



## Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand–Macromolecule Complexes. A Bottom-Up Strategy

Nohad Gresh,<sup>\*,†</sup> G. Andrés Cisneros,<sup>‡</sup> Thomas A. Darden,<sup>‡</sup> and Jean-Philip Piquemal<sup>\*,§</sup>

*Laboratoire de Pharmacochimie Moléculaire et Cellulaire, U648 INSERM, UFR Biomédicale, Université René-Descartes, 45, rue des Saints-Pères, 75006 Paris, France, Laboratory of Structural Biology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, and Laboratoire de Chimie Théorique, Université Pierre-et-Marie-Curie, UMR 7616 CNRS, case courrier 137, 4, place Jussieu, 75252 Paris, France*

Received June 4, 2007

**Abstract:** We present an overview of the SIBFA polarizable molecular mechanics procedure, which is formulated and calibrated on the basis of quantum chemistry (QC). It embodies nonclassical effects such as electrostatic penetration, exchange-polarization, and charge transfer. We address the issues of anisotropy, nonadditivity, and transferability by performing parallel QC computations on multimolecular complexes. These encompass multiply H-bonded complexes and polycordinated complexes of divalent cations. Recent applications to the docking of inhibitors to Zn-metalloproteins are presented next, namely metallo- $\beta$ -lactamase, phosphomannoisomerase, and the nucleocapsid of the HIV-1 retrovirus. Finally, toward third-generation intermolecular potentials based on density fitting, we present the development of a novel methodology, the Gaussian electrostatic model (GEM), which relies on ab initio-derived fragment electron densities to compute the components of the total interaction energy. As GEM offers the possibility of a continuous electrostatic model going from distributed multipoles to densities, it allows an inclusion of short-range quantum effects in the molecular mechanics energies. The perspectives of an integrated SIBFA/GEM/QM procedure are discussed.

### Introduction

The realm of applications of computational chemistry is considerably expanding owing to steady advances in computer power. This benefits high-level ab initio and DFT quantum chemistry (QC) as well as molecular mechanics (MM) and dynamics (MD). It is anticipated that complexes of many hundreds of thousands of atoms will soon lend themselves to MM/MD simulations. This is a compelling

incentive for refining the interaction energy potential. One example is provided by the docking of competing drugs or inhibitors in the recognition site of a protein or nucleic acid target. The correct ranking of the drugs in terms of their relative affinities depends upon binding energy differences that can be smaller than the relative errors in the interaction energies  $\Delta E_{\text{int}}$ : it is therefore critical to reduce the margins of uncertainty by refining  $\Delta E_{\text{int}}$ . The most sought-after refinement is by explicit addition of a polarization energy contribution,  $E_{\text{pol}}$ , to integrate the principal determinant of nonadditivity. The development of ‘polarizable’ molecular mechanics (PMM) is presently the object of intense efforts worldwide, as attested by the publication of review papers

\* Corresponding author e-mail: nohad.gresh@univ-paris5.fr (N.G.), jpp@lct.jussieu.fr (J.-P.P.).

<sup>†</sup> Université René-Descartes.

<sup>‡</sup> National Institute of Environmental Health Sciences.

<sup>§</sup> Université Pierre-et-Marie-Curie.

on a nearly yearly basis since 2001,<sup>1–5</sup> and the present dedicated volume.

Inclusion of an explicit  $E_{\text{pol}}$  contribution to compute interaction energies between small molecules of biological interest in the gas phase was pioneered in the mid-1960s by Claverie, Rein, and co-workers, giving rise to the so-called ‘monopole-bond polarizability approximation’, which used CNDO/2 derived atomic charges to compute the electrostatic contribution  $E_{\text{el}}$  and the polarizing field.<sup>6</sup> An MM formulation and integration of  $E_{\text{pol}}$  to compute protein–ligand–solvent complexes was due to Warshel and Levitt in the very first implementation of QM/MM methodology.<sup>7a</sup> This work introduced the evaluation of  $E_{\text{pol}}$  in condensed phases taking into account iteratively the interaction between the induced dipoles of all the molecules in the system. The impact of  $E_{\text{pol}}$  on  $\Delta E_{\text{int}}$  can be essential not only in complexes with one or more charged species but also in multiply H-bonded complexes, as exemplified by simulations of water clusters or liquid water (ref 8 and references therein).

$E_{\text{pol}}$  is generally determined by computing induced dipoles with distributed polarizabilities. Although this list is not exhaustive, and apart from SIBFA,<sup>9</sup> this is done by the MOLARIS,<sup>7b</sup> EFP,<sup>10</sup> ORIENT,<sup>11</sup> ASP-W,<sup>12</sup> SDFIII,<sup>13</sup> NEMO,<sup>14,15</sup> OPEP,<sup>16</sup> AMOEBA,<sup>8a,b</sup> AMBER,<sup>17</sup> TCPE,<sup>18</sup> Langlet et al.,<sup>19</sup> and Dang-Chang<sup>20</sup> potentials. The polarizabilities are either scalar or tensor quantities.  $E_{\text{pol}}$  can also be computed in the context of fluctuating charge models<sup>21</sup> or, more recently, using the Drude model.<sup>22</sup> The electrostatic field is screened in several MM potentials. This was done for the first time in ref 7a. SIBFA resorts to a screening by means of a Gaussian damping function. Other potentials resort to a formalism due to Thole<sup>23a,b</sup> or to an alternative Gaussian framework.<sup>23c–e</sup>

At this point it is important to recall that MM refinements have also borne on the other  $\Delta E_{\text{int}}$  MM contributions. The most important ones bore on the electrostatic contribution  $E_{\text{el}}$  upon implementing higher-order distributed multipoles (see refs 9–16 and 19 and ref 5 for discussion). We will denote below by the acronym APMM (anisotropic polarizable molecular mechanics) MM procedures which resort to distributed multipoles to compute  $E_{\text{el}}$ , on account of the strong anisotropy features that they confer to it.

As stressed in our previous review papers<sup>5,24</sup> a molecular mechanics methodology aiming to reproduce QC results should have the following features:

**(1) Separability.** The intermolecular interaction energy  $\Delta E_{\text{int}}$  should be expressed under the form of distinct separate contributions. Each contribution should be formulated and calibrated in order to closely reproduce its QC counterpart obtained from energy-decomposition analyses.<sup>25</sup>

**(2) Anisotropy.**  $\Delta E_{\text{int}}$  and its individual contributions should be able to reproduce the fine angular features of their QC counterparts, upon performing in- and out-of-plane variations in the approach of one molecule to another.

**(3) Nonadditivity.**  $\Delta E_{\text{int}}$  and its individual contributions must be able to mirror the extent of nonadditivities of their QC counterparts upon passing from bi- to multimolecular complexes. In the latter, the total interaction energies can differ substantially from the corresponding summed pairwise

interactions, being either larger or smaller in magnitude, namely in cooperative as opposed to anticooperative complexes respectively.

**(4) Transferability.** The MM potential having been calibrated on a limited training set to reproduce QC results should then be validated on a diversity of bimolecular complexes and then on multimolecular complexes without having to alter the initial calibration. Upon passing to flexible molecules, it should be able to address the issue of multipole transferability that was raised by Faerman and Price.<sup>26</sup>

Separability of  $\Delta E_{\text{int}}$  into five distinct contributions is an essential feature of the SIBFA procedure. In the present review, following the Methods section, we will investigate the extent to which requisites 2–4 above are met. This will be followed by presentations of recent SIBFA applications to molecular recognition problems.

The last section will summarize the recent advances in the development of the Gaussian electrostatic model (GEM), a force field based on density fitting.<sup>27</sup> This method resorts to Hermite Gaussian densities derived from ab initio calculations on molecules or molecular fragments. These densities constitute a continuous electrostatic model connecting distributed multipoles and electron densities.<sup>27</sup> They are used instead of the distributed multipoles at all levels allowing a direct inclusion of short-range quantum effects by means of the computation of electrostatic and repulsion integrals. Thus GEM takes into account nonclassical contributions such as the penetration energy and enables the computation of the main overlap-dependent contribution, namely short-range exchange-repulsion. As the polarization and charge-transfer contributions have been coded in the spirit of SIBFA, the use of such fitted Hermite Gaussian densities<sup>27</sup> should lead to further integration and merging of SIBFA and GEM toward third-generation molecular mechanics potentials.

**Formulation of the SIBFA Procedure.** The SIBFA intermolecular interaction energy is formulated as a sum of five contributions

$$\Delta E_{\text{int}} = E_{\text{MTP}} + E_{\text{rep}} + E_{\text{pol}} + E_{\text{ct}} + E_{\text{disp}} \quad (1)$$

denoting respectively the electrostatic multipolar ( $E_{\text{MTP}}$ ), short-range repulsion ( $E_{\text{rep}}$ ), polarization ( $E_{\text{pol}}$ ), charge-transfer ( $E_{\text{ct}}$ ), and dispersion ( $E_{\text{disp}}$ ) contributions. The analytical forms of these contributions are given in the original papers,<sup>9,28</sup> and we only review here their essential features.

**Electrostatic from Distributed Multipoles.** *Inclusion of Penetration Effects.*  $E_{\text{MTP}}$  is computed with multipoles (up to quadrupoles) that are distributed on the atoms and bond barycenters. They are extracted from the molecular orbitals (MOs) of a given molecule or molecular fragment by a procedure developed by Vigné-Maeder and Claverie.<sup>29</sup> The derivation of distributed multipoles was pioneered in the early 1970s by Dreyfus and Claverie concerning ab initio MOs<sup>30</sup> and by Rein concerning MOs resulting from Iterative Extended Huckel Theory computations.<sup>31</sup> It is useful to recall in the present context that the first applications of ab initio distributed multipoles to compute gas-phase  $\Delta E_{\text{int}}$  in biologically relevant complexes<sup>32</sup> had been published in 1979–1982, where the Dreyfus-Claverie procedure was used. The

methodology was employed on the following molecular recognition problems: the preferential Ca(II) versus Mg(II) binding in 1:2 complexes with the polar head of an anionic phospholipid, phosphatidyl serine;<sup>33</sup> the preferential binding of tetramethylammonium versus monomethylammonium in the binding site of a phosphorylcholine antibody;<sup>34</sup> the binding of nucleic acid bases by amino acid side chains;<sup>35</sup> and cation-selective binding by valinomycin,<sup>36</sup> nonactin,<sup>37</sup> and calcimycin<sup>38</sup> ionophores. In its latest refinements,  $E_{\text{MTP}}$  has been augmented with an explicit penetration term,  $E_{\text{pen}}$ .<sup>39</sup> This was shown to afford for a closer match to the Coulomb contribution,  $E_{\text{C}}$ , which is obtained from energy-decomposition analyses of the ab initio intermolecular interaction energies. Together with the developments by Vigné-Maeder and Claverie,<sup>29</sup> important advances to derive ab initio multipoles from ab initio QC MOs were pioneered in the early 1980s due to contributions of the groups of Stone et al.,<sup>40</sup> Pullman et al.,<sup>41</sup> Sokalski et al.,<sup>42</sup> and Karlstrom et al.<sup>43</sup> An interesting development is the availability on the Web of the OPEP suite of Fortran programs, interfaced to a user-friendly package to derive both distributed multipoles and polarizabilities.<sup>16a</sup> It can be also noted that promising results have been obtained using Bader's Atom in Molecules<sup>16c</sup> approach by Popelier et al.<sup>16b</sup> and as we will discuss latter using density fitting techniques.<sup>27b</sup> However, apart from ref 39, the sole other explicit introduction of  $E_{\text{pen}}$  into a multipoles treatment was within the context of the effective fragment potential (EFP) methodology<sup>44a,b</sup> implemented in GAMESS.<sup>44c</sup>

**Short-Range Exchange-Repulsion.**  $E_{\text{rep}}$  is formulated as a sum of bond–bond, bond–lone pair, and lone pair–lone pair interactions. An  $S^2/R$  representation has been used since 1994<sup>28b–e</sup> following earlier proposals by Murrell and Teixeira-Dias.<sup>45</sup> Here  $S$  denotes an approximation of the overlap between localized MOs (LMOs) of the interacting partners. Hybridization is on chemical bonds as well as on the lone pairs.  $R$  is the distance between the LMO centroids. Following the  $E_{\text{MTP}}$  refinements with inclusion of the  $E_{\text{pen}}$  term,  $E_{\text{rep}}$  is augmented with an  $S^2/R^2$  term.<sup>28e,39b</sup>

**Consistent Treatment of Induction: Polarization, Exchange-Polarization, and Charge-Transfer Energies.** In SIBFA, the induction is equivalent to the HF or DFT  $E_{\text{deloc}}$  contribution (see ref 25e and references therein).

In  $E_{\text{pol}}$  the polarizing field is computed with the same permanent multipoles as  $E_{\text{MTP}}$ . The field is screened by a Gaussian function that depends on the distance between the two interacting centers. Such a screening embodies part of short-range effects including exchange-polarization.<sup>23d</sup> The contribution of the induced dipoles to the field is computed by a self-consistent iterative procedure. Since 1991, the polarizabilities are tensors that are distributed on the bond barycenters and on the heteroatom lone pairs and are derived from the LMOs of the considered molecule or molecular fragment by a procedure due to Garmer and Stevens.<sup>46</sup> As such, both distributed multipoles and polarizabilities can be obtained from one ab initio computation performed on a molecule or constitutive molecular fragment. Each molecular entity is stored in the SIBFA library of fragments and used for subsequent assembly of molecules or molecular com-

plexes. Usually extracted from GAMESS<sup>44c</sup> computations at the HF level, they can also be calculated at the DFT level.<sup>25d</sup>

$E_{\text{ct}}$  is derived from the development of a formula due to Murrell et al.<sup>47</sup> This contribution was explicitly integrated into  $\Delta E_{\text{int}}$  in 1982–1986.<sup>48,28a</sup> A coupling with electrostatics was subsequently introduced.<sup>28b</sup> That is, the ionization potential,  $I_{\text{A}}$ , of the electron donor, on the one hand, and the electron affinity,  $A_{\text{M}}$ , and ‘self-potential’,  $V_{\text{M}}$ , of the electron acceptor, on the other hand, are modified by the electrostatic potential that each undergoes in the complex. These include the effect of the induced dipoles along with those of the permanent multipoles, thereby introducing a coupling with polarization. Such modifications of  $I_{\text{A}}$ ,  $A_{\text{M}}$ , and  $V_{\text{M}}$  were essential to account for the very strong anticooperative character of  $E_{\text{ct}}$  in polycordinated complexes of divalent cations.

To ensure for a correct inclusion of second-order polarization effects, both  $E_{\text{pol}}$  and  $E_{\text{ct}}$  components are fitted upon their RVS<sup>25b</sup> or CSOV<sup>25c–e</sup> counterparts as the two approaches do not violate the Pauli principle conserving antisymmetrized wave functions.<sup>23d,25d</sup> Concerning the polarization, one can compare its first iteration directly to the RVS results. Furthermore, the fully relaxed SIBFA energy can be related to the fully relaxed Morokuma polarization<sup>25a</sup> even though the latter approach does not embody exchange-polarization and can be seen has an upper bound to the polarization energy.<sup>23d</sup>

$$E_{\text{pol}}(\text{SIBFA}) + E_{\text{ct}}(\text{SIBFA}) \sim E_{\text{deloc}}(\text{HF/DFT}) = \Delta E(\text{HF/DFT}) - E_{\text{C}} - E_{\text{exch-rep}}$$

$$E_{\text{pol}}(\text{SIBFA, prior to iterating}) \sim E_{\text{pol}}(\text{RVS/CSOV})$$

**Dispersion and Exchange-Dispersion Components.** Finally,  $E_{\text{disp}}$  is computed as a sum of  $1/R^6$ ,  $1/R^8$ , and  $1/R^{10}$  terms.<sup>49</sup> Directionality effects are accounted for by the introduction of lone-pairs under the form of fictitious atoms. An exchange-dispersion term was also introduced. For H-bonded complexes,  $E_{\text{disp}}$  was initially calibrated on the basis of symmetry-adapted perturbation theory (SAPT)<sup>25f</sup> energy-decomposition analyses.

**Treatment of Flexible Molecules.** A flexible molecule is assembled from its constitutive rigid fragments. Following the procedure published in ref 9, the intramolecular (conformational) energy is computed as the sum of all intermolecular, interfragment interactions, using a formulation related to eq 1. Two successive fragments are connected along X–H and H–Y bonds, where X and Y denote heavy atoms. Conformational changes take place by rotations around junction bond X–Y. The multipoles of the H atoms and of the barycenters of the X–H and H–Y bonds that belonged to the upstream and the downstream fragments, respectively, disappear and are redistributed on three centers: atoms X and Y and the midpoint of the newly formed X–Y bond. SIBFA was originally validated by comparisons with QC in a series of conformational studies of small organic molecules.<sup>9,50</sup> The 1985 paper<sup>50</sup> reported gas-phase conformational studies of the Gly and Ala dipeptides and comparisons with QC computations done in parallel on



representative conformers. Ensuring consistency with gas-phase QC results is a requisite prior to simulations on larger systems and accounting for solvation effects. Reference 50 constituted to our knowledge the very first such study on peptides that used distributed multipoles and polarizabilities. This is worth recalling at this point, in view of the anticipated surge of such studies that should now resort to this kind of approach.

**Calculation of Solvation Energies**  $\Delta G_{\text{solv}}$ .  $\Delta G_{\text{solv}}$  is computed using the Langlet-Claverie (LC)<sup>51a</sup> procedure interfaced in SIBFA.<sup>51b</sup> It is formulated as a sum of electrostatic, polarization, repulsion, dispersion, and cavitation contributions. The electrostatic term is the energy due to the interaction between the electrostatic potential  $V$  created by the distributed multipoles of the solute and a fictitious charge density distributed on the cavity surface  $S$ . The charge density at a given point of  $S$  is a function of the solvent dielectric constant and of the scalar product of the electric field due the solute multipoles and of the unitary vector normal to the surface at that point. The polarization energy of each solute polarizable center is a function of its polarizability and the square of the reaction field created on that center by the charge density. Following the derivation by Huron and Claverie,<sup>51c,d</sup> the repulsion and dispersion terms are computed as sums of repulsion and dispersion energy volume integrals. The sums run on the solute atoms  $i$ , on the one hand, and on the solvent types of atoms  $j$ , on the other hand. The cavitation energy is computed as a sum of contributions from intersecting spheres, centered on the solute atoms. Following a formulation due to Pierotti,<sup>51e</sup> it is a function of a quantity  $d$ , which is the sum of the diameters of the considered atom-centered sphere and of the solvent sphere.

The possibility of constructing large, flexible molecules upon resorting to the multipoles and polarizabilities of their constitutive fragments enabled the addressing of a diversity of molecular recognition problems in 1985–1990. These bore on complexes of DNA with nonintercalating ligands<sup>52</sup> as well as intercalating drugs,<sup>53</sup> complexes of calmodulin central helix with phenothiazine drugs,<sup>54</sup> and selective binding of metal cations and biogenic amines by ionophores.<sup>55</sup> Subsequently, the availability of the restricted variational space analysis (RVS) procedure<sup>25b</sup> was instrumental to enable refinements of the  $E_{\text{rep}}$ ,  $E_{\text{pol}}$ , and  $E_{\text{ct}}$  contributions. Together with the integration of the Langlet-Claverie continuum reaction field procedure to compute  $\Delta G_{\text{solv}}$  using distributed ab initio multipoles, these have in turn enabled performing energy balances for the complexes of inhibitors with Zn-metalloenzymes. This was earlier exemplified in studies of the complexes of thermolysin with mercaptocarboxylate and phosphoramidate inhibitors.<sup>56</sup> The need for a balanced treatment of solvation and interaction energies was emphasized as early as 1976,<sup>7a</sup> and treatments in the context of classical electrostatics encompassing solvent effects were developed by Warshel and co-workers.<sup>7c</sup> The last section of this review paper will summarize some of the most recent applications in this domain.

**Further Refinements.** (a) *Quadrupolar Polarizability and Back-Donation Charge Transfer.* Significant improvements

in the representation of the monovalent Cu(I) cation were as follows:<sup>57</sup> the inclusion of its quadrupolar polarizability (QP) in addition to the dipolar one, to express the additional dependency of Cu(I) polarization energy upon the gradient of the electrostatic field; and the inclusion of charge transfer from Cu(I) to its ligands, in addition to the one taking place from the ligands to the cation.

(b) *Handling of Open-Shell Metal Cations.* Significant progress to represent open-shell metal cations took place in 2003,<sup>58</sup> upon integrating ligand field (LF) effects in SIBFA using an effective Hamiltonian in the framework of the angular overlap model (AOM).<sup>59</sup> The SIBFA-LF procedure was applied to polyligated Cu(II) complexes and was shown to enable close reproductions of QC calculations. An essential result was the preferential stabilization of square-planar arrangements in tetraligated Cu(II) complexes, in marked contrast to the tetrahedral arrangements preferentially stabilized in tetraligated Zn(II) complexes.<sup>28e</sup>

## Results and Discussion

### I. Are the Essential Features of the QC Contributions Reproduced?

Most validation computations reported in this paper have resorted to the CEP 4-31G(2d) basis set due to Stevens et al.<sup>60</sup> This ensures for consistency, since the distributed multipoles and polarizabilities were derived from QC computations on the fragments that used this very basis set. Furthermore, we have observed extremely close correlations between results obtained with this basis set and those obtained with more extended basis sets, such as the 6-311G\*\* or LACV3P\*\* ones. This is illustrated in the present paper in the case of complexes of two Zn-metalloenzymes,  $\beta$ -lactamase and phosphomannoisomerase, with their inhibitors. Thus as commented later in this paper we could observe persistent parallelisms in the evolutions of  $\Delta E(\text{QC})$  as a function of the structure of the competing inhibitor-metalloenzyme model complexes as well as closely similar magnitudes in the CEP 4-31G(2d) versus LACV3P\*\*  $\Delta E(\text{QC})$  values.

