https://doi.org/10.15407/ujpe68.10.673

T.YU. NIKOLAIENKO, L.A. BULAVIN
Taras Shevchenko National University of Kyiv
(64/13, Volodymyrs'ka Str., Kyiv 01601, Ukraine)

APPROXIMATING THE ELECTROSTATIC POTENTIAL OF MOLECULES WITH POINT CHARGES MIMICKING THE ELECTRON PAIRS

The electrostatic component used in the traditional force fields significantly impacts their accuracy in modelling the noncovalent interactions peculiar to biomolecular systems, including hydrogen bonding. In this contribution, we present a physical model for approximating the electrostatic potential of a molecule (MEP) based on the first-principle decomposition of its charge density distribution into the localized components. In contrast to conventional schemes, which typically use atom-centered charges to approximate MEP, the proposed approach locates such charges in the positions selected so as to mimic the anisotropy of the electron density distributions related to the electron pairs of atoms or covalent bonds. This peculiarity leads to a more accurate representation of the overall electrostatic potential, as verified by applying the proposed model to approximate the electrostatic component of the intermolecular interaction energy in 145 noncovalently bound molecular complexes from GMTKN55 database. This benchmark showed the root-mean-square difference between the true and approximated values of the electrostatic component of 2.7 kcal/mol, which is 2.2 times lower as compared to the traditional RESP charges method used as a baseline.

Keywords: electrostatic interaction energy, electron charge density, molecular electrostatic potential, atomic charges.

1. Introduction

The accurate determination of the intermolecular interaction energy is essential for obtaining the reliable results in the physical modeling of molecular systems with such methods as molecular dynamics [1–4], Monte Carlo [5–7], molecular docking [8–10] and hybrid QM/MM methods [11]. The core ingredient of such methods is the model of atom-atom potential functions (force fields), which approximates the dependence of the intermolecular interaction energy of molecules on system's geometry with low computational costs. In this model, the intermolecular energy is traditionally divided into additive components attributed to short-range interactions (often modeled by Lennard-Jones type potentials) and electrostatic interactions. The latter is usually approximated with the sum of pairwise interactions between point charges located on the atoms of the molecular system. Particularly, in the case of force fields that do not account for polarization effects explicitly, the charges remain constant throughout the simulation.

Citation: Nikolaienko T.Yu., Bulavin L.A. Approximating electrostatic potential of molecules with point charges mimicking the electron pairs. *Ukr. J. Phys.* **68**, No. 10, 673 (2023). https://doi.org/10.15407/ujpe68.10.673.

Цитування: Ніколаєнко Т.Ю., Булавін Л.А. Апроксимація електростатичного потенціалу молекули точковими зарядами, що імітують її електронні пари. *Укр. фіз. журн.* **68**, №10, 675 (2023).

In spite of their widespread use, the charge of the atom within a larger system is still known to be a ‘fuzzy’ concept which has no precise definition and cannot be computed uniquely, even in cases where quantum mechanics is used to describe the electronic structure of the molecule [12]. Therefore, depending on the physico-chemical considerations used in creating a particular force field model, the charge values associated with a specific atom in the system of interest can be computed by some, arbitrarily chosen, model (as, e.g., in AMBER force field [13]), or they can be considered as empirical calibration parameters (as, e.g., in the GROMOS force field [14]).

However, regardless of the method used to derive the point charge values, the use of such charges results in isotropic contributions of each of the atoms to the overal electrostatic potential of the molecular system. At the same time, it is known that, in the case of, e.g., halogen bonding [15, 16] or σ-hole bonding [17] such isotropy doesn’t hold. Such inconsistency inherent to the point atom-centered change model may further lead to incorrect contributions to intermolecular interaction energies, because their electrostatic component is fully determined by the spatial distribution of the electrostatic potential of a molecule (MEP). Therefore, more attention has recently been given to considering the anisotropic components when constructing models to approximate the molecular electrostatic potential [13].

One of the approaches in building such models is to use a multipole expansion. In this approach, each atom is associated not only with a monopole term (point charge), but also with the dipole and quadrupole terms. The calibration of such models is usually done by selecting some sample of molecules and by minimizing the difference between the total multipole moment of a particular order for each of the molecules and the sum of the corresponding moments predicted by the atomic models being parametrized. However, in this approach, there is a risk of seemingly lowering the overall approximation error due to the mutual compensation of errors, for instance, in cases where the combined contribution to a certain multipole moment from a pair of atoms can remain accurate despite both contributions in the pair having non-negligible, but opposite errors, which cancel out each other due to their similar magnitudes and opposite signs, when added together. Consequently, the obtained multipole moments for individual atoms may not be transferable from one molecule to another one, and the constructed model may further incorrectly account for the influence of the atomic environment on the multipole moments of a given atom.

In this contribution, we propose a model which addresses this drawback by firstly performing a decomposition of the electronic charge density of the molecule into components that correspond to individual atoms or specific covalent bonds and then using each of these obtained components separately to approximate their individual contributions to the overall electrostatic potential of the molecule with a dedicated set of several point charges. By using the carefully tuned sets of point charges (which are not constrained to be atom-centered in our model) for each of the electron density components, the compatibility of the proposed approximation with the existing codes of force fields is ensured. The use of atomic contributions parameterized in the proposed way can help overcome the limitations of traditional force fields that rely solely on atomic point charges and enhance their accuracy when dealing with systems involving anisotropic distributions of electron density on atoms.

