitting Approach to Force Field Parametrization with Application to the SIBFA Polarizable Force Field. *J. Comput. Chem.* **2014**, *35*, 1577−1591.
(49) Langlet, J.; Claverie, P.; Caron, F.; Boeuve, J. C. Interactions between Nucleic Acid Bases in Hydrogen-bonded and Stacked Configurations: The Role of the Molecular Charge Distribution. *Int. J. Quantum Chem.* **1981**, *20*, 299−338.
(50) Stevens, W. J.; Fink, W. Frozen Fragment Reduced Variational Space Analysis of Hydrogen Bonding Interactions. Application to the Water Dimer. *Chem. Phys. Lett.* **1987**, *139*, 15−22.
(51) Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553−566.
(52) Cammi, R.; Hofmann, H.-J.; Tomasi, J. Decomposition of the Interaction Energy between Metal Cations and Water or Ammonia with Inclusion of Counterpoise Corrections to the Interaction Energy Terms. *Theor. Chem. Acc.* **1989**, *76*, 297−313.
(53) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; et al. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(54) Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.
(55) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Surfaces. *Chem. Rev.* **1994**, *94*, 1887−1930.
(56) Jansen, G.; Heßelmann, A. J. Comment on Using Kohn-Sham Orbitals in Symmetry-Adapted Perturbation Theory to Investigate Intermolecular Interactions. *J. Phys. Chem. A* **2001**, *105*, 11156−11157.
(57) Heßelmann, A.; Jansen, G. First-Order Intermolecular Interaction Energies from Kohn-Sham Orbitals. *Chem. Phys. Lett.* **2002**, *357*, 464−470.
(58) Heßelmann, A.; Jansen, G. Intermolecular Induction and Exchange-Induction Energies from Coupled-Perturbed Kohn-Sham Density Functional Theory. *Chem. Phys. Lett.* **2002**, *362*, 319−325.
(59) Heßelmann, A.; Jansen, G. Intermolecular Dispersion Energies from Time-Dependent Density Functional Theory. *Chem. Phys. Lett.* **2003**, *367*, 778−784.
(60) Heßelmann, A.; Jansen, G. The Helium Dimer Potential from a Combined Density Functional Theory and Symmetry-Adapted Perturbation Theory Approach Using an Exact Exchange Correlation Potential. *Phys. Chem. Chem. Phys.* **2003**, *5*, 5010−5014.
(61) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schutz, M. MOLPRO: A General Purpose Quantum Chemistry Program Package. *Wires Comput. Mol. Sci.* **2012**, *2*, 242−253.
(62) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set Convergence in Correlated Calculations on Ne, N_2, and H_2O. *Chem. Phys. Lett.* **1998**, *286*, 243−252.
(63) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set Convergence of the Energy in Molecular Hartree-Fock Calculations. *Chem. Phys. Lett.* **1999**, *302*, 437−446.
(64) Eichkorn, K.; Treutler, O.; Oehm, H.; Haeser, M.; Ahlrichs, R. Auxiliary Basis Sets to Approximate Coulomb Potentials. *Chem. Phys. Lett.* **1995**, *242*, 652−660.
(65) Weigend, F.; Haeser, M. RI-MP2: First Derivatives and Global Consistency. *Theor. Chem. Acc.* **1997**, *97*, 331−340.
(66) Weigend, F.; Haeser, H.; Patzelt, H.; Ahlrichs, R. RI-MP2: Optimized Auxiliary Basis Sets and Demonstration of Efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143−152.
(67) Piquemal, J.-P.; Gresh, N.; Giessner-Prettre, C. Improved Formulas for the Calculation of the Electrostatic Contribution to Intermolecular Interaction Energy from Multipolar Expansion of the Electronic Distribution. *J. Phys. Chem. A* **2003**, *107*, 10353−10359.
(68) Stone, A. J. Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution. *Chem. Phys. Lett.* **1981**, *83*, 233−239.
(69) Stone, A. J.; Alderton, M. Distributed Multipole Analysis-Methods and Applications. *Mol. Phys.* **1985**, *56*, 1047−1064.
(70) Vigné-Maeder, F.; Claverie, P. The Exact Multicenter Multipolar Part of a Molecular Charge Distribution and its Simplified Representations. *J. Chem. Phys.* **1988**, *88*, 4934−4948.
(71) Garmer, D. R.; Stevens, W. J. Transferability of Molecular Distributed Polarizabilities from a Simple Localized Orbital Based Method. *J. Phys. Chem.* **1989**, *93*, 8263−8270.