(1) *Anisotropy.* The anisotropy features are illustrated below upon monitoring the angular dependencies of QC versus SIBFA energy contributions in two representative examples. The first is the complex of methanethiolate with the Zn(II) cation, and the second is that of carboxylate with water. Methanethiolate is the side chain of deprotonated Cys residues, which constitute an essential Zn-ligating entity in proteins. It is also encountered in the structure of several Zn-metalloenzyme inhibitors. The carboxylate anion is the most ubiquitous anion in biological systems and interacts with a diversity of polar, cationic entities as well as the majority of biologically relevant metal cations. It is therefore essential to evaluate how well the orientation sensitivity of the QC energy and its contributions can be translated by their APM counter parts. In both cases, the Zn–S<sup>−</sup> or the H(w)–O distances of approach are held fixed, and stepwise variations are done on the angle of approach  $\theta = \text{C–S–Zn}$  or  $\text{C–O–H(w)}$ .

(a) *Zn-Methanethiolate.* This complex was previously investigated in the course of the refinements of the SIBFA  $E_{\text{rep}}$ ,  $E_{\text{pol}}$ , and  $E_{\text{ct}}$  contributions.<sup>28b</sup> We report in Supp. Info 1

**Table 1.** Interaction Energies (kcal/mol) in Four Cyclic Water Tetramers<sup>a</sup>

	<i>a</i>		<i>b</i>		<i>c</i>		<i>d</i>	
	ab initio	SIBFA	ab initio	SIBFA	ab initio	SIBFA	ab initio	SIBFA
$E_1$	+2.2	+2.2	-5.1	-5.1	-6.1	-6.6	-6.4	-6.2
$E_{\text{pol}}(\text{RVS})/E_{\text{pol}}^*$	-10.3 (-3.8)	-10.9 (-4.3)	-4.3 (-0.6)	-4.0 (-0.7)	-2.4 (+0.8)	-2.1 (+0.9)	-6.4 (-1.9)	-4.6 (-1.3)
$E_{\text{pol}}(\text{KM})/E_{\text{pol}}$	-14.0 (-7.3)	-15.8 (-7.7)	-5.1 (-0.9)	-5.1 (-1.2)	-2.8 (+0.7)	-2.3 (+0.9)	-5.7 (-1.8)	-6.4 (-1.9)
$E_{\text{ct}}$	-8.0 (-0.6)	-6.6 (-1.3)	-3.5 (0.0)	-3.2 (-0.1)	-2.5 (+0.4)	-2.6 (+0.1)	-3.7 (-0.3)	-3.7 (+0.2)
$\delta E(\text{MP2})/E_{\text{disp}}$	-11.9 (+0.1)	-11.5	-8.7 (0.0)	-7.3	-8.4 (+0.1)	-6.3	-10.2 (+0.4)	-8.5
$\Delta E(\text{MP2})/\Delta E_{\text{tot}}$	-30.0	-31.9	-21.9	-20.7	-19.6	-17.8	-25.9	-23.9

<sup>a</sup> See text for definitions. Nonadditivities are given in parentheses. Negative values indicate cooperativity.

in the Supporting Information the corresponding evolutions in light of the latest refinements.<sup>28c</sup>

(*b*) *Formate-Water*. The angularity features of this complex have been analyzed and reported in a former study<sup>28c</sup> and, regarding the newest SIBFA refinements, for  $E_{\text{rep}}$  in a recent paper.<sup>39b</sup> For completeness, Supp. Info 2 in the Supporting Information displays the corresponding evolutions of the second-order RVS contributions and of their SIBFA counterparts. Both  $E_{\text{pol}}$  and  $E_{\text{ct}}$  now have shallower behaviors than in the methanethiolate-Zn(II) complex.

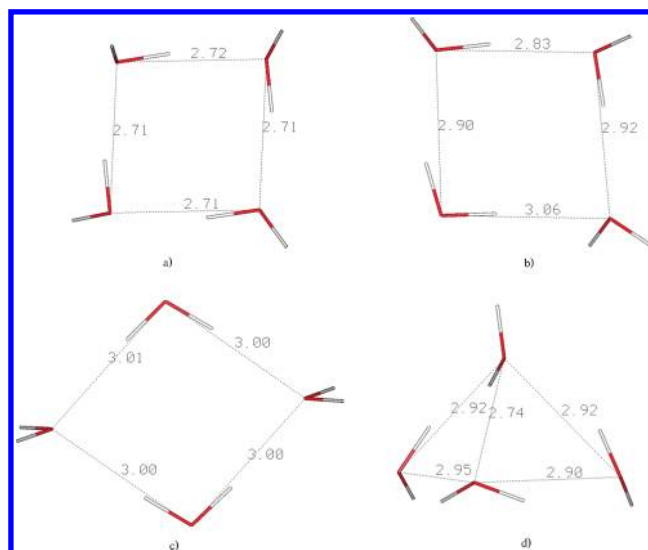
(*c*) *Stacked Formamide Dimer*. This complex is commented on in Supp. Info 3 in the Supporting Information.

The anisotropy features of  $E_{\text{pol}}$  stem from the Garmer-Stevens polarizabilities, which are tensors rather than scalars. Furthermore, heteroatoms are endowed with off-centered lone-pair polarizabilities. The corresponding  $E_{\text{pol}}(\text{lp})$  is maximized when a polarizing center approaches closer to the location of the lone pair centroid. The necessity of off-centered as opposed to atom-centered polarizabilities was recently shown in studies of water-chain complexes designed to maximize the cooperativity response.<sup>23d</sup>

The energy minimizations of the multimolecular complexes reported below used the ‘Merlin’ software.<sup>61</sup>

(2) *Nonadditivity*. In multimolecular complexes, the total interaction energy is not equal to the summed pairwise intermolecular interactions between individual molecules. Thus, the magnitude of  $\Delta E_{\text{int}}$  can be larger than such a sum: *cooperativity* is a feature of the majority of multiply H-bonded complexes or chains. It can, alternatively, be smaller in magnitude than it. *Anticooperative* complexes are mostly encountered in the polycoordinated complexes of a charged species, particularly in the complexes of divalent metal cations. It is critical for polarizable potentials to account equally well for both features. While this has been recognized for a long time, there have been surprisingly few QC analyses of the energy origins of nonadditivity,  $\delta E_{\text{nadd}}$ : i.e., to what an extent could  $\delta E_{\text{nadd}}$  be traced back essentially to the second-order contributions, what are the separate contributions stemming from  $E_{\text{pol}}$  and from  $E_{\text{ct}}$ , and how well could the APM contributions reproduce the nonadditive behaviors of their QC counterparts. RVS energy-decompositions on multimolecular complexes are an invaluable asset for such a quantification.

(*a*) *Cooperativity*. QC and SIBFA studies were performed on multiply hydrogen-bonded water oligomers<sup>62</sup> and models of peptide H-bonded networks.<sup>63</sup> The amounts of QC-computed cooperativities were closely reproduced by SIBFA. RVS analyses showed  $\delta E_{\text{nadd}}$  to originate predominantly from

**Figure 1.** Representation of the four cyclic water tetramers *a*–*d*.

$E_{\text{pol}}$ , while  $E_{\text{ct}}$  contributed little to it, and the SIBFA analyses were fully consistent with the RVS ones.

As an illustration, Table 1 reports a comparison between QC and SIBFA results on four cyclic water tetramers initially designed by Hodges et al.<sup>64</sup> and further considered by Masella et al.<sup>62b</sup> to probe nonadditivity from QC computations and how well these could be translated by polarizable molecular mechanics. These tetramers are represented in Figure 1. In the first, *a*, each water acts in an alternating pattern as an H-bond acceptor to one neighbor and as an H-bond donor to the other. In the second, *b*, one of these waters acts as an H-bond acceptor from both its neighbors, with one of the neighbors acting as an H-bond donor to its own two neighbors. In *c*, two opposite waters act as double H-bond donors, while the two other opposite waters act as double H-bond acceptors. *d* is an alternating three-dimensional arrangement. Table 1 shows a close numerical agreement of QC and SIBFA values in terms of total energies as well as individual contributions, the  $\Delta E(\text{MP2})$  and  $\Delta E(\text{SIBFA})$  ordering being the following:  $a > d > b > c$ . It is instructive to compare the amounts of anticooperativity of  $E_{\text{pol}}$  and  $E_{\text{ct}}$ , as given in parentheses in Table 1.  $E_{\text{pol}}$  is the essential determinant of nonadditivity, consistent with ref 64. Complex *c* is the sole anticooperative complex, with similar QC and SIBFA  $\delta E_{\text{nadd}}$  values.  $E_{\text{pol}}(\text{KM})$  and  $E_{\text{pol}}(\text{RVS})$  denote the values of  $E_{\text{pol}}$  that result from the Kitaura-Morokuma<sup>25a</sup> and the RVS<sup>25b</sup> energy decomposition analyses, respectively.  $E_{\text{pol}}^*$  and  $E_{\text{pol}}$  denote the values of

**Table 2.** RVS and SIBFA Interaction Energies (kcal/mol) in Four 12–20 Water Clusters

	number of waters							
	12		16		16 (MC)		20	
	SIBFA*	RVS	SIBFA*	RVS	SIBFA*	RVS	SIBFA*	RVS
$E_{\text{MTP}}^*/E_{\text{c}}$	−167.6	−168.5	−230.9	−231.4	−179.5	−179.8	−293.2	−294.3
$E_{\text{rep}}^*/E_{\text{exch}}$	151.9	151.4	207.9	207.5	149.8	149.9	263.6	263.2
$E_1$	−15.8	−17.1	−23.1	−23.9	−29.7	−29.9	−30.6	−31.1
$E_{\text{pol}}^*/E_{\text{pol}} \text{ RVS}$	−30.6	−34.7	−42.0	−47.8	−32.7	−35.5		
$E_{\text{pol}}/E_{\text{pol}}$	−41.3	−44.7	−56.5	−61.7	−44.1	−45.1	−71.3	−78.6
$E_{\text{ct}}$	−22.1	−23.1	−30.2	−31.3	−22.6	−23.1	−37.3	−39.4
$\Delta E(\text{SIBFA})/\Delta E(\text{RVS})$	<b>−79.2</b>	<b>−80.1</b>	<b>−109.8</b>	<b>−110.4</b>	<b>−96.4</b>	<b>−94.8</b>	<b>−139.2</b>	<b>−139.1</b>

SIBFA polarization in which the polarizing field is computed with the sole permanent multipoles and with the permanent + induced dipoles, respectively. As discussed in ref 62,  $E_{\text{pol}}^*(\text{SIBFA})$  has close numerical values to  $E_{\text{pol}}(\text{RVS})$ , and  $E_{\text{pol}}(\text{SIBFA})$  has values close to  $E_{\text{pol}}(\text{KM})$ . Such agreement also carries over to the corresponding  $\delta E_{\text{nadd}}$  values.  $E_{\text{ct}}$  is weakly nonadditive, its  $\delta E_{\text{nadd}}$  values being the largest in absolute magnitude for the most strongly bound tetramer *a*.

Table 2 reports the results of parallel RVS and SIBFA computations on four 12–20 water clusters.<sup>39b</sup> It is instructive to re-emphasize the impact of second-order terms in such complexes. Complexes *a*, *b*, and *d* are three-dimensional aggregates in three-dimensional cubic arrangements having 12, 16, and 20 water molecules, respectively, and complex *c* is a small aggregate extracted from an ongoing Monte Carlo (MC) simulation on a water box of  $n = 64$  molecules. The numerical values of  $E_{\text{pol}}(\text{SIBFA})$  outweigh those of the summed first-order contributions  $E_1$ , for which the large stabilizing values of  $E_{\text{MTP}}$  are strongly opposed by those of  $E_{\text{rep}}$ , on account of the shortening of the O–O H-bonding distances (in the 2.7–2.9 Å range) due to cooperativity. In fact, for all three cubic arrangements, *a*, *b*, and *d*, even  $E_{\text{ct}}(\text{SIBFA})$  has larger absolute values than  $E_1$ . All these trends are found in the RVS computations. For all four complexes,  $\Delta E(\text{SIBFA})$  reproduces  $\Delta E(\text{RVS})$  with a relative error <2%. As in Table 1 above, a close correspondence is seen between  $E_{\text{pol}}(\text{RVS})$  and  $E_{\text{pol}}^*(\text{SIBFA})$ , on the one hand, and  $E_{\text{pol}}(\text{KM})$  and  $E_{\text{pol}}(\text{SIBFA})$ , on the other hand.  $E_{\text{pol}}(\text{KM})/E_{\text{pol}}(\text{SIBFA})$  have larger magnitudes than  $E_{\text{pol}}(\text{RVS})/E_{\text{pol}}^*(\text{SIBFA})$ , a signature for cooperativity.

(*b*) *Anticooperativity*. The first concurrent RVS and SIBFA computations on polycordinated cation complexes were performed in the course of SIBFA refinements and bore on polyhydrated complexes of Zn(II), Mg(II), Ca(II), and Cd(II).<sup>65</sup> These were followed by studies on polycordinated Zn(II) complexes in ‘hard’ and ‘soft’ binding protein binding sites<sup>56b,66a</sup> as well as in Zn(II)-metalloenzyme sites including different inhibitor anionic moieties.<sup>56b</sup> The presence of two anions in these sites resulted in very large increases of the magnitudes of  $\Delta E$  and its contributions. The SIBFA computations were nevertheless able to closely reproduce the QC  $\Delta E$  values, in terms of both the total energies and their individual contributions. Subsequent analyses of anticooperativity were done on complexes of formate with penta- and hexahydrated Zn(II) complexes<sup>66b</sup> and on the above-mentioned polycordinated Zn(II) complexes.<sup>65,66a</sup> In these

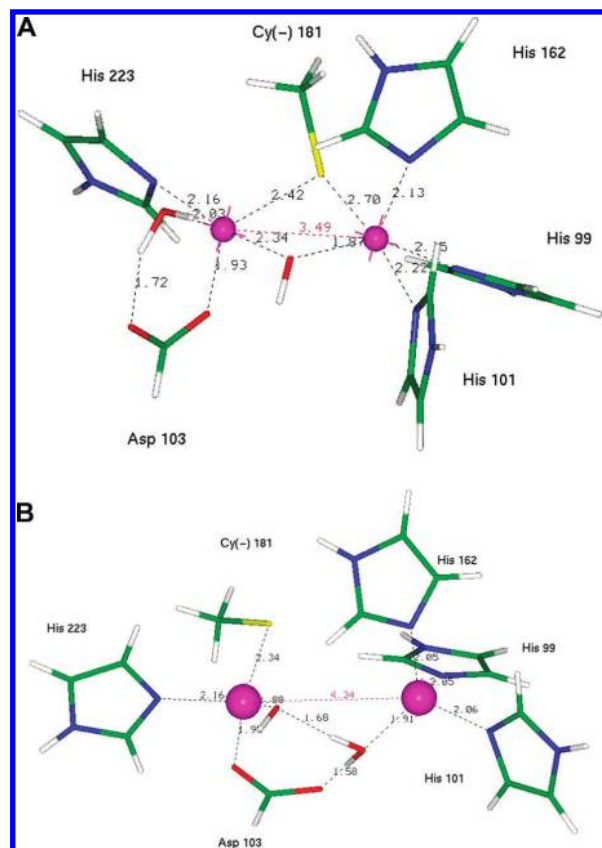
studies the values of QC and SIBFA  $E_1$ ,  $E_{\text{pol}}$ , and  $E_{\text{ct}}$  were compared to their summed values in the separate pairwise complexes that make up the multimolecular complexes. While  $E_1$  showed very little nonadditivity,  $E_{\text{pol}}$  and mostly so  $E_{\text{ct}}$  were strongly anticooperative. It was observed that  $E_{\text{pol}}(\text{SIBFA})$  reproduced well the anticooperativity of  $E_{\text{pol}}(\text{RVS})$ , while  $E_{\text{ct}}(\text{SIBFA})$  somewhat overestimated that of  $E_{\text{ct}}(\text{RVS})$ , particularly upon accumulation of negatively charged ligands (up to four) in the first Zn(II) coordination shell. The anticooperativity of  $E_{\text{ct}}(\text{SIBFA})$  could be reduced by a very simple concerted change of Zn-parameters to allow for the best match to  $E_{\text{ct}}(\text{RVS})$  upon passing from the monoligated  $[\text{Zn}-\text{H}_2\text{O}]^{2+}$  complex to the hexaligated  $[\text{Zn}(\text{H}_2\text{O})_6]^{2+}$  one (see ref 28e for details). As compared to ref 66a, this then resulted in a notably closer agreement of  $E_{\text{ct}}(\text{SIBFA})$  values to the  $E_{\text{ct}}(\text{RVS})$  ones in the representative complexes of Zn(II) with three and four methanethiolate ligands.<sup>28e</sup> This leaves open the issue of the nonadditivity of the contribution of correlation to  $\Delta E$ ,  $\delta \Delta E_{\text{corr}}(\text{MP2})$ , in polycordinated Zn(II) complexes, while in contrast  $E_{\text{disp}}(\text{SIBFA})$  is purely additive. Inclusion of triple-dipole interactions<sup>67</sup> could be considered in future studies to endow  $E_{\text{disp}}(\text{SIBFA})$  with nonadditivity.

The correspondence between QC and SIBFA computations is illustrated below in two examples. The first is that of Zn(II) complexes with six water molecules, and the second is a binuclear Zn(II) complex with a metallo- $\beta$ -lactamase binding site.

In Supp. Info 4 in the Supporting Information are reported the results of parallel QC and SIBFA computations that bore on three competing complexes of Zn(II) with six water molecules.

*Binuclear Zn(II) Binding Sites*. These sites constitute stringent tests for APMM procedures because dramatic enhancements of nonadditivity can be expected. This is due to the proximity of the two cations (in the 3–4.5 Å range) and to the buildup of charged and highly polarizable ligands. Previously investigated complexes<sup>66a,28e</sup> bore on models of Gal4, a binuclear Zn-finger with six cysteinate residues, and on Zn(II)-metallo- $\beta$ -lactamase, an enzyme responsible for the acquired resistance of bacteria to antibiotics. High-resolution X-ray diffraction studies on the *B. fragilis* strain<sup>68</sup> showed the first Zn(II) to be ligated by three His side chains and a hydroxy anion, while the second was ligated by three anionic residues: the hydroxy, an aspartate, and a cysteinate as well as by one His side chain and a water molecule.





**Figure 2.** Representation of the complexes with two Zn(II) cations in the binding site of metallo- $\beta$ -lactamase at the Zn–Zn distances of (a) 3.5 Å and (b) 4.3 Å. Reprinted with permission from Gresh et al. *Journal of Computational Chemistry* **2005**, 26, 1113. Copyright 2005 John Wiley.

Starting from the X-ray structure, SIBFA energy minimizations were performed, after constraining the Zn–Zn distances at 3.0, 3.5, and 3.8 Å (structures *a–c*). QC energy minimizations were subsequently performed starting from the SIBFA minima. While these confirmed the shallow dependence of  $\Delta E$  upon the Zn–Zn distance that was found by SIBFA, they also derived an alternative minimum, denoted as *d*, with the two Zn cations now at  $>4$  Å; the His-bound Zn(II) is now bound to water instead of hydroxy, as a consequence of proton transfer that took place during QC energy minimization. The other cation is now bound to all three anionic ligands and to one His side chain.<sup>69</sup> Complex *d* was reprocessed and energy-minimized using SIBFA and standard internal SIBFA fragment coordinates.<sup>28c</sup> Complexes *b* and *d* are represented in parts a and b, respectively, of Figure 2. The results of concurrent parallel SIBFA computations and RVS analyses at the SIBFA minima are reported in Table 3. In keeping with the results from the previous HF energy minimizations, the RVS analysis shows complexes *b* and *d* to have very close  $\Delta E$  values, differing by 6 out of 1200 kcal/mol, namely less than 1%. Such a small difference is due to compensations of large energy differences between individual contributions. Thus  $E_l$  favors *b* over *d* by a large amount (57 kcal/mol), while both  $E_{pol}$  and  $E_{ct}$  favor *d* over *b* by a total of 64 kcal/mol. The SIBFA computations have very close agreements with the RVS ones. These concern the numerical values of the total energies as well as of their

individual contributions, the opposed trends of first- versus second-order contributions, and the  $d > b$  energy ordering. Such trends remain the same if the LACV3P\*\* basis set<sup>70</sup> is used instead of the CEP 4-31G(2d) one as well as upon going to correlated levels, namely, DFT, LMP2,<sup>71</sup> or MP2.

(3) *Transferability. Interactions Involving Flexible Molecules.* There are several aspects to transferability. The first is the need for a molecular mechanics potential to be applied on a diversity of complexes other than the ‘training set’ on which it was initially calibrated. The separability feature of an APM potential, whose individual contributions are each formulated on the basis of quantum chemistry, should, if their formulations are correct, ensure such transferability. Thus, e.g., if water is properly calibrated on the basis of a limited training set of water dimer complexes, it should be possible to subsequently investigate not only all possible water dimer complexes but also water oligomers of virtually any size as well. Extension of the calibration to any other chemical entity should enable the investigation of all possible complexes that involve this entity in combination with all other ones present in the library. In SIBFA, such ‘entities’ are the constitutive molecular fragments with their internal geometries and distributed multipoles and polarizabilities, which are stored in a library of fragments. Another aspect of transferability relates to the recurrence of well-defined atomic ‘species’ within the molecular fragments. Each atom is identified according to its hybridization state, the number and nature of its neighbors, and the net charge and type of fragment to which it belongs. As an example, O atoms can be assigned as belonging to a hydroxyl or ether-like group, to a carbonyl, a carboxylate, a phosphate, or to a methoxy group, etc. According to its class, a given O is given effective radii for  $E_{rep}$ ,  $E_{pol}$ ,  $E_{ct}$ , and  $E_{disp}$ . These radii are calibrated once and for all to reproduce the radial behavior of the corresponding RVS contribution on a model bimolecular complex. There is a third aspect to transferability that is critical to handling flexible molecules of arbitrarily large size, ranging from pharmacologically relevant ligands up to macromolecules. Such molecules are assembled from their constitutive fragments given the knowledge of the sequence, the length of the junction bond, and the torsion angle along that bond. The multipoles are redistributed along the junction bond following a procedure published in ref 9. This gives rise to the following issue: what is the loss of accuracy due to assembling. That is, is it possible to account in terms of interaction energies for the fact that the multipoles on the fragments undergo changes in their intensities upon integration in a large molecule? With the increase of computer power, it becomes now possible to perform an ab initio computation on large molecular entities of 200 atoms and more and derive their distributed multipoles and polarizabilities. Denoting by A–B a saturated chemical bond between heavy atoms A and B, a large molecule can be subsequently split into smaller fragments by breaking bond A–B and replacing it by two junction bonds A–H\* and H\*–B, with two fictitious hydrogen atoms H\* having null multipoles along the A–B direction, the A–B distance being the same as in bond A–B. This enables for rotations around A–B of the two newly created fragments. How then to energetically

**Table 3.** Interaction Energies (kcal/mol) in the  $\beta$ -Lactamase Binding Sites<sup>a</sup>

	<i>a</i>		<i>b</i>		<i>c</i>		<i>d</i>	
	ab initio	SIBFA	ab initio	SIBFA	ab initio	SIBFA	ab initio	SIBFA
$E_0/E_{MTP}$	-1351.8	-1373.4	-1346.3	-1367.1	-1330.7	-1364.7	-1321.0	-1345.4
$E_{\text{exch}}/E_{\text{rep}}$	362.3	393.9	344.3	370.0	350.4	390.5	375.9	398.8
$E_1$	<b>-989.5</b>	<b>-979.5</b>	<b>-1002.0</b>	<b>-996.2</b>	<b>-980.4</b>	<b>-974.2</b>	<b>-945.1</b>	<b>-946.6</b>
$E_{\text{pol}}(\text{RVS})/E_{\text{pol}}^*$	-223.9	-224.9	-203.3	-202.5	-209.9	-216.6	-252.9	-250.2
$E_{\text{pol}}(\text{HF})/E_{\text{pol}}$	-184.9	-165.7	-173.6	-152.4	-185.6	-172.9	-216.9	-199.2
$E_{\text{pol}}(\text{Zn}(\text{II}))$	-6.1	-3.7	-6.0	-3.6	-7.8	-5.4	-8.0	-3.4
$E_{\text{ct}}$	-56.8	-65.5	-57.2	-66.0	-60.9	-61.7	-75.2	-70.6
$E_{\text{ct}}^*$	-35.7		-36.5		-40.1		-56.3	
BSSE	-21.1		-20.7		-20.8		-19.0	
$\Delta E$	<b>-1210.2</b>	<b>-1207.0</b>	<b>-1212.1</b>	<b>-1211.9</b>	<b>-1206.0</b>	<b>-1203.4</b>	<b>-1218.8</b>	<b>-1213.0</b>
$\Delta E(\text{MP2})/\Delta E_{\text{tot}}$	<b>-1327.6</b>	<b>-1324.0</b>	<b>-1324.3</b>	<b>-1323.5</b>	<b>-1313.5</b>	<b>-1311.2</b>	<b>-1325.7</b>	<b>-1325.1</b>
$\delta E(\text{MP2})/E_{\text{disp}}$	-117.4	-116.1	-112.2	-110.8	-107.5	-107.1	-106.9	-111.6
$\Delta E(\text{HF}/\text{LACV3P}^{**})$	-1241.0		-1242.6		-1237.4		-1248.3	
$\Delta E(\text{LMP2})$	-1270.5		-1270.6		-1270.2		-1272.5	
$\delta E(\text{LMP2})$	-29.5		-28.0		-32.8		-24.2	
$\Delta E(\text{B3LYP}/\text{LACV3P}^{**})$	-1292.1		-1292.7		-1284.9		-1296.6	

<sup>a</sup> *a*–*c*: standard complexes from the *B. fragilis* binding site; *d*: complex derived from HF energy minimization. In *a*–*c*, the Zn–Zn distances are 3.0, 3.5, and 3.8 Å, respectively. In *d*, the Zn–Zn distance is 4.3 Å. The electrostatic potential used in the computation of  $E_{\text{ct}}$  is computed with a full multipolar expansion and with the induced dipoles. In ref 28e, it was mistakenly limited to the sole monopoles.

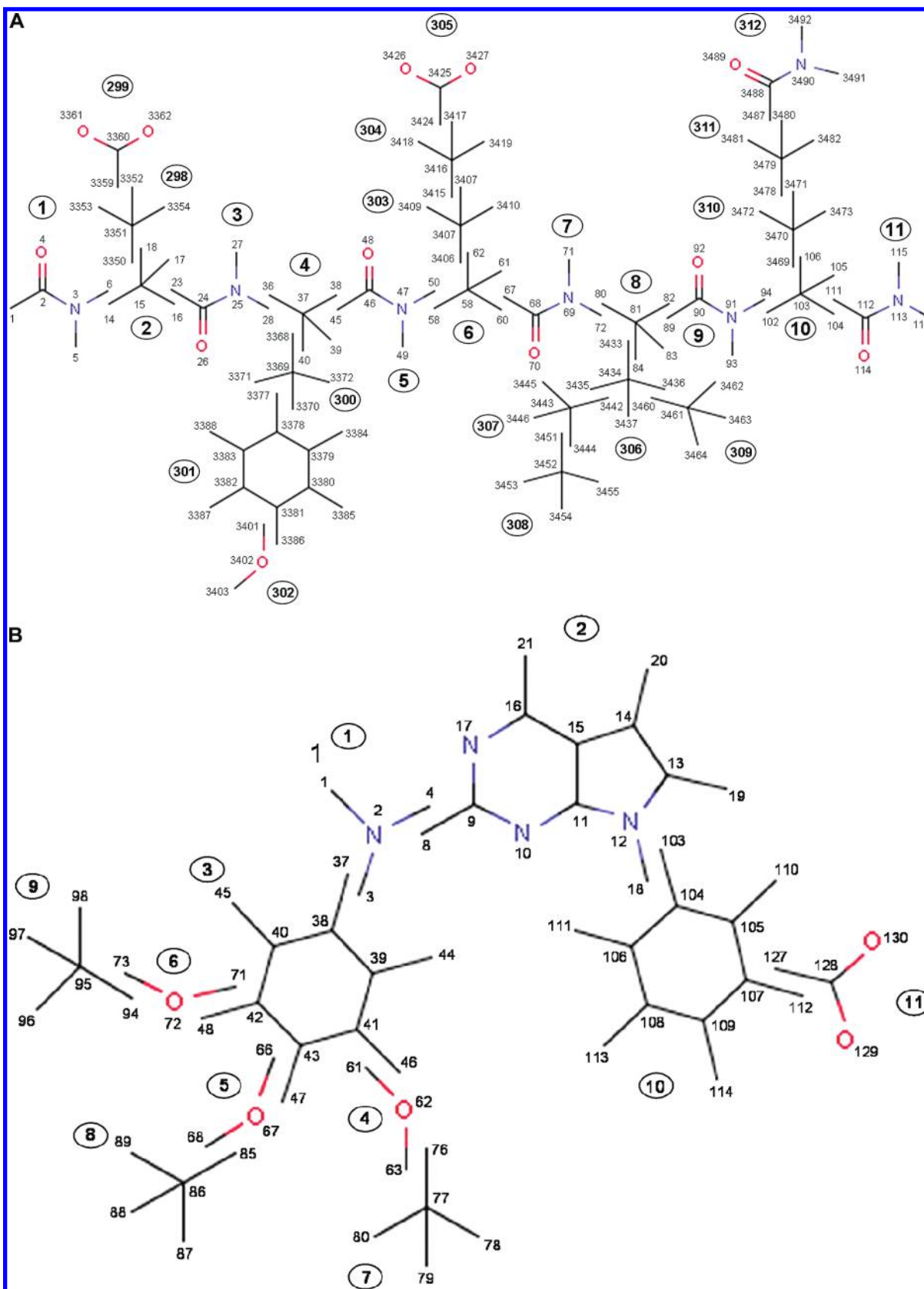
account for the fact that the multipoles on the fragments undergo changes in their intensities upon conformational changes so as not to bias any particular set of conformers. Such an issue was raised for the first time by Faerman and Price<sup>26</sup> upon constructing oligopeptides from the multipoles of their constitutive fragments.

In perturbation or variation theories, the impact of changes of multipole intensities due to complex formation is translated by the second-order contributions  $E_{\text{pol}}$  and  $E_{\text{ct}}$ , while the first-order electrostatic contribution is computed with the multipolar distributions that retain the intensities they have in the isolated molecule or molecular fragment. The electrostatic field giving rise to the polarization contribution is itself computed with the permanent multipolar distribution augmented with induced dipoles derived by a self-consistent iterative procedure. We have extended this representation to the case of intramolecular interactions. Since the inception of the SIBFA procedure,<sup>9</sup> these are computed as the sum of intermolecular interactions between the constitutive fragments of the molecule. In the procedure that is presently used,  $E_{\text{MTP}}$  is computed with junction multipoles that are redistributed along the junction bond, namely its origin, its extremity, and its barycenter. These junction multipoles do not interact with the two connected fragments, since such interactions are large and constant. To compute  $E_{\text{pol}}$ , on the other hand, an alternative set of multipoles is used, for which no redistribution along the junctions is done. In this fashion, each individual fragment retains the net charge it has prior to the assembling procedure, namely 0 if neutral,  $-1$  if anionic, and  $1$  if cationic, whereas it is not retained following redistribution. This prevents an imbalance of  $E_{\text{pol}}$  between two successive fragments that have lost their net charges, and that could be amplified in the complete molecule due to the nonadditivity of  $E_{\text{pol}}$ . It was also necessary to prevent overlaps involving the H atoms belonging to the X–H junction bonds. Such bonds were shrunk by carrying back the end H atoms on the X atom whence the bond originates.

Finally, upon computing the intermolecular interactions between flexible molecules, *inter*- and *intramolecular inter*-fragment interactions have to be computed simultaneously and consistently as a single integrated energy. This need is a consequence of the nonadditivity of  $E_{\text{pol}}$  and  $E_{\text{ct}}$ . It illustrates the connections between nonadditivity and transferability.  $\Delta E_{\text{int}}$  between two or more interacting molecules can be subsequently derived by subtracting from such a total energy all sums of interfragment interactions within each individual molecule.

An illustration of the manner flexible molecules are constructed from their fragments is given in Figure 3a,b. Parts a and b relate respectively to the assembly of the five first amino acids of protein Fak (focal adhesion kinase), a target for the design of antitumor drugs, and of an inhibitor belonging to the pyrrolopyrimidine series (de Courcy et al., to be published). Part a represents the first ten fragments making up the backbone. The side chains are assembled after completion of the 140 amino acid backbone. Thus Asp414 is built out from its methane and formate fragments, Tyr415 from methane, benzene, and phenol, etc. Part a also gives the numbering of the atoms that takes into account the presence of the additional centers along the chemical bonds. All individual peptide and nucleic acid fragments being stored in a library with the relevant information concerning the internal geometry, the types of atoms, the distributed multipoles and polarizabilities on proteins and nucleic acids can be constructed using software that uses in addition the information regarding the sequence and torsional angles. Part b shows the inhibitor as constructed from its constitutive pyrimidine, sp<sup>2</sup> amine, benzene, water, methane, and formate fragments. To account for conjugation effects, a prior QC computation was performed on an aminopyrimidine molecule, which was then broken up into pyrimidine and HNH<sub>2</sub>, these two entities retaining the same multipolar expansion as in the original molecule, the fictitious H atoms on their junctions having null multipoles, and while the junction





**Figure 3.** (a) Fragments making up the backbone of the five first amino acids of protein Fak (focal adhesion kinase). (b) Fak protein inhibitor as constructed from its constitutive pyrimidine, sp<sup>2</sup> amine, benzene, water, methane, and formate fragments.

bonds CH and HN have each half of the multipoles of the broken C–N bond. The other fragments already belong to the library of SIBFA fragments. Thus a new molecule can be constructed from fragments that are already present in

the library. If this is not the case, a QC computation is done on it enabling to derive its distributed multipoles and polarizabilities, and the fragment can be stored for future uses. Most QC computations are done with the GAMESS

package.<sup>44c</sup> In the general case, SIBFA energy minimizations are done in internal coordinates. Conformational changes thus take place by torsions around the junction bonds. The approach of a given molecule toward another is governed by six intermolecular variables. Molecular dynamics are done in Cartesian coordinates, while standard bond lengths and valence angles are enforced by stretching and bending harmonic restraints.

Our first studies on the intermolecular interactions of flexible molecules bore on the complexes of Zn(II) with glycine and the glycine zwitterion,<sup>72</sup> on the one hand, and with  $\alpha$ - and  $\beta$ -mercaptocarboxamides, on the other hand.<sup>73</sup> The latter constitute the Zn-binding moieties of several potent Zn-metalloenzyme inhibitors.<sup>74</sup> Following the procedures outlined above, it was possible to closely reproduce the QC values of Zn(II) binding in different configurations of approach or as a function of the zwitterionic state<sup>72</sup> and its conformational dependencies.<sup>73</sup> These studies were extended to complexes of Cu(I) with flexible molecules involved in the formation of supramolecular assemblies<sup>57</sup> and to those of Cu(II) with a new class of HIV-1 inhibitors that can fit the protease dimer binding site.<sup>75</sup> SIBFA was also used to study of the conformation-dependent intermolecular interactions of the triphosphate anion, the tetra-anionic end of ATP, with Zn(II) used as a probe.<sup>76</sup> The results are commented on in Supp. Info 5 in the Supporting Information. We next considered the high-resolution X-ray structure of the complex of HPPK with a nonhydrolyzable ATP analog, that has one central ester O replaced by methylene.<sup>77</sup> The results are commented on in Supp. Info 6 in the Supporting Information.

*Conformational Studies of Oligopeptides. Test on the Alanine Tetrapeptide.* Most previous analyses of transferability had borne on charged flexible ligands and their interactions with divalent cations. The predominant effects of divalent cation binding on the ligands involved the polar/charged heteroatoms and their connecting bonds, since these were the most exposed to the incoming cation and involved simultaneously the mutual interactions between these sites. The junction bonds, being less accessible, were expected to play a lesser role. It was then important to evaluate the impact of the approximations done for the handling of the interfragment junctions in the case of neutral molecules and in the absence of external charge. This is exemplified by the case of oligopeptides of alanine, which has the simplest side chain, namely a methyl group. The oligopeptide backbones are assembled in SIBFA as a succession of formamides and methyl groups, and the Ala side chain is represented by a methyl group.<sup>50</sup> For pure intramolecular interactions, the interactions involving the junction bonds are expected to have weights comparable to those involving the nonjunction bonds or the atoms. For the evaluation of the SIBFA conformational energies in such molecules, we have in ref 78 computed the energies of ten alanine tetrapeptide conformers, that were used by Beachy et al.<sup>79a</sup> to benchmark standard molecular mechanics potentials against ab initio computations. The structures of these ten conformers are recalled in Supp. Info 7 in the Supporting Information. Starting from these, SIBFA energy minimizations were performed as a function of the  $\phi$ ,  $\psi$ , and  $\chi$  dihedral angles with fixed standard internal

coordinates. At the converged minima, single-point QC computations were performed with three different basis sets: CEP 4-31G(2d), 6-311G\*\*, and cc-pvtz(-f). The results are reported in Table 4a,b. Table 4a reports the QC results at the HF level and the SIBFA ones in the absence of the  $E_{\text{disp}}$  contribution. Table 4b reports the results in the presence of correlation, namely at the DFT level with different functionals for the exchange-correlation terms, namely Becke88/Perdew 86,<sup>80</sup> PLAP3,<sup>81</sup> K2-BVWN,<sup>82</sup> and B3LYP;<sup>83</sup> at the LMP2 level;<sup>71</sup> and at the MP2 level. The SIBFA results are given with two different scalings of  $E_{\text{disp}}$  by 1.0 and by 0.8. The latter value was previously found<sup>28d</sup> to enable the reproduction by  $\Delta E_{\text{tot}}(\text{SIBFA})$  of the  $-5.1$  kcal/mol water–water dimerization energy that resulted from a large basis set MP2 study of this dimer by Feyereisen et al.,<sup>79b</sup> with  $\Delta E(\text{SIBFA})$  in the absence of  $E_{\text{disp}}$  providing a very close agreement to the corresponding HF value by these authors ( $-3.9$  versus  $-3.6$  kcal/mol, respectively). Table 4a shows  $\delta E(\text{SIBFA})$  to give the same ordering of conformer stability as the CEP 4-31G(2d) and 6-311G\*\* basis sets. The  $\delta E(\text{SIBFA})$  values are close to those found with the CEP 4-31G(2d) basis set. Such agreements also carry out to the cc-pVTZ(-f) basis set, with a maximal error of 1.4 kcal/mol for high-lying conformer 7, and an rms of 0.7 kcal/mol. These results indicate that the introduction of  $E_{\text{pol}}$  in pure intramolecular interaction energies, with the same calibration as for intermolecular interactions, could be realized in a balanced fashion, without overestimating the stabilities of the most folded conformers. The values of  $\delta E^*(\text{SIBFA})$ , namely without  $E_{\text{pol}}$ , have a downgraded agreement with the  $\delta E(\text{QC})$  ones. The values of  $\delta E_{\text{mono}}(\text{SIBFA})$ , computed by limiting  $E_{\text{MTP}}$  to the sole monopole–monopole term, even though in the presence of  $E_{\text{pol}}$ , have an even worse agreement. Thus, in the framework of SIBFA, explicit introduction of a polarization contribution is clearly insufficient to restore the agreement with QC computations if the electrostatic contribution were to be limited to the sole monopole–monopole term. Table 4b shows that correlation brings a reduction of the  $\delta E$  values, the folded conformations having their relative stabilities improved with respect to the extended ones. However, the extent of  $\delta E$  reduction depends upon the procedure, the basis sets, and, for the DFT computations, upon the exchange-correlation functional as well. The LMP2 computations bring  $\delta E$  reductions that are intermediate between the DFT and MP2 ones. The results of Table 4b were commented on in more detail in ref 78. It is observed that the  $\delta E_{\text{tot}}(\text{SIBFA})$  values with a scaling of 0.8 for  $E_{\text{disp}}$  (conform to the value adopted in ref 28d concerning the water dimer) agree best with the 6-311G\*\* LMP2 calculations, with which they give a 1.3 kcal/mol rms. It is presently difficult to trace back to a specific contribution the origin of the 0.7–1.3 kcal/mol rms increase upon passing from the uncorrelated to correlated levels, since there are no QC energy-decomposition analyses for intramolecular interactions. It could be instructive in future calculations to resort to correlated rather than uncorrelated multipoles and polarizabilities to construct the fragments as recently initiated for intermolecular interaction energies.<sup>39b</sup> The results of Table 4 should not be compared to those published by Beachy et

**Table 4.** Ala Tetrapeptide: (a) Values of the HF and SIBFA (without the Dispersion Component) Conformational Energy Differences  $\delta E$  and (b) Values of the DFT, LMP2, MP2 Quantum-Chemical, and SIBFA Conformational Energy Differences  $\delta E^a$ 

(a)						
conformer	ab initio HF			SIBFA		
	4-31G(2d)	6-311G**	cc	$\delta E$	$\delta E^b$	$\delta E_{\text{mono}}$
1	0.0	0.0	0.0	0.0	0.0	0.0
2	1.0	1.1	1.3	0.8	1.0	-0.7
3	10.2	9.3	10.5	11.3	14.9	6.1
4	3.2	3.1	3.4	3.3	3.2	1.9
5	7.3	7.3	7.6	7.7	6.5	6.3
6	7.5	6.1	7.7	8.1	9.8	9.2
7	13.4	12.2	13.7	12.3	13.4	12.4
8	17.6	16.2	17.8	18.9	21.8	22.0
9	30.0	28.4	29.8	29.6	35.7	27.4
10	28.2	26.6	28.5	28.9	34.6	38.1

(b)										
conformer	DFT				B3LYP/ cc	LMP2		MP2	SIBFA	
	B88/PD86	PLAP3	K2	B3LYP/6-311G**		6-311G**	cc		6-311G**	c d
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.7	1.0	0.7	0.7	1.0	0.6	3.4	-0.1	0.3	0.4
3	6.9	11.4	7.7	6.6	7.7	5.8	10.8	1.5	3.4	4.8
4	3.5	3.7	2.8	2.7	3.0	1.9	2.1	1.1	2.3	2.5
5	7.7	8.1	7.0	7.2	7.3	5.8	5.8	4.4	6.5	6.7
6	7.2	9.5	6.8	5.4	6.6	3.9	4.2	0.8	3.6	4.5
7	11.5	14.7	11.2	9.8	11.1	7.5	9.7	3.1	5.2	6.6
8	15.3	19.9	15.0	13.6	15.0	11.7	16.3	6.6	10.7	12.1
9	22.3	33.0	23.9	22.5	23.7	21.5	28.3	15.4	17.5	19.8
10	24.0	32.5	24.0	21.4	23.6	17.6	20.2	10.7	17.5	19.7

<sup>a</sup> Single-point ab initio computations are performed on the SIBFA minima. The  $\delta E$  values (kcal/mol) are computed with respect to the energy of the most stable conformer taken as energy zero. <sup>b</sup>  $\delta E$ : SIBFA energy value in the absence of  $E_{\text{pol}}$ . <sup>c</sup> A multiplicative factor of 1 is used for the  $E_{\text{disp}}$  component. <sup>d</sup> A multiplicative factor of 0.8 is used for the  $E_{\text{disp}}$  component.

al. since as mentioned above, energy minimization was only along the torsion angles, while valence angles and bond lengths were not relaxed. Toward this aim, angle bending and bond stretching force constants have to be recalibrated in the framework of SIBFA. Because the formulation of the energy is different than in standard molecular mechanics procedures, such constants can significantly differ from the 'classical' ones. This was actually undertaken regarding the peptide  $\text{sp}^3$   $\text{C}_\alpha$ -centered angle, and the results were commented on.<sup>78</sup> While the ten Ala tetrapeptide conformers had been reinvestigated for the first time in the context of polarizable potentials,<sup>84</sup> the results reported in ref 78 were the first such investigation that used distributed ab initio multipoles and polarizabilities. The very first conformational studies of dipeptides that resorted to distributed multipoles date back to 1985, during the inception of SIBFA.<sup>50</sup> The polarizabilities then used were scalar polarizabilities, and the contribution of  $E_{\text{pol}}$  was smaller than in the present studies; this was due to the use of much smaller basis sets. Further studies on the Ala dipeptide as well as on  $\beta$ -turn forming peptides were published in 1998 using the CEP 4-31G(2d) basis set as a follow-up to the 1995 SIBFA refinements.<sup>85</sup> Addressing the issue of multipole transferability leading to that of an appropriate representation of interfragment  $E_{\text{pol}}$

and  $E_{\text{ct}}$  was done subsequently<sup>78</sup> which then led to the study reported here.