(72) Dupuis, M.; Marquez, A.; Davidson, E. R. HONDO 99.6, 1999, based on HONDO 95.3; Dupuis, M.; Marquez, A.; Davidson, E. R. *Quantum Chemistry Program Exchange (QCPE)*; Indiana University: Bloomington, IN, 1999; 47405.
(73) Becke, A. K.; Edgecombe, K. E. A Simple Measure of Electron Localization in Atomic and Molecular Systems. *J. Chem. Phys.* **1990**, *92*, 5397−5403.
(74) Silvi, B.; Savin, A. Classification of Chemical Bonds Based on Topological Analysis of Electron Localization Functions. *Nature* **1994**, *371*, 683−686.
(75) Chaudret, R.; Gresh, N.; Cisneros, A.; Scemama, A.; Piquemal, J.-P. Further Refinements of Next Generation Force Fields. Non-empirical Localization of Off-centered Points in Molecules. *Can. J. Chem.* **2013**, *91*, 804−810.
(76) Law, M. M.; Hutson, J. M. I-nolls: a Program for Interactive Non-linear Least-squares Fit of the Parameters of Physical Models. *Comput. Phys. Commun.* **1997**, *102*, 252−268.
(77) Garmer, D. R.; Gresh, N. A Comprehensive Energy Component Analysis of the Interaction of Hard and Soft Dications with Biological Ligands. *J. Am. Chem. Soc.* **1994**, *116*, 3556−3567.
(78) Gresh, N.; Claverie, P.; Pullman, A. Computations of Intermolecular Interactions. Expansion of a Charge-transfer Energy Contribution in the Framework of an Additive Procedure. Applications to Hydrogen-bonded Systems. *Int. J. Quantum Chem.* **1982**, *22*, 199−215.
(79) Gresh, N.; Claverie, P.; Pullman, A. Intermolecular Interactions. Elaboration on an Additive Procedure Including an Explicit Charge-Transfer Contribution. *Int. J. Quantum Chem.* **1986**, *29*, 101−118.
(80) Creuzet, S.; Langlet, J.; N. Gresh, N. Adjustment of the SIBFA Procedure for Potential Maps to Study Hydrogen-Bonding Vibrational Frequencies. *J. Chim. Phys.* **1991**, *88*, 2399−2409.
(81) Claverie, P. In *Inter-Molecular Interactions. From Diatomics to Biopolymers*; Wiley: New York, 1976.
(82) Sponer, J.; Jurecka, P.; Marchan, I.; Luque, F. J.; Orozco, M.; Hobza, P. Nature of Base Stacking. Reference Quantum-Chemical Stacking Energies in Ten Unique B-DNA Base-Pair Steps. *Chem. - Eur. J.* **2006**, *12*, 2854−2865.
(83) El Hage, K.; Piquemal, J. P.; Hobaika, Z.; Maroun, R.; Gresh, N. Substituent-Modulated Affinities of Halobenzne Derivatives to the HIV-1 Integrase Binsing Site. Analysis of the Interaction Energies by Parallel Quantum Chemistry and Polarizable Molecular Mechanics. *J. Phys. Chem. A* **2014**, *118*, 9772−9782.
(84) Lipparini, F.; Lagardère, L.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: I. Toward Massively Parallel Direct Space Computations. *J. Chem. Theory Comput.* **2014**, *10*, 1638−1651.
(85) Fonsecca Guerra, C.; Zijlstra, H.; Paragi, G.; Bickelhaupt, F. M. Telomere Structure and Stability. Covalency in Hydrogen Bonds, Not Resonance Assistance, Causes Cooperativity in Guanine Quartets. *Chem. - Eur. J.* **2011**, *17*, 12612−12622.
(86) Yurenko, Y. P.; Novotny, J.; Sklenar, V.; Marek, R. Exploring Non-Covalent Interactions in Guanine- and Xanthine-Based Model DNA Quadruplex Structures: a Comprehensive Quantum-Chemical Approach. *Phys. Chem. Chem. Phys.* **2014**, *16*, 2072−2084.
(87) Novotny, J.; Yurenko, Y. P.; Kulhanek, P.; Marek, R. Tailoring the Properties of Quadruplex Nucleobases for Biological and Nanomaterial Applications. *Phys. Chem. Chem. Phys.* **2014**, *16*, 15241−15248.
(88) Gresh, N.; Leboeuf, M.; Salahub, D. R. Energetics and Structure in Neutral, Cationic, and Anionic H-bonded Complexes. A Combined

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495

Ab Initio SCF/MP2 Supermolecular, Density Functional Theory, and Molecular Mechanics Investigation. In *Modeling the Hydrogen Bond*; Smith, D. A., Ed.; ACS Symposium Series No. 569; American Chemical Society: Washington, DC, 1994; pp 82−112.
(89) Piquemal, J.-P.; Giessner-Prettre, C. Unpublished work.

DOI: 10.1021/acs.jpcb.5b01695
J. Phys. Chem. B 2015, 119, 9477−9495