At this stage the existence of dependencies between anisotropy, nonadditivity, and transferability is worth mentioning. Such dependencies thus exist *between anisotropy and nonadditivity*. A recent example was provided by water chains of up to 12 molecules.<sup>23d</sup> Thus off-center lone pair polarizabilities not only are a determinant of anisotropy but also enhance cooperativity due to their closer distances to the polarizing partners. By contrast, atom-centered polarizabilities give rise to underestimated  $E_{\text{pol}}$  values with respect to QC computations. There are also dependencies *between nonadditivity and transferability* as occurs upon handling flexible molecules, namely regarding the issue of multipole transferability. Thus it was shown that both nonadditive  $E_{\text{pol}}$  and  $E_{\text{ct}}$  contributions, which resort to permanent multipoles and induced dipoles, enabled for the accounting of the impact of changes in multipole intensities upon building a large molecule from fragments and upon conformational changes. On the other hand, the existence of connections *between separability and transferability* is not clear. While such connections are obvious in the case of intermolecular interactions between rigid fragments, they could be questioned for intramolecular interactions. In this case, separabil-



ity of the contributions could only be considered regarding the interfragment interaction energies. Thus as was shown above for the Zn(II) complexes of triphosphate,<sup>76</sup> while  $\Delta E(\text{QC})$  can be correctly reproduced by  $\Delta E(\text{SIBFA})$ , this is not the case for the individual contributions.

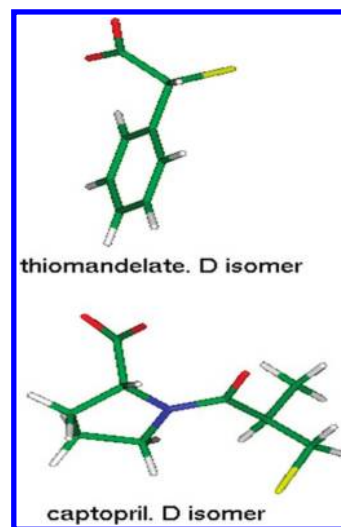
**II. Extension to Molecular Recognition Problems.** In addition to the above-mentioned applications to Cu(I) and Cu(II) complexes in the context of supramolecular chemistry, SIBFA was applied to the following systems:

*Toward APM Applications to DNA and RNA.* The binding of hydrated Zn(II) and Mg(II) cations to guanine, adenine, and the G-C and A-T base pairs was investigated in parallel by SIBFA and QC, showing close numerical agreements in  $\Delta E_{\text{int}}$  values.<sup>86</sup> Direct as well as through-water binding of the cations to the bases was investigated. SIBFA was able to account for the significant cooperativity ( $-15$  kcal/mol) of Zn(II) binding to the G-C base pair. These studies were extended to 5'-guanosine monophosphate, a basic building block of DNA/RNA helices.<sup>87</sup> With the ribose in either a C2'endo or a C3'endo conformation, three competing binding modes were investigated. They involved the following: (a) simultaneous cation binding to both phosphate O<sub>1</sub> and guanine N<sub>7</sub>; (b) direct binding to O<sub>1</sub> and through-water binding to N<sub>7</sub>; (c) and, conversely, through-water binding to O<sub>1</sub> and direct binding to N<sub>7</sub>. At both HF and DFT levels, close agreements were observed between the SIBFA and the QC energy values, both regarding the magnitudes of the binding energies and the ranking of the different binding modes. These studies will be extended to oligonucleotides of increasingly larger sizes and to their complexes with metal cations and ligands.

*Toward de Novo Predictions of the Conformations of Short Zn-Metallo-Oligopeptides.* We have resorted to a hierarchical procedure which, starting from random conformations, selects candidate conformers by a Monte Carlo approach with a potential of mean-force<sup>88</sup> and then postprocesses them using SIBFA.<sup>89</sup> This procedure was applied to the 18-residue Zn-finger of the HIV-1 nucleocapsid protein having a CCHC core (three Cy<sup>-</sup> residues and a His one) and its CCHH mutant. rms deviations of the C <sub>$\alpha$</sub>  backbones of 3.5 Å and in the 2.2–3 Å range were found for these two Zn-fingers, respectively. Extensions of the procedure to include algorithms for global minimum searches<sup>90</sup> will be considered for future applications.

*Complexes of Zn-Metalloproteins with Inhibitors.* The targeted proteins are two bacterial enzymes, the Zn-metallo- $\beta$ -lactamase from *B. fragilis* and phosphomannoisomerase (PMI) from *C. albicans*, and the C-terminal Zn-finger from HIV-1 nucleocapsid.

(a) *Complexes of Metallo- $\beta$ -Lactamase (MBL) with Captopril and Thiomandelate Mercaptocarboxylate Inhibitors.* There are presently no inhibitors with sufficient affinity to MBL so as to be clinically useful, which raises a serious health concern. Mercaptocarboxylate derivatives endowed with micromolar affinity to MBL could be used as possible leads for the design of more efficient inhibitors. These are D- and L-captopril and D- and L-thiomandelate (Figure 4). While binding to *B. fragilis* MBL is known to occur upon removal of the Zn-chelating hydroxy anion and its replace-

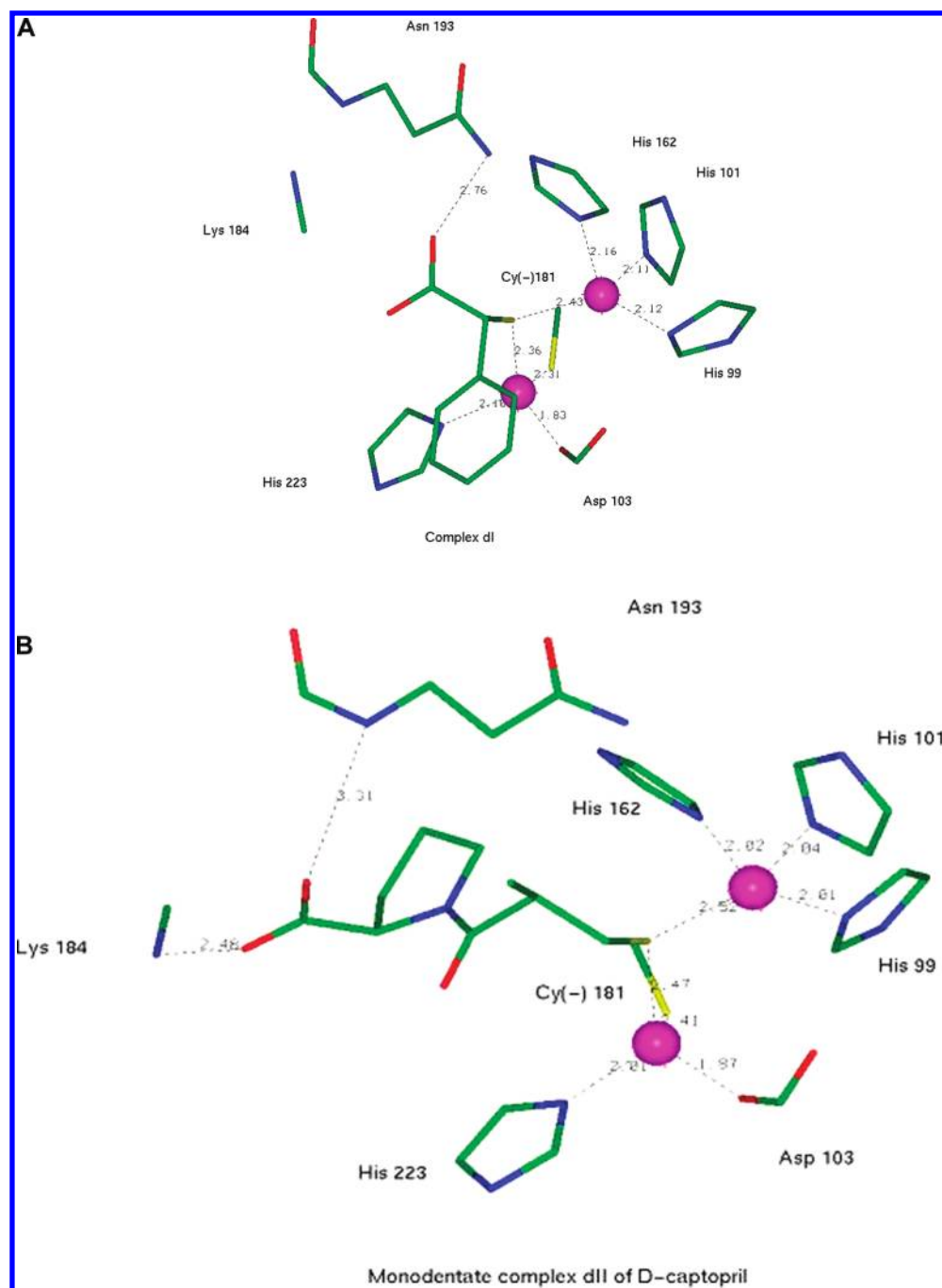


**Figure 4.** Molecular structures of D-captopril and D-thiomandelate. Reprinted with permission from Antony et al. *Journal of Computational Chemistry* **2005**, 26, 1131. Copyright 2005 John Wiley.

ment by one or by both anionic moieties of the inhibitor, there was no high-resolution structural information regarding the actual structures of their complexes with MBL. We have in refs 91 and 92 modeled a 108-residues model of MBL on the basis of the high-resolution X-ray structure by Concha et al. of uninhibited MBL.<sup>68</sup> Thiomandelate was built from methanethiolate, methane, benzene, and formate fragments. Captopril was built from methanethiolate, methane, proline, and formate fragments. Energy minimization (EM) was performed on the side chains of the residues making up the binding site, on all inhibitor torsion angles, and on the six inhibitor intermolecular variables as well as on the positions of the two Zn(II) cations. Different starting positions for EM were chosen, that were obtained from an exploratory docking that used constrained MD with the Accelrys software and the Cff91 force field,<sup>93a</sup> the constraints corresponding to enforcements of mono- or bidentate binding.

*Thiomandelate Complexes.* Seven and four distinct complexes were characterized for D- and L-thiomandelate, respectively.<sup>92</sup> Figure 5a represents the d-I D-thiomandelate complexes. In d-I, thiomandelate binds monodentately to the two Zn(II) cations through its S<sup>-</sup> atom, and the carboxylate binds to the Asn193 side chain. d-IIb is a bidentate binding mode in which the carboxylate binds to one Zn(II) cation. d-III is an alternative binding mode in which the carboxylate has replaced S<sup>-</sup> in the Zn(II)-chelating position and binds simultaneously through its second O atom to the Asn193 side chain. At the converged unconstrained SIBFA minima, the energy balances were completed upon computing the solvation energy  $\Delta G_{\text{solv}}$  using the Langlet-Claverie Continuum reaction field procedure.<sup>51a</sup> Energy balances including  $\Delta G_{\text{solv}}$  were more favorable for D-thiomandelate than for L-thiomandelate binding, consistent with experimental results, and for both isomers, more favorable for mono- than bidentate binding.

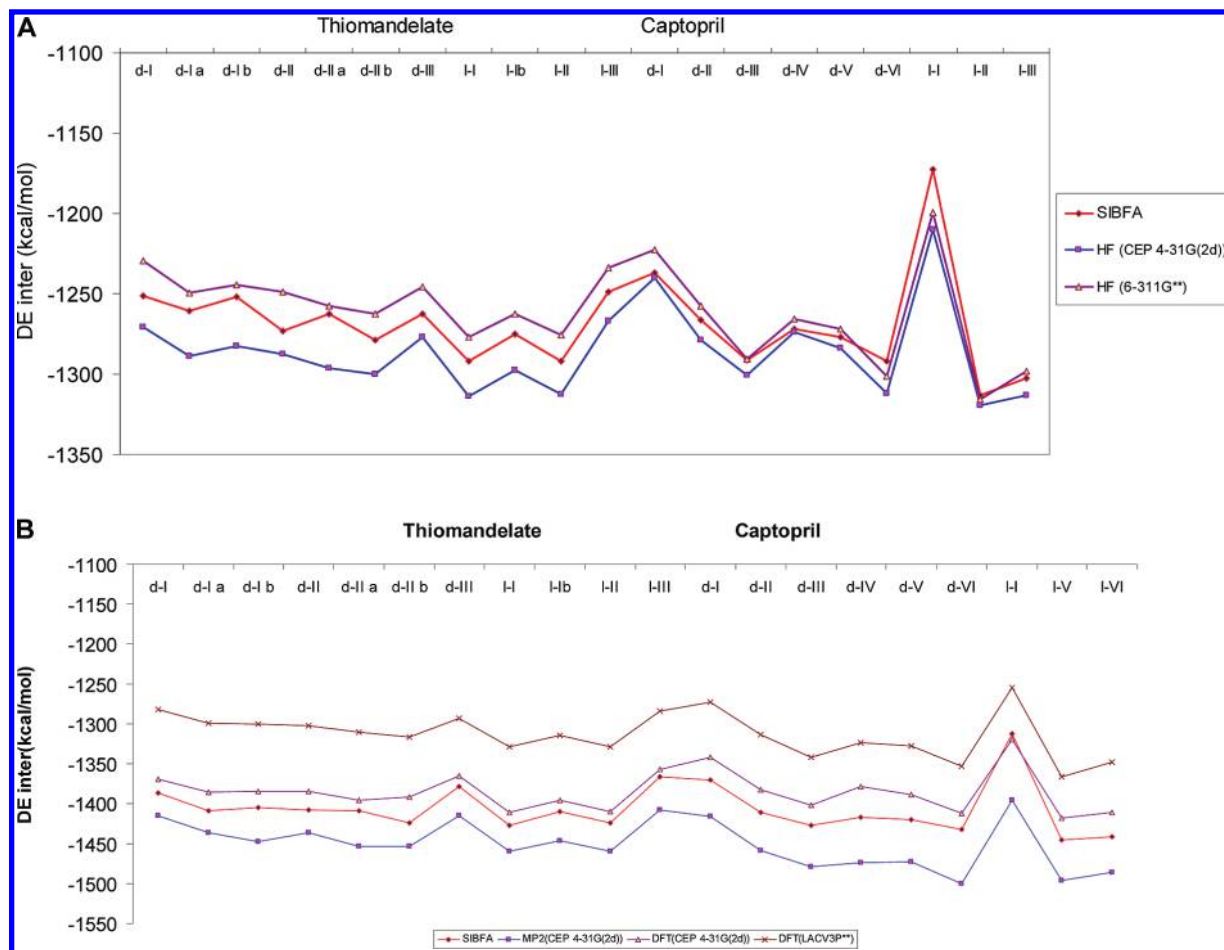
*D- and L-Captopril Complexes.* The competing modes can be either monodentate with binding of the sole S<sup>-</sup> to the two Zn(II) cations, or bidentate, involving additional Zn(II)-



**Figure 5.** (a) Representative complexes of D-thiomandelate with metallo- $\beta$ -lactamase. Reprinted with permission from Antony et al. *Journal of Computational Chemistry* **2005**, 26, 1131. Copyright 2005 John Wiley. (b) Representative complexes of D-captopril with metallo- $\beta$ -lactamase. Reprinted with permission from Gresh *Current Pharmaceutical Design* **2006**, 12, 2121. Copyright 2006 Bentham Science Publisher, Ltd.

binding by either the formate or the carbonyl group. Up to nine distinct complexes could be characterized, as discussed in more detail in the preceding papers.<sup>5,91</sup> Thus monodentate complex d-II is stabilized, in addition to Zn(II) chelation by  $S^-$ , by interactions of the carboxylate with both the Lys184 side chain and the Asn193 main chain (see Figure 5b). In complex d-III, it is the carbonyl that now interacts with the Asn193 main chain. In complex d-IV, the carbonyl binds to one Zn(II) cation, and the formate is bound to the Lys184 side chain. In complex d-VI, the formate binds simultaneously to the Zn(II) cation and the Lys184 side chain, while the carbonyl binds the Asn193 side chain. The energy

balances showed D-captopril to be more favorably bound by MBL than L-captopril, consistent with experimental results, and that the best binding mode was monodentate mode d-II. Although as mentioned above, there are no X-ray structures of *B. fragilis* MBL complexes with captopril, it is worth mentioning that a high-resolution X-ray structure on the complex of a MBL from a *P. aeruginosa* strain with a mercaptocarboxamide inhibitor analogous to D-captopril had shown very similar binding modes: monodentate binding of  $S^-$  to the two Zn(II) cations, and the terminal carboxylate simultaneously bound to the side chain of Lys161 and the main chain of Asn167, two residues that occupy positions



**Figure 6.** (a) Compared evolutions of  $\Delta E(\text{SIBFA})$  and  $\Delta E(\text{HF})$  in the 19 complexes of captopril and thiomandelate with the two Zn(II) cations and the eight residues modeling the metallo- $\beta$ -lactamase binding site. SIBFA vs HF interaction energies (kcal/mol). (b) Compared evolutions of  $\Delta E_{\text{tot}}(\text{SIBFA})$ ,  $\Delta E(\text{MP2})$ , and  $\Delta E(\text{DFT})$  in the 19 complexes of captopril and thiomandelate with the two Zn(II) cations and the eight residues modeling the metallo- $\beta$ -lactamase binding site. Values (kcal/mol) of  $\Delta E(\text{SIBFA})$  with  $E_{\text{disp}}$  and correlated quantum-chemical interaction energies. Reprinted with permission from Antony et al. *Journal of Computational Chemistry* **2005**, 26, 1131. Copyright 2005 John Wiley.

similar to the respective Lys184 and Asn193 ones of *B. fragilis* MBL.<sup>94</sup> While this should lend credence to the APMM calculations, an equally demanding test relates to comparing the  $\Delta E_{\text{int}}$  values to parallel  $\Delta E(\text{QC})$  ones in model binding sites extracted from the cavity. Such models total 98 atoms, a size rendering them amenable to QC computations. The binding cavity has a very high local concentration of ionic charges. In addition to the two Zn(II) dications at 3.5 Å from one another, these include the anionic charges of Asp104 and  $\text{Cy}^{-181}$ , those of the inhibitor methanethiolate and formate groups, and the cationic charge of Lys184. Thus, similar to the kinase binding site, very important nonadditivity effects can be anticipated, underlining again the need to correctly account for the simultaneous interplay of inter- and intramolecular polarization and charge transfer. The SIBFA/QC comparisons were done at both uncorrelated and correlated levels. At the HF level,  $\Delta E(\text{SIBFA})$  was compared to  $\Delta E(\text{HF})$  using either CEP 4-31G(2d) or LACV3P\*\* basis sets. At the correlated level,  $\Delta E_{\text{tot}}(\text{SIBFA})$  was compared to  $\Delta E(\text{DFT})$  with both basis sets and to  $\Delta E(\text{MP2})$  with the

CEP 4-31G(2d) basis set. Such comparisons are discussed below together with those done for the thiomandelate complexes.

We have regrouped in Figure 6a,b the captopril and thiomandelate results under the form of graphs representing the evolutions of  $\Delta E(\text{SIBFA})$  and  $\Delta E(\text{QC})$  values for all 20 complexes, namely d-I up to I-III for captopril and d-I up to I-III for thiomandelate. Figure 6a shows  $\Delta E(\text{SIBFA})$  to have values consistently intermediate between the  $\Delta E(\text{HF})$  ones with the CEP 4-31G(2d) and LACV3P\*\* basis sets, with the sole exception of the highest-lying complex I-I. The SIBFA curve shows very good agreement with the QC one, except at the level of complex d-II for thiomandelate. This is because d-II is computed in SIBFA to have a more favorable  $\Delta E$  than d-IIa, while the reverse occurs with the HF calculations. Such an inversion involves differences of 10 kcal/mol out of 1260, namely less than 1%. At the correlated level,  $\Delta E_{\text{tot}}(\text{SIBFA})$  has values intermediate between the MP2 and the DFT ones with the CEP-431G(2d) basis set. The SIBFA curve displays very good correlation with the QC ones (Figure 6b).



Such results are highly encouraging, notwithstanding further SIBFA refinements. They could be used to benchmark other polarizable molecular mechanics procedures. The structures of the 20 complexes are available as Supporting Information to ref 92 as well as on the Web at <http://www.lct.jussieu.fr/pagesperso/jpp/SIBFA.html>.

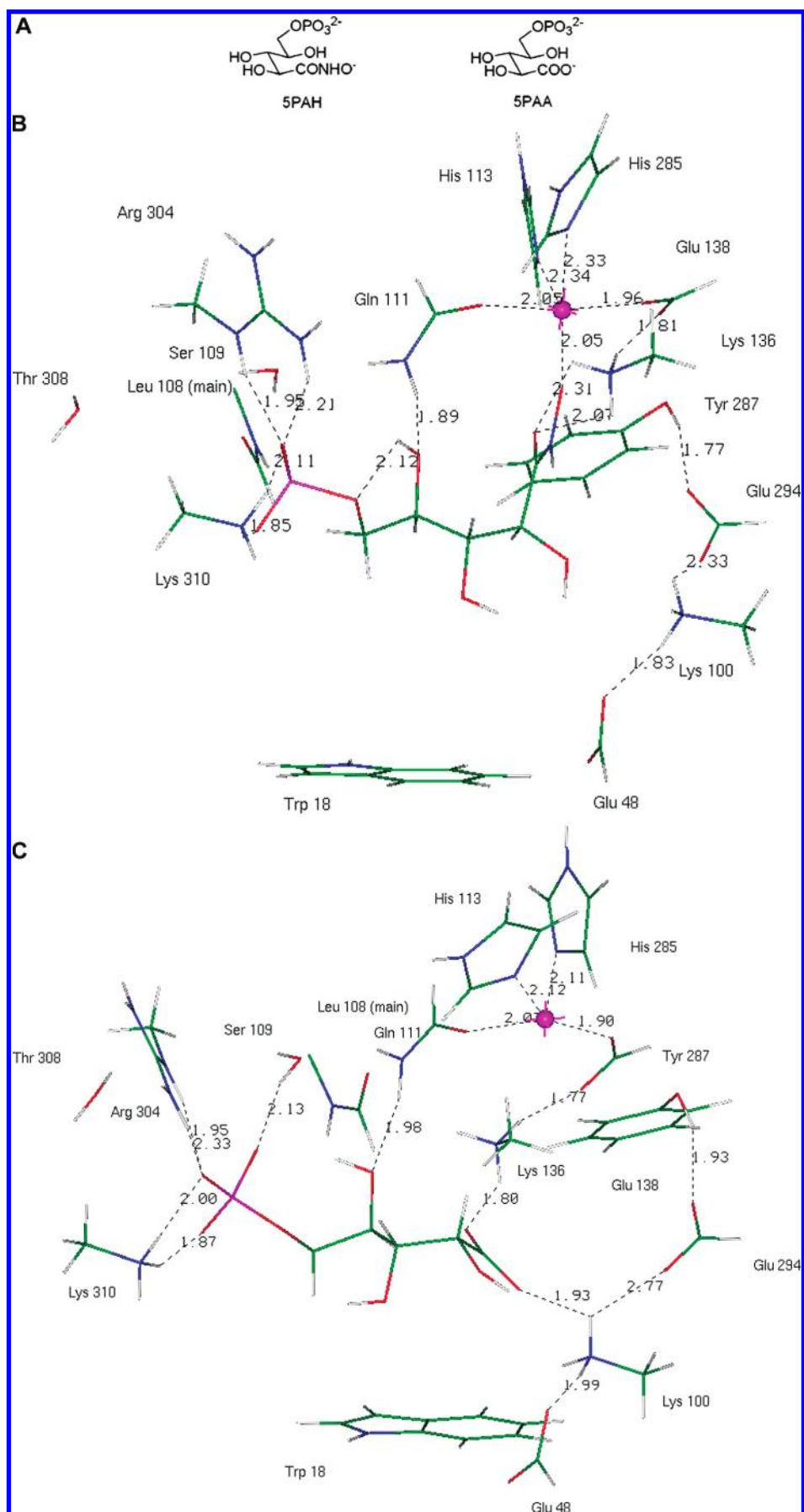
As concerns the energy balances done in the 108-residue model, we wish to note that while the D isomers of both captopril and thiomandelate are predicted to be the better-bound isomers, the energy differences between competing complexes are likely to be overestimated, since the interaction energy values (without  $\Delta G_{\text{solv}}$ ) represent enthalpies, not free energies, as they do not presently include entropy effects due to the reduction of translational and rotational motions of the ligand upon complex formation (for a recent discussion, see ref 93b). The captopril versus thiomandelate energy balances should not be presently compared, essentially because of a different calibration of  $\Delta G_{\text{solv}}$  that was adopted in the thiomandelate study, conforming to the one used in ref 87.

(b) *Complexes of Phosphomannosomerase (PMI) to 5-Phospho-D-arabinohydroxamate and 5-Phospho-D-arabinonate Inhibitors.* PMI is a Zn(II)-dependent isomerase that catalyzes the reversible isomerization of D-mannose 6-phosphate and D-fructose-6-phosphate. It plays an essential role in the metabolism of bacteria and microorganisms. It is involved in several pathologies, such as leishmaniasis, cystic fibrosis, and opportunistic infections in immuno-depressed individuals.<sup>95</sup> There are no PMI inhibitors presently in use clinically. 5-Phospho-D-arabinohydroxamate (5-PAH, Figure 7a) was recently reported as the most potent PMI inhibitor, displaying nanomolar affinity.<sup>96</sup> Replacing hydroxamate by carboxylate yielding 5-phospho-D-arabinonate (5-PAA, Figure 7a) resulted in loss of inhibitory potency. Zn(II) binding was experimentally shown to occur through hydroxamate rather than phosphate, despite the latter's dianionic character. We have performed SIBFA energy minimizations on the complexes of 5-PAH and 5-PAA with a 164-residue model of PMI.<sup>97</sup> We used the X-ray crystal structure of uninhibited PMI<sup>98</sup> as a starting point. As in the MBL studies, the PMI backbone was held rigid, and the side chains of the residues making up the binding site were relaxed. Two different starting points were considered, with either hydroxamate/carboxylate or phosphate bound to the Zn(II)-binding site. The non-Zn(II)-bound anionic moiety interacted with two basic residues, Arg304 and Lys310, at the entrance of the receptor cavity. In addition, the 5-PAH minima were used as new starting points for energy minimization of the 5-PAA complexes and conversely. This yielded a total of eight complexes. One more 5-PAH complex was investigated in which bidentate Zn(II)-binding of hydroxamate through both O atoms was enforced and subsequently relaxed. These energy minimizations were performed first in vacuo and then refined resumed by including  $\Delta G_{\text{solv}}(\text{LC})$  in the total energies. In modes A and A', hydroxamate is bound bidentately and monodentately, respectively. Bidentate binding occurs at the expense of Zn(II) binding to His285. In mode B, phosphate binding displaces both His residues from Zn(II). In mode B', on the other hand, phosphate is bound to Zn(II) only

indirectly, namely through Lys136 that is itself H-bonded to Zn-bound Glu138. Figure 7b gives a representation of the most stably bound complex of 5-PAH complex with PMI, namely A', limited to the binding site. It was similarly found that in the 5-PAA complexes the carboxylate could bind either directly to Zn(II) or indirectly through the Lys136-Glu138 salt bridge. Figure 7c represents the most stably bound complex of 5-PAA, which corresponds to mode C'.

In agreement with experiment, the final energy balances indicated 5-PAH to have a significantly larger affinity than 5-PAA and that Zn(II) binding should occur through hydroxamate/carboxylate rather than phosphate. However, as in the MBL case, it was necessary to validate the values of  $\Delta E_{\text{int}}$  by comparisons with parallel QC computations on the model binding site, now encompassing up to 140 atoms. The results reported in Table 5 indicate at both uncorrelated and correlated levels close agreements with the QC results. As for the model MBL complexes, the nine structures could serve to benchmark other PMM approaches. They are provided as Supporting Information for ref 97 as well as at the above-mentioned Web site. Extensions of the present work are in the design of novel 5-PAH analogs, in order to further improve their binding affinities.

(c) *Binding of a Mercaptobenzamide Thioester to the C-Terminal Zn-Finger of HIV-1 Nucleocapsid.* The HIV-1 nucleocapsid (NCp7) plays a pivotal role in HIV-1 metabolism. It has two highly structured Zn-binding domains with the C(X2)C(X4)H(X4)C motif (where X is any amino acid). It is a potential target for the development of novel antiviral drugs, because, in contrast to the HIV-1 protease and reverse transcriptase, mutations can impair its structure and function. This has led to the design of 'Zn-ejector' molecules that can disrupt Zn(II) binding.<sup>99</sup> Recently, mercaptobenzamide thioesters have been designed.<sup>99d,e</sup> One compound, denoted as C-247 (Figure 8a) has an S-connected carbonyl group that could make a covalent bond with the S<sup>-</sup> atom of a Zn-coordinating NCp7 residue. Thus if the proximity between the carbonyl C and one Cys S were sufficient (in the 3.0–3.6 Å range), and if the S–C–O angle were adequate, a covalent bond could be formed, entailing loss of Zn-binding. We have performed SIBFA energy minimization on the binding of compound C-247 with residues Arg32-Gln53 of the C-terminal Zn-finger.<sup>100</sup> Both main-chain and side-chain torsion angles were relaxed. One of the most stable structures, represented in Figure 8b, complies with such requirements. It is stabilized by a double H-bond of the carboxamide chain with the side chain of Gln45 and by partial stacking of the benzene ring over the Trp37 ring. Two additional H-bonds are between the Lys34 main-chain carbonyl and the end carboxamide side chain and between the Gln45 side-chain N and the thioester carbonyl O. The energy balances including  $\Delta G_{\text{solv}}$  are reported in Table 6. They are computed as the difference between the minimized energies of the C-247–NCp7 complex, on the one hand, and those of isolated C-247 and NCp7, separately minimized prior to complexation, on the other hand. They indicate the predominant role of the second-order terms and, in particular, of  $E_{\text{pol}}$  and  $E_{\text{disp}}$ , in complex stabilization. By contrast,  $E_1$  is destabilizing. As a continuation of this work, we will seek

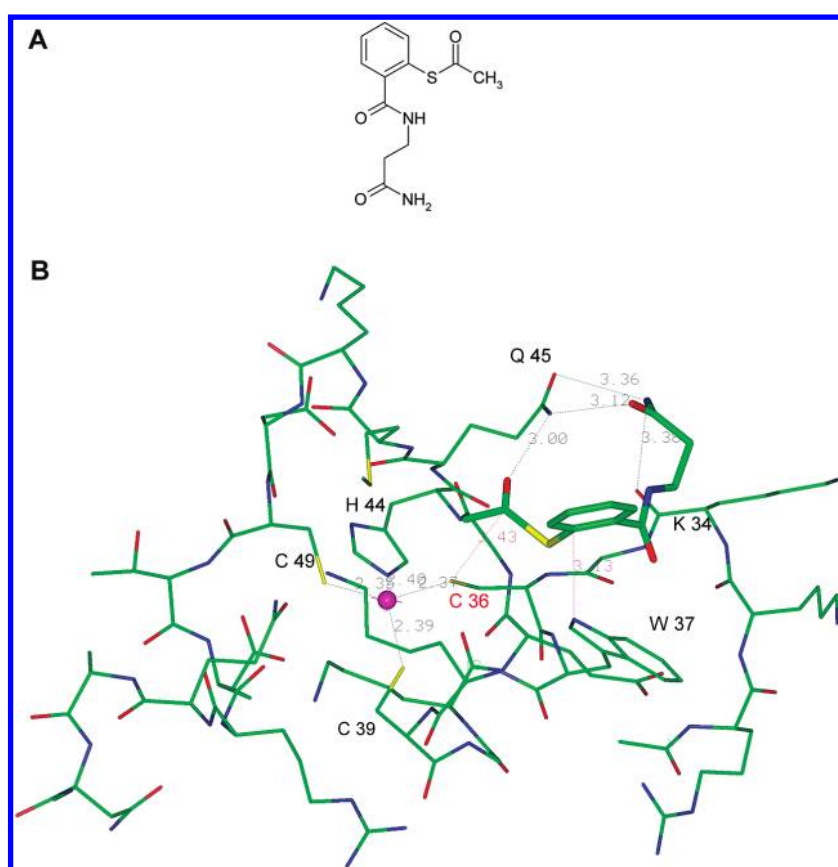


**Figure 7.** Representation of 5-PAH and 5-PAA PMI inhibitors as well as representative complexes of 5-PAH with the model binding site of PMI. Reprinted with permission from Roux et al. *Journal of Computational Chemistry* **2007**, 28, 938. Copyright 2007 John Wiley.

**Table 5.** Interaction Energies (kcal/mol) of the Bifunctional Inhibitors in the Model Binding Site (MBS) Consisting of 14 Residues (See Text) Extracted from Their PMI Complexes in the Two Competing Arrangements<sup>a</sup>

	PMI-5PAH					PMI-5PAA				PMI
	A	A'	A''	B	B'	C	C'	D	D'	
$E_{\text{MTP}}$	-1396.2	-1417.3	-1383.8	-1377.1	-1341.9	-1359.9	-1300.7	-1353.3	-1367.4	-625.6
$E_{\text{rep}}$	270.3	270.5	261.4	269.9	264.5	254.6	266.7	284.8	277.7	170.6
$E_1$	-1125.9	-1146.7	-1122.4	-1107.2	-1077.4	-1105.3	-1033.0	-1068.5	-1089.7	-454.9
$E_{\text{pol}}$	-122.0	-123.4	-122.8	-113.7	-89.2	-122.8	-147.5	-113.5	-102.5	-110.4
$E_{\text{ct}}$	-40.0	-25.1	-40.2	-33.5	-46.5	-39.8	-42.5	-40.0	-39.2	-30.9
$\Delta E$	-1287.9	-1310.8	-1285.4	-1242.9	-1224.6	-1267.9	-1224.0	-1222.0	-1231.4	-596.3
$\Delta E^b$	-1283.6	-1310.8	-1278.2	-1243.2	-1240.4	-1250.7	-1227.5	-1221.0	-1233.7	-601.6
$\Delta E^c$	-1315.0	-1344.9	-1308.6	-1264.9	-1266.1	-1278.7	-1252.0	-1242.4	-1256.6	-618.8
$E_{\text{disp}}$	-86.1	-87.0	-85.4	-79.0	-76.5	-79.6	-78.4	-78.5	-76.6	-57.3
$\Delta E_{\text{tot}}$	-1374.0	-1397.8	-1370.8	-1321.9	-1301.1	-1347.5	-1302.4	-1300.5	-1308.0	-653.6
$\Delta E(\text{DFT})^c$	-1358.5	-1386.9	-1349.8	-1295.1	-1300.9	-1324.2	-1295.4	-1288.0	-1299.9	-653.0

<sup>a</sup> (a) 5PAH with hydroxamate in the Zn-binding site; (b) 5PAH with phosphate in the Zn-binding site; (c) 5PAA with carboxylate in the Zn-binding site; (d) 5PAA with phosphate in the Zn-binding site; (e) unligated PMI with one water molecule replacing the inhibitor in the Zn(II) coordination sphere. <sup>b</sup> CEP 4-31G(2d) basis set. <sup>c</sup> LACV3P\*\* basis set.

**Figure 8.** (a) Molecular structure of a 2-mercaptobenzamide thioester inhibitor (compound C-247) of the HIV-1 nucleocapsid protein. (b) Representation of the complex of inhibitor C-247 with the second Zn-finger of HIV-1 NcP7.

to improve the binding energies of C-247 by local modifications. We are also simultaneously performing QM/MM studies to elucidate the mechanism of S–C bond formation using the structure of Figure 8b as a starting point.

**III. Toward Condensed Phase and Higher Accuracy: The Gaussian Electrostatic Model.** As quantum calculations are able to give quantitative results and have shown the importance of short-range effects on intermolecular interaction energies, the development of the SIBFA equations constitutes a notable step toward a quantitative description of intermolecular interactions in molecular mechanics en-

abling a separate reproduction of the individual physical components of the total interaction energy. However, as SIBFA attempts to mimic the anisotropy of the density, a second more natural option can be by means of interacting frozen electron densities. As demonstrated several years ago by Kim and Gordon<sup>101</sup> for atom–atom potentials based on Density Functional Theory, these could improve the description of short-range quantum effects.

Indeed, some of us recently introduced<sup>27</sup> a methodology termed Gaussian Electrostatic Model (GEM) which is able to compute molecular interaction energies in the spirit of



**Table 6.** Interaction Energies (kcal/mol) of C-247 with the Arg32-Asn55 Zn-Finger of HIV-1 NCp7

	complex	finger	C-247	summed
$E_{\text{MTP}}$	-3647.5	-3411.8	-183.5	-3595.3
$E_{\text{MTP}}^a$	-52.2			
$E_{\text{rep}}$	2773.0	2613.0	88.4	2701.4
$E_{\text{rep}}^a$	71.6			
$E_1$	-874.5	-798.8	-95.1	-893.9
$E_1^a$	19.4			
$E_{\text{pol}}$	-543.6	-507.2	-20.7	-527.9
$E_{\text{pol}}^a$	-15.7			
$E_{\text{ct}}$	-79.4	-69.3	-0.3	-69.6
$E_{\text{ct}}^a$	-9.8			
$E_{\text{disp}}$	-951.3	-855.7	-53.9	-909.6
$E_{\text{disp}}^a$	-41.7			
$E_{\text{tor}}$	59.1	53.3	+4.2	57.5
$E_{\text{tor}}^a$	1.6			
$E_{\text{tot}}$	-2389.7	-2177.7	-165.8	-2343.5
$\delta E_{\text{tot}}^a$	-46.2			
$\Delta G_{\text{solv}}$	-572.7	-560.0	-39.7	-599.7
$\delta \Delta G_{\text{solv}}^a$	+27.0			
$\delta E_{\text{tot}} + \delta \Delta G_{\text{solv}}$	-19.2			

<sup>a</sup> After subtraction of the energies of the Zn-finger and of C-247 separately minimized.

the SIBFA approach but using the formalism of density fitting<sup>102</sup> (DF) methods usually devoted to the fast evaluation of Coulomb integrals for ab initio codes. We present here an overview of recent achievements concerning GEM. We will first summarize the initial steps of the development by addressing the important issue of the calculations of the required integrals to derive intermolecular Coulomb energies from fitted densities.<sup>103</sup> Results of a first GEM version that calculates intermolecular interaction energies from isolated monomer electron densities will then be detailed.<sup>27a</sup> To conclude, a generalized GEM density fitting scheme<sup>27b</sup> will be presented as well as its extension to periodic boundary conditions (PBC)<sup>27b</sup> and to QM/MM.<sup>104</sup>

(I) *Methods. (A) From a Density Fitting Procedure to Intermolecular Coulomb Energies.* We have used the formalism of the variational density fitting method,<sup>102</sup> an approach which is usually devoted to a fast approximation of the Coulomb interaction.

This method relies on the use of an auxiliary Gaussian basis set (ABS) to fit the molecular electron density obtained from a relaxed one-electron density matrix using a linear combination of atomic orbitals (LCAO).<sup>105</sup>

$$\tilde{\rho} = \sum_{k=1}^N x_k k(r) \approx \rho = \sum_{\mu\nu} P_{\mu\nu} \phi_{\mu}(r) \phi_{\nu}^*(r) \quad (2)$$

The determination of the coefficients requires the use of a modified singular value decomposition (SVD) procedure in which the inverse of an eigenvalue is set to zero if it is below a certain cutoff.<sup>27,102</sup>

Using the fitted electronic densities, it has been shown<sup>103</sup> that it is possible to accurately compute the intermolecular Coulomb interaction energy (see eq 3) from frozen monomer densities in the direct spirit of ab initio energy decomposition schemes (see for example refs 25b,e).

$$E_{\text{Coulomb}} = \frac{Z_A Z_B}{r_{AB}} - \int \frac{Z_A \tilde{\rho}^B(r_B)}{r_{AB}} dr - \int \frac{Z_B \tilde{\rho}^A(r_A)}{r_{AB}} dr + \int \frac{\tilde{\rho}^A(r_A) \tilde{\rho}^B(r_B)}{r_{AB}} dr \quad (3)$$

By using density fitting, both long-range multipolar and short-range penetration electrostatic energies (missing in a distributed multipole treatment) are included, the errors being relatively small compared to reference ab initio data using the same density matrices.<sup>103</sup>

All the required integrals (electron–electron and electron–nuclear) were computed based on the McMurchie–Davidson recursions<sup>106</sup> enabling the use of higher angular moment Gaussian functions if required. It is important to point out that the formalism also enables an accurate representation of both electrostatic potentials and fields (Figure 9)

(B) *From a Density Fitting Procedure to Intermolecular Interaction Energies.* The reproduction of total interaction energy from fitted densities was studied based on the capability of the DF approach to compute accurate intermolecular Coulomb energies, thereby offering the possibility of a direct application of the methodology to molecular mechanics.<sup>27a</sup>

The total interaction was computed as the sum of four *separate* contributions: electrostatic (Coulomb), exchange–repulsion, polarization, and charge transfer. The central idea is that each contribution should match its Density Functional Theory (DFT) counterpart obtained using the Constrained Space Orbital Variation (CSOV) approach<sup>25d,e</sup> at the DFT level.

$$\Delta E_{\text{tot}} = E_{\text{Coulomb}} + E_{\text{exch-repulsion}} + E_{\text{pol}} + E_{\text{ct}} = E_{\text{Frozen Core}} + E_{\text{pol}} + E_{\text{ct}} \quad (4)$$

At this point, no long-range dispersion contribution was added since we focused on reproducing DFT/ B3LYP<sup>83</sup> interaction energies, but the SIBFA  $E_{\text{disp}}$  contribution could also be included.<sup>39</sup>

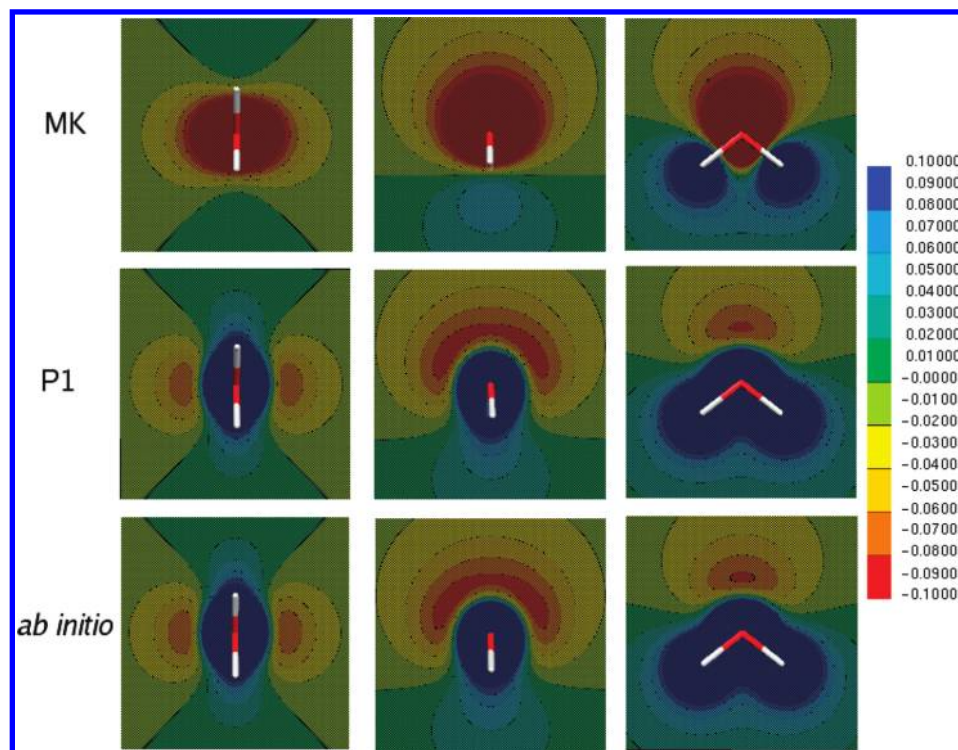
As mentioned above  $E_{\text{Coulomb}}$  is directly computed from the integrals computed using the fitted densities. Extending the approach, we followed an idea put forth by Wheatley and Price<sup>107a</sup> and computed a two-body exchange repulsion based on the overlap model. This model relies on the observed proportionality between the exchange–repulsion energy and<sup>107b</sup> the overlap of the charge density, the calculation of the latter quantity being straightforward in the framework of our density fitting approach.

$$E_{\text{exch-repulsion}} \approx K S_{\rho} \quad (5)$$

where

$$S_{\rho} = \int \rho_a(r) \rho_b(r) dr \approx \int \tilde{\rho}_a(r) \tilde{\rho}_b(r) dr \quad (6)$$

The value of the parameter  $K$  can be easily determined and corresponds to the slope of a linear regression of the overlap of charge density versus the corresponding ab initio exchange–repulsion energy values. Finally, the charge transfer and polarization energies were computed following the SIBFA



**Figure 9.** Electrostatic potential maps for the water molecule calculated from Merz–Kollman-generated charges MK, GEM fitted density, and ab initio calculation. All errors are in kcal/mol (see ref 101 for details).

**Table 7.** Intermolecular Coulomb Energies (in kcal/mol) for Ten Water Dimer Geometries for the GEM-0 Approach Fitted on B3LYP (or CCSD)/aug-cc-pVTZ Densities<sup>a</sup>

level of theory for $E_{\text{Coulomb}}$	water dimer geometry									
	1	2	3	4	5	6	7	8	9	10
CSOV (DFT)	-8.11 (-6.15)	-6.85 (-5.08)	-6.64 (-4.91)	-6.73 (-4.86)	-5.77 (-4.17)	-5.44 (-3.97)	-4.87 (-3.47)	-1.64 (-1.09)	-4.95 (-3.42)	-2.87 (-2.04)
GEM-0 (DFT)	-8.14	-6.89	-6.55	-6.77	-5.77	-5.48	-5.05	-1.77	-4.76	-2.74
CCSD (DCBS)	-7.96	-6.69	-6.48	-6.69	-5.71	-5.33	-4.89	-1.55	-4.77	-2.72
GEM-0 (CCSD)	-8.07	-6.75	-6.55	-6.58	-5.79	-5.56	-5.01	-1.68	-4.66	-2.70
SAPT (CCSD)	-8.02	-6.73	-6.49	-6.70	-5.69	-5.33	-4.96	-1.55	-4.81	-2.70

<sup>a</sup> Results in parentheses are interaction energies from a distributed multipole approach. CCSD reference calculation using the aug-cc-pVTZ basis set are provided and compared to the SAPT results (for details see ref 27a).

equations but using the electrostatic potentials and fields computed from the fitted densities.

**Results from a First Force Field Implementation: GEM-0.** We present here GEM results<sup>27a</sup> using fitted densities with an auxiliary basis set restricted to s-type ( $l = 0$ ) Gaussians on water dimers and water clusters of up to 64 molecules. As the use of s-type Gaussian functions enables the rotation of the frozen fitted monomer densities, we term this method the Gaussian Electrostatic model (GEM-0).<sup>27a</sup> Using a nine-center spherical Gaussian density model for water, we demonstrated that accurate calculations could be performed on electrostatic energies. Table 7 gives results of our model on the ten minima of the total energy surface of the water dimer determined in previous studies<sup>108</sup> to investigate the accuracy of the intermolecular electrostatic energy at the B3LYP/aug-cc-pVTZ level. With respect to QC, a first striking result is that the values of the Coulomb interaction energy are notably improved compared to those from distributed multipoles (without  $E_{\text{pen}}$ )<sup>29</sup> obtained at the same level of theory. If we compare the results to the CSOV

references values, we can see in Table 7 that the penetration energy is recovered by the molecular mechanics as in our previous study.<sup>103</sup> The average absolute error of the ten configurations is 0.089 kcal/mol. The transferability of the auxiliary coefficients is demonstrated, and each of the dimers is correctly described. Regarding the reproduction of reference exchange-repulsion energies, the results were encouraging. They showed the robustness of the overlap model and were strongly correlated to reference B3LYP ab initio calculations of  $E_{\text{exch}}$  with a correlation factor of 0.9986 as displayed in Figure 10. The model has an average absolute error of 0.12 kcal/mol as shown in Table 8.

For all ten water dimers, close agreements were similarly found concerning the polarization and charge-transfer contributions (parts a and b, respectively, of Table 9), for which average absolute errors of 0.096 and 0.097 kcal/mol were found with respect to CSOV.

The final step consisted of comparisons of the sums of the GEM-0 energy components to the corresponding DFT interaction energies. For the ten water dimers, and with

**Table 8.** Intermolecular Exchange-Repulsion Energies (in kcal/mol) for 10 Water Dimer Geometries for GEM-0 Fitted on B3LYP (or CCSD-BD)/aug-cc-pVTZ Densities vs CSOV at the B3LYP/aug-cc-pVTZ Level<sup>a</sup>

level of theory for $E_{\text{exch-rep}}$	water dimer geometry									
	1	2	3	4	5	6	7	8	9	10
CSOV (DFT)	6.84	5.63	5.37	5.08	4.22	3.85	3.59	1.18	3.59	1.89
GEM-0 (DFT)	6.84	5.71	5.36	4.86	3.95	3.95	3.94	1.27	3.64	1.97

level of theory for $E_{\text{exch-rep}}$	water dimer geometry									
	1	2	3	4	5	6	7	8	9	10
SAPT(CCSD)	8.01	6.62	6.31	6.12	5.06	4.64	4.26	1.28	4.35	2.22
GEM-0 (CCSD)	8.05	6.76	6.32	5.77	5.01	4.75	4.54	1.24	4.14	2.19

<sup>a</sup> GEM-0 results in parentheses are exchange-repulsion energies obtained with GEM-0 using auxiliary coefficients obtained by averaging fits of the density using a  $10^{-10}$  cutoff (ref 27a).

**Table 9.** (a) Polarization Energies (kcal/mol) for Ten Water Dimer Geometries for the GEM-0 Approach Fitted on B3LYP/augcc-pVTZ Densities Compared to CSOV B3LYP/augcc-pVTZ Results (Ref 27a) and (b) Charge-Transfer Energies (kcal/mol) for Ten Water Dimer Geometries for the GEM-0 Approach Fitted on B3LYP/augcc-pVTZ Densities Compared to CSOV B3LYP/augcc-pVTZ Results (Ref 27a)

(a)										
level of theory for $E_{\text{pol}}$	water dimer geometry									
	1	2	3	4	5	6	7	8	9	10
CSOV/DFT	-1.33	-1.14	-1.12	-0.69	-0.64	-0.62	-0.37	-0.12	-0.44	-0.28
GEM-0/DFT	-1.22	-1.03	-0.92	-0.55	-0.53	-0.50	-0.27	-0.08	-0.42	-0.29

(b)										
level of theory for $E_{\text{ct}}$	water dimer geometry									
	1	2	3	4	5	6	7	8	9	10
CSOV	-1.77	-1.48	-1.42	-0.96	-0.80	-0.68	-0.53	-0.20	-0.54	-0.26
GEM-0/SIBFA	-1.86	-1.42	-1.31	-0.94	-0.73	-0.63	-0.44	-0.11	-0.56	-0.29

**Table 10.** Total Interaction Energies (kcal/mol) for Ten Water Dimer Geometries for the GEM-0 Approach Fitted on B3LYP/aug-cc-pVTZ Densities Compared to CSOV B3LYP/aug-cc-pVTZ Results Corrected from the Basis Set Superposition Error (Ref 27a)

level of theory for $\Delta E_{\text{int}}$	water dimer geometry									
	1	2	3	4	5	6	7	8	9	10
CSOV	-4.39	-3.82	-3.80	-3.38	-3.00	-2.91	-2.36	-0.78	-2.30	-1.56
GEM-0/SIBFA	-4.30	-3.71	-3.28	-3.32	-3.13	-2.88	-1.98	-0.45	-2.29	-1.59

respect to the BSSE-corrected CSOV total interaction energies, a 0.16 kcal/mol average absolute error was obtained, limited to 0.038 kcal/mol in terms of the relative average error (Table 10). Such a result thus confirms this methodology to reproduce realistic interactions.

We have also applied the model to 16–64 water clusters, as extracted from Monte Carlo simulations in ice or in bulk water that resorted to SIBFA. In all cases, the accuracy of the method appears very good. Thus for the 16, 20, and 64 water clusters, the values of the Coulomb interaction energy amounting to -186.84, -309.38, and -449.52 kcal/mol compare closely to the corresponding CSOV values of -186.38, -307.20, and -446.12 kcal/mol, respectively (Table 11). For the exchange-repulsion energies, the model also performs very well with errors below 1% (Table 11). The polarization and charge-transfer terms have also close agreements with available QC results (Table 12). Therefore, in order to evaluate the overall accuracy of our model, we

**Table 11.** Coulomb and Exchange-Repulsion Intermolecular Interaction Energies (kcal/mol) for Water Clusters ( $n = 16, 20, 64$ ) for the GEM Approach Fitted on B3LYP (or CCSD)/aug-cc-pVTZ vs ab Initio CSOV/B3LYP/aug-cc-pVTZ Values<sup>a</sup>

$n$	$E_{\text{coulomb}}$ GEM-0 (DFT)	$E_{\text{coulomb}}$ CSOV (DFT)	$E_{\text{coulomb}}$ GEM-0 (CCSD)	$E_{\text{exch-rep}}$ GEM-0 (DFT)	$E_{\text{exch-rep}}$ CSOV (DFT)
16	-186.84	-186.38	-184.80	164.95	166.54
20	-309.38	-307.20	-305.84	292.25	292.16
64	-449.52	-446.12	-443.54	336.48	NC

<sup>a</sup> NC = not computed (ref 27a).

computed the total BSSE-corrected interaction energies at the same level of theory for the 16 and 20 molecule clusters. Relative errors of +3.16 out of -114.02 kcal/mol and of -3 out of -168.1 kcal/mol were found for these two respective clusters, confirming the good transferability of the



**Table 12.** Polarization Energies (kcal/mol) for Water Clusters ( $n = 16, 20, 64$ ) for the GEM-0 Approach Fitted on B3LYP /aug-cc-pVTZ vs ab Initio CSOV/B3LYP/ aug-cc-pVTZ Values<sup>a</sup>

$n$	$E_{\text{pol}}$ two-body GEM-0	$E_{\text{pol}}$ two-body CSOV	$E_{\text{pol}}$ GEM-0 ( $E_{\text{pol}}$ GEM-0 initial guess)	$E_{\text{pol}}$ KM/HF ( $E_{\text{pol}}$ RVS/HF)
16	−30.75	−31.03	−48.53 (−36.82)	−45.11 (−35.50)
20	−47.53	−48.01	−82.79 (−62.60)	−78.6 (NC)
64	−57.97	NC	−77.89 (−64.78)	NC (NC)

<sup>a</sup> For the GEM-0 column, results in parentheses correspond to the polarization energy of the first set of induced dipoles. RVS polarization results are given in parentheses in the KM column. Both are computed at the CCP 4-31G(2d) level. NC = not computed (ref 27a).

different approximations. GEM-0 has been also tested for metals for electrostatic and exchange-repulsion contributions. Accurate results are obtained even at a very short range.<sup>27a</sup> It is important to point out that this density fitting procedure is not limited to Hartree–Fock or DFT energies. Thus Tables 7 and 8 had also shown close agreements<sup>27a</sup> between GEM fitted on relaxed CCSD-Bruckner-Double densities and reference CCSD SAPT calculations.

(C) *Extension to Higher Angular Momenta, Computational Speedup, and Periodic Boundary Conditions.* At this point, GEM-0 showed a very good accuracy but requires several nonatomic centers as it uses s-type ( $l = 0$ ) Gaussian functions only. In order to reduce the number of sites, an extension of the formalism to higher angular momenta ( $l > 0$ ) was required.

(a) *Extension to Higher Angular Momenta: Accuracy of Forces and Energies.* One advantage of using fitted densities expressed in a linear combination of Gaussian functions is that the choice of Gaussian functions for the ABS needs not be restricted to Cartesian Gaussians. In order to extend GEM to higher order angular momenta,<sup>27b</sup> we have chosen to use normalized *Hermite Gaussian functions* for the calculation of the intermolecular interactions. Thus, the use of Hermite Gaussians in the calculation of the intermolecular interactions results in improved efficiency by the use of the McMurchie-Davidson (McD) recursion<sup>106</sup> since the expensive Cartesian-Hermite transformation is avoided. Obtaining the Hermite expansion coefficients from the fitted Cartesian coefficients is straightforward since Hermite polynomials form a basis for the linear space of polynomials.

We have also implemented noise reduction techniques<sup>27b</sup> for the fitting procedure in addition to the already discussed cutoff in the eigenvalues during the SVD procedure. Indeed, this method produces undesirable numerical instabilities (noise) when the number of basis functions starts to grow with Gaussian functions as commented above. In addition, we have observed that these instabilities are also present when using only s-type spherical functions<sup>27a</sup> albeit to a lower extent. In the present implementation we have opted to use the Tikhonov regularization formalism. Additionally, Jung et al.<sup>109</sup> have recently shown that the use of a damped Coulomb operator  $\hat{O} = \text{erfc}(\beta r)/r$  can be used for the fitting procedure. These authors have employed this kernel to localize the integrals in order to increase the calculation speed of three-center Coulomb integrals in a quantum mechanical

**Table 13.** Relative rms Force Deviation with Respect to CSOV for the Ten Water Dimers (Ref 27b)

level of theory	6-31G*			aug-cc-pVTZ		
	A1	P1	G03	A1	P1	G03
Coulomb	0.06	0.03	0.01	0.15	0.04	0.05
exchange	0.22	0.08	0.07	0.12	0.04	0.07

program. For our purposes, the implemented damped Coulomb operator could be employed to attenuate the near-singular behavior due to long-range interactions.<sup>27b</sup>

With such procedures, Coulomb and exchange-repulsion have been calculated with higher angular momenta that allow for a reduction of the number of sites compared to GEM-0 for the ten water dimers as well as representative benzene dimers. Excellent agreement was obtained in all cases for the intermolecular interactions with errors below 0.1 kcal/mol for electrostatic and around 0.15–0.2 kcal/mol for exchange repulsion.<sup>27b</sup> In practice in the MD community, the measure of the accuracy has been the forces since this is the quantity that determines the trajectories. Upon using GEM with three ABSs, such an accuracy could be evaluated by comparing the calculated GEM forces with those obtained with CSOV using the finite difference method.<sup>27b</sup> For both Coulomb and overlap interactions, Table 13 shows that for the ten water dimers small rms deviations are observed between forces calculated with A1, P1,<sup>110</sup> and g03 ABSs compared to the CSOV forces computed at the B3LYP level with the same basis sets, namely 6-31G\* and aug-cc-pVTZ. The errors in the exchange-repulsion forces are also very satisfactory considering the simplicity of the overlap model compared to ab initio.

(b) *From Densities to Site Multipoles: A Continuous Electrostatic Model.* Challacombe et al.<sup>111</sup> have shown that Hermite Gaussians have a simple relation to elements of the Cartesian multipole tensor. Expanding on that work, once the Hermite coefficients have been determined, they may be employed to calculate multipoles centered at the expansion sites.<sup>27b</sup> Thus we have been able to obtain distributed multipoles centered at the ABS's sites that connect naturally with an accurate evaluation of the exact Coulomb interaction energy. This connection will be useful for the direct use of such multipoles into SIBFA as well as in the generation of damping functions<sup>39a</sup> that accounts for the penetration error when using these multipoles. Unlike conventional multipole expansions, the spherical multipole expansion obtained from Hermite Gaussians has an intrinsic finite order, namely, the highest angular momentum in the ABS. This is thus similar to the multipolar expansions derived by Volkov and Cop-pens.<sup>112</sup>

This connection between multipoles and Hermite densities is important. Indeed, unlike s-type functions ( $l = 0$ ), fitting coefficients with  $l > 0$  (sp, spd ...) are not invariant by rotation. These coefficients must be transformed for each molecular fragment orientation in order to compute interaction energies. Such a transformation can be achieved by defining both a *global* orthogonal coordinate system frame and a *local* orthogonal coordinate frame for each fragment fitting site. Hermite Gaussians in the two coordinate systems

can be related using the chain rule.<sup>27b</sup> Such a method has been previously developed for point dipoles and generalized to higher order multipoles.<sup>113</sup> These frame definitions are similar to those in the OPEP code<sup>16a</sup> and could be applied to SIBFA as well. It is important to point out that the same chain rule approach works also for the transformation to scaled fractional coordinates which will be important toward the extension to PBC where the use of Particle Mesh Ewald (PME)<sup>114</sup> requires that the coefficients be transformed to scaled fractional coordinates.

(c) *Increasing Computational Efficiency Using Reciprocal Space Methods.* Additionally, a significant computational speedup can be achieved using reciprocal space methods.<sup>27b</sup> Indeed, it is possible to split the integrals required for the frozen-core contribution into direct and reciprocal space contributions.

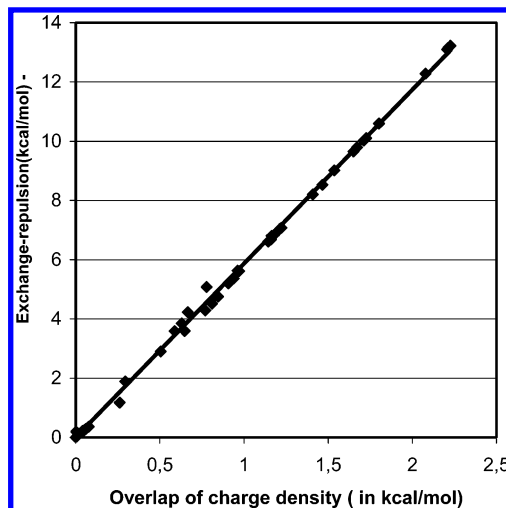
The direct sum corresponds to full computation of integrals between two centers at a distance below a chosen cutoff. Such integrals are computed using a generalized McD recursion applicable to Gaussian derivatives of any smooth function of  $r$  and so thus to all the direct space integrals used in this study, i.e., overlap, Coulomb, and damped Coulomb.<sup>27b</sup>

The rest of the integrals are treated using reciprocal space. Three methods were implemented: regular Ewald, Particle Mesh Ewald (PME),<sup>114</sup> and Fast Fourier Poisson (FFP)<sup>115</sup> Denoting by  $N$  the number of molecules, since the regular Ewald approach scales as  $N^2$ , the use of fast Fourier transformations (FFT) is necessary to improve the scaling and reach  $N \log(N)$ .

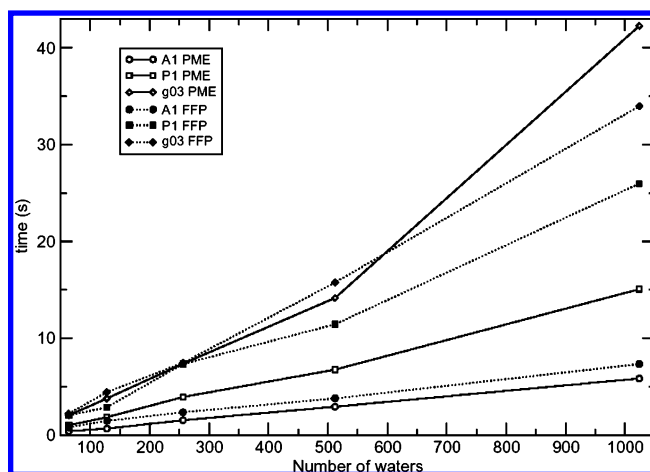
PBC GEM implementation with reciprocal space methods has been tested by calculating the intermolecular Coulomb energies and forces for a series of water boxes  $[\text{H}_2\text{O}]_N$ ,  $N = 64, 128, 256, 512$ , and 1024. The reciprocal space methods are quite efficient. The calculation of the energies and forces for the largest system was tested at the highest accuracy, i.e., the 1024 water box with a very extended g03 ABS which corresponds to 654 336 Hermite coefficients (located on atoms, bonds, and lone pairs) and takes only 34 s with FFP and 42 s with PME (see Figure 11) using rms accuracies of  $10^{-4}$  on a dual Xeon 3.3 Ghz processor.<sup>27b</sup> It is further noted that both reciprocal space methods are highly parallelizable, which would increase computational efficiency.

Moreover, we have recently shown that thanks to a numerical approach to the Hermite fitting<sup>116</sup> using molecular properties calculated on grids as well as an improved splitting procedure for the compact and diffuse functions it was possible to improve the accuracy and so to diminish the number of auxiliary functions. For example, it has been possible to use the small A2 basis set restricted to atoms and to perform a calculation on a 4096 water box GEM calculation fitted on a B3LYP/6-31G\* reference level. Such calculation took 2.6 s on a *single processor*, which is about an order of magnitude slower than the corresponding point charges Amber calculations which took 0.2 s on the same computer.

(D) *QM/MM and Future Developments.* A QM/MM implementation has been recently performed<sup>104</sup> using GEM as the MM force field. This method has been used, parallel

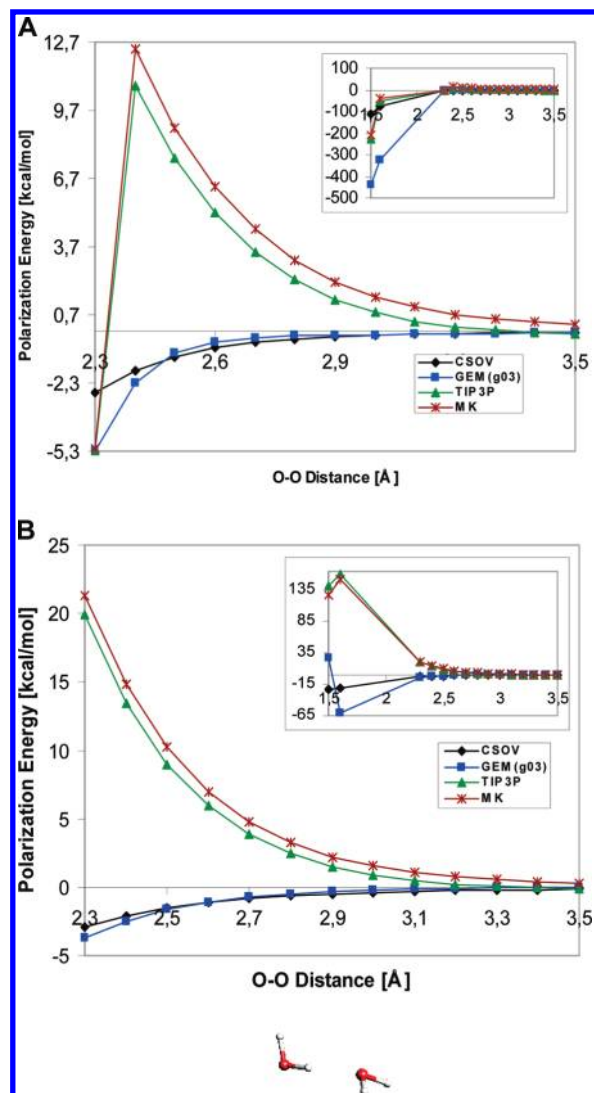


**Figure 10.** Correlation of the overlap of charge density (kcal/mol) computed with GEM-0 vs exchange-repulsion energy (kcal/mol) obtained at a CSOV/B3LYP/aug-cc-pVTZ level for 200 orientations of the water dimer.



**Figure 11.** Timings for water boxes with rms force tolerance of 10-4. Closed circles: A1 PME; closed squares: P1 PME; closed diamonds: G03 PME; open circles: A1 FFP; open squares: P1 FFP; open diamonds: G03 FFP.

to conventional QM/MM using point charges, to evaluate the polarization on the QM subsystem by the MM environment for the ten water dimers. GEM was found to give the correct polarization response compared to reference CSOV polarization energies. By contrast, point charges produced significant underpolarization of the QM subsystem, in several cases actually presenting an opposite sign of the polarization contribution (see Figure 12). This approach prefigures a prospective multilevel implementation of a SIBFA/GEM/QM strategy. Indeed, it is important to mention that results obtained with both PME and FFP can be mixed. This opens up novel possibilities for QM/MM implementation: thus the GEM section proximal to the QM could be calculated with PME or FFP, the remaining MM subsystem could be represented via GEM multipoles, and these could be used for SIBFA and calculated in reciprocal space using PME. As most of the gradients of the SIBFA energy function are available, this opens up the possibility of long condensed-phase SIBFA MD/PBC. Such an implementation<sup>8b</sup> in



**Figure 12.** Polarization of the QM water molecule in the geometry of the linear water dimer at various distances for a QM/MM calculation using GEM (molecule A = QM, top; molecule B = QM, bottom). Inset shows a range from 1.5 to 3.5 Å.

AMBER 9.0 was recently achieved by some of us for AMOEBA. We plan to perform it in the context of SIBFA as well.

*Present Status of the Software.* At this point it is necessary to mention some present possibilities and limitations of the SIBFA software

(a) *Timings.* An in vacuo single-point computation on the complex of the 5-PAH inhibitor with a PMI model encompassing 164 residues (about 2700 atoms totaling about 8000 centers) requires about 3 min CPU time on a single-processor IBM sp4 computer (there are no cutoffs for the energy computations). Merlin can resort to nongradient minimizers such as ROLL or SIMPLEX or to numerical evaluations of the gradients using the BFGS, the Davidson-Fletcher-Powell, or the Conjugate Gradients Algorithms (see ref 61 for details). For most applications, we found the ROLL algorithm as the most effective, although it entails a significantly larger number of energy evaluations. Energy minimizations on about 200–300 internal variables requiring about 5000

energy computations thus take about 10 days on a single processor. They are postprocessed for one or two additional rounds to ensure for convergence of the energy. Subsequent energy minimizations encompassing  $\Delta G_{\text{solv}}$  are about 6-fold more time-consuming but can now be done on a version of the code that parallelizes on four to eight processors.

(b) *Availability of the Gradients.* Presently, most analytical gradients have been coded and checked. The principal gradients presently not available are those of  $E_{\text{ct}}$  and of  $\Delta G_{\text{solv}}$ . The coding of  $E_{\text{ct}}$  is underway and will be reported shortly. In the present context, a simplified version of  $E_{\text{pol}}$  and its gradients has been coded for which, similar to the  $\Delta G_{\text{solv}}$  computations, scalar instead of tensor polarizabilities are used. The availability of the analytical gradients should enable for more efficient searches of the potential energy surface although with a simpler energy function, since the minima could be reprocessed for a last round with the complete function. This availability has also enabled us to start preliminary MD simulations with the simplified SIBFA potential. These will be used to locate alternative docking modes in ligand–macromolecule complexes. On the other hand, however, condensed-phase simulations will depend upon the merging of SIBFA with PME and/or GEM methodologies as discussed below.

*Present Scope of Applications.* Several ongoing applications of SIBFA bear on the docking of inhibitors with protein targets and are carried out in close collaboration with experimentalists. While Zn-metalloproteins constitute a privileged target, the extension to other targets, such as signaling proteins and kinases, is underway. The size of the target proteins can encompass up to 200 amino acid residues. Optimization of the code to handle larger systems is underway, including its porting to Fortran 90 and parallelization.

## Conclusions and Perspectives

The availability of energy-decomposition analyses of QC intermolecular interactions is essential for the development of APMM procedures. We have shown in this review that the separable SIBFA potential can reproduce the anisotropy and nonadditivity features of  $\Delta E_{\text{int}}(\text{QC})$  and of its contributions. A particularly challenging test was provided by binuclear Zn(II) complexes, as in the binding site of bacterial metallo- $\beta$ -lactamase (MBL).<sup>28c</sup>  $\Delta E_{\text{int}}(\text{SIBFA})$  could closely reproduce  $\Delta E(\text{QC})$  and the contrasting behaviors of  $E_1$ , on the one hand, and of  $E_{\text{pol}}$  and  $E_{\text{ct}}$ , on the other hand, in two structurally very distinct and competing arrangements. Multipole transferability is a critical issue in order to be able to handle flexible molecules. We have shown that the separable character of the potential, encompassing both polarization and charge transfer, were necessary to compute intra- and intermolecular interactions of a flexible molecule assembled from rigid fragments. This was illustrated in two extreme cases, divalent cation binding by triphosphate and mercapcarboxamides, on the one hand, and the conformational energies of ten Ala and Glu tetramers, on the other hand. SIBFA has been applied to investigate inhibitor binding to Zn-dependent metalloenzymes. Two recent examples are MBL and phosphomannoisomerase (PMI).<sup>91,92,97</sup> Energy



balances were performed including the contribution of continuum  $\Delta G_{\text{solv}}(\text{LC})$  for different inhibitors in several competing arrangements. Validations by parallel QC computations were done on model binding sites of MBL and PMI totaling up to 140 atoms. Twenty and nine complexes were thus evaluated in these respective sites. The evolutions of the SIBFA interaction energies paralleled the QC ones, with relative errors <3%. The last application bore on a nonenzymatic Zn-metalloprotein, the HIV-1 nucleocapsid (NCp7), a novel target for the design of new-generation anti-HIV inhibitors. One of the low-energy minima had the nucleophilic S-connected carbonyl group at an appropriate distance (3.4 Å) and orientation from Cys36 S<sup>−</sup> to initiate covalent bond formation followed by Zn-ejection.  $E_{\text{pol}}$  and  $E_{\text{disp}}$  were the main contributors to the final energy balances, while  $E_1$  was destabilizing. The SIBFA-derived complexes are being reprocessed by QM/MM procedures, indicating the connectedness between classical MM, APMM, and QM. SIBFA is being extended to a diversity of metal cations. Such extensions benefit from the integration<sup>58</sup> of Ligand Field (LF) effects, on the one hand, and the availability of energy-decomposition procedures<sup>25</sup> and the possibility of quantifying correlation as well as relativistic effects,<sup>117,118</sup> on the other hand. The coupling with Particle Mesh Ewald (PME) methodologies<sup>27b,113,8b</sup> should significantly widen its scope toward large macromolecular complexes and condensed phase. The interface with GEM,<sup>27</sup> which can itself be coupled to QM<sup>104</sup> should give rise in the near future to a multilevel QM/GEM/SIBFA methodology since GEM offers a direct connection between multipoles and densities. This approach could be applied to biomolecular systems such as 4-oxalocrotonate tautomerase.<sup>119,120</sup> GEM offers increased accuracy and full separability of its components as well as improved cooperative effects by the inclusion of native short-range quantum effects. Finally GEM Hermite Gaussian densities can be derived for any element of the periodic classification where ab initio relaxed densities at Hartree–Fock, post Hartree–Fock, or DFT levels are available. In two forthcoming papers, we describe new fitting improvements for the hermites as well as the generalized energy function for small molecules and flexible peptides.

**Abbreviations.** 5-PAH, 5-phospho-D-arabinohydroxamate; 5-PAA, 5-phospho-D-arabinonate; ABS, auxiliary basis set; AMBER, assisted model building with energy refinements; AMOEBA, atomic multipole optimized energetics for biomolecular applications; AOM, angular overlap model; APMM, anisotropic polarizable molecular mechanics; ASP-W, anisotropic site potential for water; ATP, adenosine triphosphate; A1,A2, DGauss DFT Coulomb fitting auxiliary basis set; B3LYP, Becke–Lee–Yang Parr functional; BVWN, Becke–Vosko–Wilk–Nusair functional; CEP 4-31(2d), coreless effective potential double-zeta and two 3d polarization functions on heavy atoms; CNDO, complete neglect of differential overlap; CSOV, constrained space orbital variations; DF, density fitting; DFT, density functional theory; EFP, effective fragment potential; EM, energy minimization; FFP, fast Fourier Poisson; GEM, Gaussian electrostatic model; G03, automatically generated Gaussian 03 Coulomb fitting auxiliary basis set; HF, Hartree–Fock; HPPK, dihy-

dropter pyrophosphokinase; KM, Kitauro–Morokuma energy-decomposition procedure; LACV3P\*\*, Los Alamos compact valence potentials; LCAO, linear combination of atomic orbitals; LF, ligand field effects; LMO, localized molecular orbitals; LMP2, localized Moller–Plesset 2; MBL, metallo- $\beta$ -lactamase; MC, Monte-Carlo; McD, McMurchie–Davidson; MD, molecular dynamics; MM, molecular mechanics; MO, molecular orbitals; MP2, Moller–Plesset 2; NCp7, nucleocapsid protein from HIV-1 retrovirus; NEMO, nonempirical molecular orbital; OPEP, optimally partitioned electric properties; PBC, periodic boundary conditions; PME, particle mesh Ewald; PMI, phosphomannoisomerase; PMM, polarizable molecular mechanics; Pvtz, polarized valence triple zeta; QC, quantum chemistry; QM, quantum mechanics; QP, quadrupolar polarizability; RMS, root mean square; RVS, restricted variational space; SAPT, symmetry-adapted perturbation theory; SDFF, spectroscopically determined force field; SIBFA, sum of interactions between fragments ab initio computed; TCPE, topological and classical many-body polarization effects.

Inception and developments of SIBFA were done in close collaboration with Drs. Pierre Claverie, Alberte Pullman, Jacqueline Langlet, and Claude Giessner-Prettre. We also wish to thank Drs. Walter J. Stevens, Morris Krauss, David R. Garmer, Lee G. Pedersen, and Dennis R. Salahub, for very stimulating collaborations. One of us (N.G.) thanks Benoit de Courcy for his help during preparation of the manuscript. We thank the computer centers of CINES (Montpellier, France), CRIHAN (Rouen, France), CCRE (Paris, France), and NIEHS (RTP, NC) and the Advanced Biomedical Computer Center, NCI-FCRDC, U.S.A. This research was supported in part by CNRS (France), the Ligue Nationale contre le Cancer (Comité Ile-de-France), and by the Intramural Research Program of the NIH and NIEHS (U.S.A.).

**Supporting Information Available:** Zn-methanethiolate, water-formate, and stacked formamide complexes, interaction energies for Zn(II) complexes with water ligands, interaction of triphosphate with Zn(II) probe cation, and interaction of triphosphate anion with the binding site of HPPK kinase. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236.
- (2) Rick, S. W.; Stuart, S. J. *Rev. Comput. Chem.* **2002**, *18*, 89.
- (3) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27.
- (4) McKerell, J. J. *Comput. Chem.* **2004**, *25*, 1584.
- (5) Gresh, N. *Curr. Pharm. Des.* **2006**, *12*, 2121.
- (6) (a) Pullman, B.; Claverie, P.; Caillet, J. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 1663. (b) Rein, R.; Claverie, P.; Pollack, M. *Int. J. Quantum Chem.* **1968**, *2*, 1129. (c) Claverie, P.; Rein, R. *Int. J. Quantum Chem.* **1969**, *3*, 537.



- (7) (a) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227. (b) Lee F. S.; Chu Z. Y.; Warshel A. *J. Comput. Chem.* **1993**, *14*, 161, and references therein. (c) Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283.
- (8) (a) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933, and references therein. (b) Piquemal, J. P.; Perera, L.; Cisneros, G. A.; Ren, P.; Pedersen, L. G.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*, 054511.
- (9) Gresh, N.; Claverie, P.; Pullman, A. *Theor. Chim. Acta* **1984**, *66*, 1.
- (10) (a) Day, P.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D. R.; Basch, H.; Cohen, D. *J. Chem. Phys.* **1996**, *105*, 1968. (b) Slipchenko, L. V.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276.
- (11) (a) Price, S. L.; Stone, A. J. *Mol. Phys.* **1984**, *51*, 569. (b) Price, S. L.; Stone, A. J. *J. Chem. Soc., Faraday Soc.* **1992**, *88*, 1755.
- (12) Millot, C. J.; Soetens, J. C.; Martins, Costa, N. T. C.; Hodges, M. P.; Stone, A. J. *J. Phys. Chem. A* **1998**, *102*, 754.
- (13) Mannfors, B.; Mirkin, N. G.; Palmo, K.; Krimm, S. *J. Comput. Chem.* **2001**, *22*, 1933.
- (14) Hermida-Ramón, J. M.; Brdarski, S.; Karlström, G.; Berg, U. *J. Comput. Chem.* **2003**, *24*, 161.
- (15) (a) Gagliardi, L.; Lindh, R.; Karlstrom, G. *J. Chem. Phys.* **2004**, *121*, 4494. (b) Soderhjelm, P.; Krogh, J. W.; Karlstrom, G.; Ryde, U.; Lindh, R. *J. Comput. Chem.* **2007**, *28*, 000.
- (16) (a) Angyan, J. G.; Chipot, C.; Dehez, F.; Hattig, C.; Jansen, G.; Millot, C. *J. Comput. Chem.* **2003**, *24*, 997. (b) Popelier, P. L. A.; Joubert, L.; Kosov D. S. *J. Phys. Chem. A* **2001**, *105*, 8254. (c) Bader, R. F. W. *Atoms in Molecules: a Quantum Theory*; Oxford University Press: 1990.
- (17) Case, D. A.; Cheatham, T. E., II; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *26*, 1668.
- (18) (a) Masella, M.; Flament, J.-P. *J. Chem. Phys.* **1997**, *107*, 9105. (b) Masella, M.; Flament, J.-P. *Mol. Phys.* **1998**, *95*, 97. (c) Cuniasse, P.; Masella, M. *J. Chem. Phys.* **2003**, *119*, 1874.
- (19) (a) Langlet, J.; Claverie, P.; Caron, F.; Boeue, J.-C. *Int. J. Quantum Chem.* **1981**, *19*, 299. (b) Derepas, A. L.; Soudan, J. M.; Brenner, V.; Dognon, J. P.; Millié, P. *J. Comput. Chem.* **2002**, *23*, 1013.
- (20) Dang, L. X.; Chang, T. M. *J. Chem Phys.* **1997**, *106*, 8149.
- (21) (a) Rappé, A. K.; Goddard, W. A., III *J. Phys. Chem.* **1991**, *95*, 3358. (b) Liu, Y. P.; Kim, K.; Berne, B. J.; Friesner, R. A.; Rick, S. W. *J. Chem. Phys.* **1998**, *108*, 4739. (c) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741. (d) Chelli, R.; Procacci, P. *J. Chem. Phys.* **2002**, *117*, 9175.
- (22) (a) Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185. (b) Yu, H.; Hansson, T.; van Gunsteren, W. L. *J. Chem. Phys.* **2003**, *118*, 221. (c) Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1587.
- (23) (a) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341. (b) van Duijnen, P. T.; Swart, M. J. *J. Phys. Chem. A* **1998**, *102*, 2399. (c) Chelli, R.; Procacci, P. *J. Chem. Phys.* **2002**, *117*, 9175. (d) Piquemal, J. P.; Chelli, R.; Procacci, P.; Gresh, N. *J. Phys. Chem. A* **2007**, *111*, 8170. (e) Elking, D.; Darden, T. A.; Woods R. J. *J. Comput. Chem.* **2007**, *28*, 1261.
- (24) Gresh, N. *J. Chim.-Phys. Chim. Biol.* **1997**, *94*, 1365.
- (25) (a) Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325. (b) Stevens, W. J.; Fink W. *Chem. Phys. Lett.* **1987**, *139*, 15. (c) Bagus, P. S.; Hermann, K.; Bauschlicher, C. W., Jr. *J. Chem. Phys.* **1984**, *80*, 4378. (d) Bagus, P. S.; Illas F. *J. Chem. Phys.* **1992**, *96*, 896. (e) Piquemal, J.-P.; Márquez, A.; Parisel, O.; Giessner-Prettre, C. *J. Comput. Chem.* **2005**, *26*, 1052.. (f) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887. (g) Langlet, J.; Caillet, J.; Bergès, J. Reinhardt, P. *J. Chem. Phys.* **2003**, *118*, 6157.
- (26) Faerman, C. H.; Price, S. L. *J. Am. Chem. Soc.* **1990**, *112*, 4915.
- (27) (a) Piquemal, J. P.; Cisneros, A.; Reinhardt, P.; Gresh, N.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*, 104101. (b) Cisneros, A.; Piquemal, J. P.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*, 184101.
- (28) (a) Gresh, N.; Claverie, P.; Pullman, A. *Int. J. Quantum Chem.* **1986**, *29*, 101. (b) Gresh, N. *J. Comput. Chem.* **1995**, *16*, 856. (c) Gresh, N.; Leboeuf, M.; Salahub, D. R. In *Modelling the Hydrogen Bond*; ACS Symposium Series 569; Smith, D. A., Ed.; 1994; p 82. (d) Gresh, N.; Guo, H.; Kafafi, S. A.; Salahub, D. R.; Roques, B. P. *J. Am. Chem. Soc.* **1999**, *121*, 7885. (e) Gresh, N.; Piquemal, J.-P.; Krauss, M. *J. Comput. Chem.* **2005**, *26*, 1113.
- (29) Vigné-Maeder, F.; Claverie, P. *J. Chem Phys.* **1988**, *88*, 4934.
- (30) (a) Dreyfus, M. Ph.D. Thesis, University of Paris, 1970. (b) Claverie, P. Ph.D. Thesis, Paris, 1973, CNRS library number A.O. 8214. (c) Claverie, P. In *Localization and Delocalization in Quantum Chemistry*; Chalvet, O., Daudel, R., Diner, S., Malrieu, J. P., Eds.; Reidel: Dordrecht, Vol. II, p 127.
- (31) (a) Rein, R.; Rabinowitz, J. R.; Swisler, T. J. *J. Theor. Biol.* **1972**, *34*, 215. (b) Rein, R. *Adv. Quantum Chem.* **1973**, *7*, 335.
- (32) Gresh, N.; Claverie, P.; Pullman, A. *Int. J. Quantum Chem.* **1979**, *Symp. 11*, 253.
- (33) Gresh, N. *Biochim. Biophys. Acta* **1980**, *597*, 345.
- (34) Gresh, N.; Pullman, B. *Biochim. Biophys. Acta* **1980**, *625*, 356.
- (35) Gresh, N.; Pullman, B. *Biochim. Biophys. Acta* **1980**, *608*, 47.
- (36) Gresh, N.; Etchebest, C.; de la Luz, Rojas, O.; Pullman, A. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1981**, *8*, 109.
- (37) Gresh, N.; Pullman, A. *Int. J. Quantum Chem.* **1982**, *22*, 709.
- (38) Gresh, N.; Pullman, A. *Int. J. Quantum Chem., Quant. Biol. Symp.* **1983**, *10*, 215.
- (39) (a) Piquemal, J.-P.; Gresh, N.; Giessner-Prettre, C. *J. Phys. Chem. A* **2003**, *107*, 10353. (b) Piquemal, J.-P.; Chevreau, H.; Gresh, N. *J. Chem. Theory Comput.* **2007**, *3*, 824.
- (40) (a) Stone, A. J. *Chem. Phys. Lett.* **1981**, *83*, 233. (b) Stone A. J.; Alderton, M. *Mol. Phys.* **1985**, *56*, 1047.
- (41) Etchebest, C.; Lavery, R.; Pullman, A. *Theor. Chim. Acta* **1982**, *62*, 17.
- (42) (a) Sokalski, W. A.; Poirier, R. A. *Chem. Phys. Lett.* **1983**, *98*, 86. (b) Sokalski, W. A.; Sawaryn, A. *J. Chem. Phys.* **1987**, *87*, 526.

- (43) (a) Karlstrom, G.; Linse, P.; Wallqvist, A.; Jonsson, B. *J. Am. Chem. Soc.* **1983**, *105*, 3777. (b) Andersson, M.; Karlstrom, G. *J. Phys. Chem.* **1985**, *89*, 4957.
- (44) (a) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, A.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293. (b) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. *J. Chem. Phys.* **2003**, *112*, 7300. (c) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.
- (45) Murrell, J. N.; Teixeira-Dias, J. J. C. *Mol. Phys.* **1970**, *19*, 521.
- (46) Garmer, D. R.; Stevens, W. J. *J. Phys. Chem. A* **1989**, *93*, 8263.
- (47) Murrell, J. N.; Randic, M.; Williams, D. R. *Proc. R. Soc. London, Ser. A* **1966**, *284*, 566.
- (48) Gresh, N.; Claverie, P.; Pullman, A. *Int. J. Quantum Chem.* **1982**, *22*, 199.
- (49) Creuzet, S.; Langlet, J.; Gresh, N. *J. Chim.-Phys. Phys. Chim. Biol.* **1991**, *88*, 2399.
- (50) Gresh, N.; Pullman, A.; Claverie, P. *Theor. Chim. Acta* **1985**, *67*, 11.
- (51) (a) Langlet, J.; Claverie, P.; Caillet, J.; Pullman, A. *J. Phys. Chem.* **1988**, *92*, 1617. (b) Langlet, J.; Gresh, N.; Giessner-Prettre, C. *Biopolymers* **1995**, *36*, 765. (c) Huron, M.-J.; Claverie, P.; *J. Phys. Chem.* **1972**, *76*, 2123. (d) Huron, M.-J.; Claverie, P. *J. Phys. Chem.* **1974**, *78*, 1853. (e) Pierotti, R. A. *J. Phys. Chem.* **1965**, *69*, 2813.
- (52) Gresh, N. *Biopolymers* **1985**, *24*, 1527.
- (53) (a) Chen, K.-X.; Gresh, N.; Pullman, B. *Nucleic Acids. Res.* **1986**, *14*, 3799. (b) Chen, K.-X.; Gresh, N.; Pullman, B. *Mol. Pharmacol.* **1986**, *30*, 279. (c) Chen, K.-X.; Gresh, N.; Pullman, B. *Nucleic Acids. Res.* **1988**, *16*, 3061. (d) Gresh, N.; Pullman, B.; Arcamone, F.; Menozzi, M.; Tonani, R. *Mol. Pharmacol.* **1989**, *35*, 251.
- (54) (a) Gresh, N.; Pullman, B. *Mol. Pharmacol.* **1986**, *29*, 355. (b) Gresh, N. *Mol. Pharmacol.* **1987**, *31*, 617.
- (55) (a) Gresh, N.; Pullman, A. *New. J. Chem.* **1986**, *10*, 405. (b) Gresh, N. *New. J. Chem.* **1986**, *11*, 61.
- (56) (a) Gresh, N.; Roques, B.-P. *Biopolymers* **1997**, *41*, 145. (b) Garmer, D. R.; Gresh, N.; Roques, B.-P. *Proteins* **1998**, *31*, 42.
- (57) Gresh, N.; Policar, C.; Giesner-Prettre, C. *J. Phys. Chem. A* **2002**, *106*, 5660.
- (58) Piquemal, J. P.; Williams-Hubbard, B.; Fey, N.; Deeth, R. J.; Gresh, N.; Giessner-Prettre, C. *J. Comput. Chem.* **2003**, *24*, 1963.
- (59) (a) Schäffer, C. E.; Jørgensen, C. K. *Mol. Phys.* **1964**, *9*, 401. (b) Gerloch, M.; Harding, J.-H.; Wooley, R. G. *Struct. Bonding* **1981**, *46*, 1. (c) Bridgeman, A. J.; Gerloch, M. *Prog. Inorg. Chem.* **1997**, *45*, 179. (d) Burton, V. J.; Deeth, R. J.; Kemp, C. M.; Gilbert, P. J. *J. Am. Chem. Soc.* **1995**, *117*, 8407. (f) Deeth, R. J. *Coord. Chem. Rev.* **2001**, *212*, 11.
- (60) Stevens, W. J.; Basch, H.; Krauss, M. *J. Chem. Phys.* **1984**, *81*, 6026.
- (61) Evangelakis, G. A.; Rizos, J. P.; Lagaris, I. E.; Demetropoulos, I. N. *Comput. Phys. Commun.* **1987**, *46*, 401.
- (62) (a) Gresh, N. *J. Phys. Chem. A* **1997**, *101*, 8690. (b) Masella, M.; Gresh, N.; Flament, J. P. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 2745.
- (63) Guo, H.; Gresh, N.; Roques, B. P.; Salahub, D. R. *J. Phys. Chem. B* **2000**, *104*, 9746.
- (64) Hodges, M. P.; Stone, A. J.; Xantheas, S. J. *Phys. Chem. A* **1997**, *101*, 9163.
- (65) Gresh, N.; Garmer, D. R. *J. Comput. Chem.* **1996**, *17*, 1481.
- (66) (a) Tiraboschi, G.; Gresh, N.; Giessner-Prettre, C.; Pedersen, L. G.; Deerfield, D. W. *J. Comput. Chem.* **2000**, *21*, 1011. (b) Tiraboschi, G.; Roques, B. P.; Gresh, N. *J. Comput. Chem.* **1999**, *20*, 1379.
- (67) Axilrod, B. M.; Teller, E. *J. Chem. Phys.* **1943**, *11*, 299.
- (68) Concha, N. O.; Rasmussen, B. A.; Bush, K.; Herzberg, O. *Structure (London)* **1996**, *4*, 623.
- (69) Krauss, M.; Gilson, H. S. R.; Gresh, N. *J. Phys. Chem. B* **2001**, *105*, 8040.
- (70) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.
- (71) (a) Saebo, S.; Pulay, P. *J. Chem. Phys.* **1987**, *86*, 914. (b) Murphy, R. B.; Beachy, M. D.; Friesner, R. A. *J. Chem. Phys.* **1995**, *103*, 1481.
- (72) Rogalewicz, F.; Gresh, N.; Ohanessian, G. *J. Comput. Chem.* **2000**, *21*, 963.
- (73) Tiraboschi, G.; Fournie-Zaluski, M. C.; Roques, B. P.; Gresh, N. *J. Comput. Chem.* **2001**, *22*, 1038.
- (74) Roques, B. P.; Noble, F.; Fournié-Zaluski, M. C.; Beaumont, A. *Pharmacol. Rev.* **1993**, *45*, 88.
- (75) Ledecq, L.; Lebon, F.; Durant, F.; Giessner-Prettre, C.; Marquez, A.; Gresh, N. *J. Phys. Chem. B* **2003**, *107*, 10640.
- (76) Gresh, N.; Shi, G.-B. *J. Comput. Chem.* **2004**, *25*, 160.
- (77) Biasczyk, J.; Shi, G. B.; Yan, H.; Ji, X. *Structure (London)* **2000**, *8*, 1049.
- (78) Gresh, N.; Kafafi, S. A.; Truchon, J.-F.; Salahub, D. R. *J. Comput. Chem.* **2004**, *25*, 823.
- (79) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908. (b) Feyereisen, M. W.; Feller, D.; Dixon, D. A. *J. Phys. Chem.* **1996**, *100*, 2993.
- (80) Becke, A. D. *J. Chem. Phys.* **1988**, *88*, 1053.
- (81) Proynov, E. I.; Sirois, S.; Salahub, D. R. *Int. J. Quantum Chem.* **1997**, *64*, 427.
- (82) Kafafi, S. A.; El-Gharkawy, E. R. H. *J. Phys. Chem. A* **1998**, *102*, 3202.
- (83) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev.* **1988**, *B37*, 785. (b) Becke, A. *J. Chem. Phys.* **1993**, *98*, 5648.
- (84) (a) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741. (b) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515.
- (85) Gresh, N.; Tiraboschi, G.; Salahub, D. R. *Biopolymers* **1998**, *45*, 405.
- (86) Gresh, N.; Sponer, J. *J. Phys. Chem. B* **1999**, *103*, 11415.
- (87) Gresh, N.; Sponer, J. E.; Spacková, N.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2003**, *107*, 8669.

- (88) (a) Derreumaux, P. *J. Chem. Phys.* **1999**, *111*, 2301. (b) Derreumaux, P. *J. Chem. Phys.* **2002**, *117*, 3499.
- (89) Gresh, N.; Derreumaux, P. *J. Phys. Chem. B* **2003**, *107*, 4862.
- (90) Yun, M. R.; Lavery, R.; Mousseau, N.; Zakrzewska, K.; Derreumaux, P. *Proteins: Struct., Genet., Bioinformatics* **2006**, *63*, 967.
- (91) Antony, J.; Gresh, N.; Olsen, L.; Hemmingsen, L.; Schofield, C.; Bauer, R. *J. Comput. Chem.* **2002**, *23*, 1281.
- (92) Antony, J.; Piquemal, J.-P.; Gresh, N. *J. Comput. Chem.* **2005**, *26*, 1131.
- (93) (a) *Accelrys software*; 9645 Scranton Road, San Diego, CA, U.S.A. (b) Wu, H. J.; Roux, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825.
- (94) Concha, N. O.; Janson, C. A.; Rowling, P.; Pearson, S.; Cheever, C. A.; Clarke, B. P.; Lewis, C.; Galleni, M.; Frère, J. M.; Payne, D. J.; Bateson, J. H.; Abdel-Meguid, S. S. *Biochemistry* **2000**, *39*, 4288.
- (95) (a) Payton, M. A.; Rheinneck, M.; Klig, L. S.; DeTiani, M.; Bowden, E. *J. Bacteriol.* **1991**, *173*, 2006. (b) Smith, D. J.; Proudfoot, A. E. I.; De Tiani, M.; Wells, T. N. C.; Payton, M. A. *Yeast* **1995**, *11*, 301. (c) Patterson, J. H.; Waller, R. F.; Jeevarajah, D.; Billman-Jacobe, H.; McConville, M. J. *Biochem. J.* **2003**, *372*, 77. (d) Garami, A.; Ilg, T. *J. Biol. Chem.* **2001**, *276*, 6566. (e) Shinabarger, D.; Berry, A.; May, T. B.; Rothmel, R.; Fialho, A.; Chakrabarty, A. M. *J. Biol. Chem.* **1991**, *266*, 2080.
- (96) Roux, C.; Lee, J. H.; Jeffery, C. J.; Salmon, L. *Biochemistry* **2004**, *43*, 2926.
- (97) Roux, C.; Gresh, N.; Perera, L. E.; Piquemal, J. P.; Salmon, L. *J. Comput. Chem.* **2007**, *28*, 938.
- (98) Cleasby, A.; Wonacott, A.; Skarzynski, T.; Hubbard, R. E.; Davies, G. J.; Proudfoot, A. E. I.; Bernard, A. R.; Payton, M. A.; Wells, T. N. C. *Nature Struct. Biol.* **1996**, *3*, 470.
- (99) (a) Rice, W. G.; Turpin, J. A.; Huang, M.; Clanton, D.; Buckheit, R. W., Jr.; Covell, D. G.; Wallqvist, A.; McDonnell, N. B.; DeGuzman, R. N.; Summers, M. F.; Zalkow, L.; Bader, J. P.; Haugwitz, R. D.; Sausville, E. A. *Nat. Med.* **1997**, *3*, 341. (b) Huang, M.; Maynard, A.; Turpin, J. A.; Graham, L.; Janini, G. M.; Covell, D. G.; Rice, W. G. *J. Med. Chem.* **1998**, *41*, 1371. (c) Goel, A.; Mazur, S. J.; Fattah, R. J.; Hartman, T. L.; Turpin, J. A.; Huang, M.; Rice, W. G.; Appella, E.; Inman, J. K. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 767. (d) Turpin, J. A.; Song, Y.; Inman, J. K.; Huang, M.; Wallqvist, A.; Maynard, A.; Covell, D. G.; Rice, W. G.; Appella, E. *J. Med. Chem.* **1999**, *42*, 67. (e) Srivastava, P.; Schito, M.; Fattah, R. J.; Hartman, T.; Buckheit, R. W.; Turpin, J. A.; Appella, E. *Bioorg. Med. Chem.* **2004**, *12*, 6437.
- (100) Miller, Jenkins, L. M.; Hara, T.; Durell, S. R.; Hayashi, R.; Inman, J. K.; Piquemal, J.-P.; Gresh, N.; Appella, E. *J. Am. Chem. Soc.* **2007**, *129*, 11067.
- (101) Gordon, R. G.; Kim, Y. S. *J. Chem. Phys.* **1972**, *56*, 3122.
- (102) Boys, S. F.; Shavitt, I. A. *Fundamental Calculation of the Energy Surface for the System of Three Hydrogens Atoms*; NTIS: Springfield, VA, 1959; AD212985. (b) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. *J. Chem. Phys.* **1979**, *71*, 4993.
- (103) Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A. *J. Chem. Phys.* **2005**, *123*, 044109.
- (104) Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A. *J. Phys. Chem. B (Letter)* **2006**, *110*, 13682.
- (105) Roothaan, C. C. J. *Rev. Mod. Phys.* **1960**, *23*, 69.
- (106) McMurchie, L. E.; Davidson, E. R. *J. Comput. Phys.* **1978**, *26*, 218.
- (107) (a) Wheatley, R. J.; Price, S. *Mol. Phys.* **1990**, *69*, 50718. (b) Kita S.; Noda K.; Inouye, H. *J. Chem. Phys.* **1976**, *64*, 3446.
- (108) (a) van Duijneveldt-van de Rijdt, J. G. C. M.; Mooij, W. T. M.; van Duijneveldt, F. B. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1169. (b) Tschumper G. S.; Leininger M. L.; Hoffman B. C.; Valeev, E. F.; Quack M.; Schaffer, H. F., III *J. Chem. Phys.* **2002**, *116*, 690.
- (109) Jung, Y.; Sodt, A.; Gill, P. M. W.; Head-Gordon, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6692.
- (110) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. *Can. J. Chem.* **1992**, *70*, 560.
- (111) Challacombe, M.; Schwegler, E.; Almlöf, J. *Computational Chemistry: Review of Current Trends*; World Scientific: Singapore, 1996.
- (112) Volkov, A.; Coppens, P. *J. Comput. Chem.* **2004**, *25*, 921.
- (113) (a) Sagui, C.; Pedersen, L. G.; Darden, T. A. *J. Chem. Phys.* **2004**, *120*, 73. (b) Toukmaji, a.; Sagui, C.; Board, J.; Darden, T. A. *J. Chem. Phys.* **2000**, *113*, 10913.
- (114) (a) Essmann, M.; Perera, L.; Berkowitz, M.; Darden, T. A.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577. (b) Sagui, C.; Pedersen, L. E.; Darden, T. A. *J. Chem. Phys.* **2004**, *120*, 73.
- (115) York, D.; Yang, W. *J. Chem. Phys.* **1994**, *101*, 3298.
- (116) Cisneros, G. A.; Elking, D.; Piquemal, J.-P.; Darden, T. A. *J. Phys. Chem. A* **2007**, DOI: 10.1021/jp074817r.
- (117) Gourlaouen, C.; Piquemal, J.-P.; Parisel, O. *J. Chem. Phys.* **2006**, *124*, 174311.
- (118) Gourlaouen, C.; Piquemal, J. P.; Saue, T.; Parisel, O. *J. Comput. Chem.* **2006**, *27*, 142.
- (119) Cisneros, G. A.; Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2003**, *125*, 10384.
- (120) Cisneros, G. A.; Wang, M.; Silinski, P.; Yang, W.; Fitzgerald, M. C. *Biochemistry* **2004**, *43*, 6885.

CT700134R