Although the underlying decomposition of the spatial electronic charge density of the molecule into the localized contributions can be performed within different physical models, by using a model that preserves the properties of the resulting components within molecules of different atomic structures as much as possible, it can be expected that the ultimate multipole moments will remain relatively universal, transferable, and applicable to various molecules. In the proposed model, the decomposition of the spatial electronic charge density of the molecule into the components, corresponding to the electron pairs of individual atoms or covalent bonds, is performed using the CLPO method [18, 19]. This method allows for the decomposition of the overall charge density of the molecule, obtained through non-empirical quantum chemical methods, into additive components localized on individual atoms or covalent bonds within the studied system. For each component obtained from this decomposition, the presented model enables the construction of a system of point charges approximating the electrostatic potential of the considered component, taking its possible anisotropy into account.

2. Theory

2.1. Decomposition of molecular electron charge density into the localized components

The spatial distribution of the electrostatic potential $\Phi(\mathbf{r})$ of a molecule is determined by the spatial distribution of its total charge density $\rho(\mathbf{r})$ according to the expression

$$\Phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int\frac{\rho(\mathbf{r}')\,d\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}, \tag{1}$$

which is the solution of the Poisson equation

$$\Delta\Phi(\mathbf{r}) = -\frac{\rho(\mathbf{r})}{\epsilon_0}.$$

In turn, the spatial distribution of the charge density of the molecule can be obtained with quantum-mechanical calculations which can be performed without involving empirical parameters other than a few fundamental physical constants. However, the densities obtained in this way typically take the form of an expansion containing products of pairs of basis functions. Although each function is localized on certain atoms of the system, two functions contributing to a given product can 'belong' to different atoms, making the overall term non-localized. This peculiarity makes it impossible to directly decompose the charge density $\rho(\mathbf{r})$ (and, thus, its electrostatic potential $\Phi(\mathbf{r})$) into localized components that would correspond to individual atoms of the molecule or its covalent bonds. Instead, such decomposition becomes possible, if localized orbitals are used as the basis functions.

Indeed, the spatial distribution of the electron charge density $\rho(\mathbf{r})$ is the diagonal part of the reduced one-particle density matrix (1-RDM) $\rho(\mathbf{r}) = \gamma(\mathbf{r},\ \mathbf{r})$ [20–22]:

$$\gamma(\mathbf{r},\mathbf{r}') = N_e \sum_{\sigma_1,\dots,\sigma_N}\int \Psi^{*\sigma_1\dots\sigma_N}(\mathbf{r},\mathbf{r}_2,\dots,\mathbf{r}_N)\times$$
$$\times\,\Psi^{\sigma_1\dots\sigma_N}(\mathbf{r}',\mathbf{r}_2,\dots,\mathbf{r}_N)\,d\mathbf{r}_2\dots d\mathbf{r}_N$$

where Ψ represents the many-particle electron wave function, N_e is the total number of electrons, and σ_i denote the electron spin indices in the wave function. In turn, for the 1-RDM, an approximate representation $\gamma(\mathbf{r},\mathbf{r}') \approx \gamma^{\mathrm{loc}}(\mathbf{r},\mathbf{r}')$ can be constructed in the form of expansion over localized orbitals as

$$\gamma^{\mathrm{loc}}(\mathbf{r},\mathbf{r}') = \sum_i n_i\,\varphi_i^{\mathrm{loc}}(\mathbf{r})\,\varphi_i^{\mathrm{loc}}(\mathbf{r}'), \tag{2}$$

where $\varphi_i^{\mathrm{loc}}(\mathbf{r})$ denotes the localized orbitals (which can be selected to be purely real for the sake of the present discussion), each of which can be associated with a specific atom or covalent bond of the molecule, and n_i represents the electron population of orbital φ_i^{loc}. Among various possible ways of constructing the localized orbitals as a basis for expansion (2), the CLPO method [18] provides the orbitals suitable for the optimal (in terms of minimizing the deviation $\left\|\gamma-\gamma^{\mathrm{loc}}\right\|_F^2$) approximation of the reduced one-particle density matrix. For that reason, these orbitals were chosen for the further use.

Thus, using (2), we obtain the decomposition of the spatial distribution of the electron charge density $\rho_{\mathrm{el}}(\mathbf{r})$ of the molecule in the form of an expansion:

$$\rho_{\mathrm{el}}(\mathbf{r}) \approx \sum_i n_i\left|\varphi_i^{\mathrm{loc}}(\mathbf{r})\right|^2 = \sum_i \rho_i(\mathbf{r}), \tag{3}$$

where $n_i\left|\varphi_i^{\mathrm{loc}}(\mathbf{r})\right|^2$ represents a component of the spatial electron charge density corresponding to the orbital φ_i^{loc}, which, in turn, is associated with a specific atom or covalent bond of the molecule. Due to their localization, it is convenient to rearrange the terms in expression (3) and to present the constructed approximation for $\rho_{\mathrm{el}}(\mathbf{r})$ in the form:

$$\rho_{\mathrm{el}}(\mathbf{r}) \approx \sum_a \rho_a^{at}(\mathbf{r}) + \sum_j \rho_j^{\mathrm{bond}}(\mathbf{r}), \tag{4}$$

where terms of the form $\rho_X(\mathbf{r}) = \sum_{i\in X} n_i\left|\varphi_i^{\mathrm{loc}}(\mathbf{r})\right|^2$ (with respective upper indices) represent the sum of components of the electron charge density corresponding to all orbitals localized on the a-th atom ($\rho_a^{at}(\mathbf{r})$) or the j-th covalent bond of the system ($\rho_j^{\mathrm{bond}}(\mathbf{r})$).

Due to the linear dependence of the potential $\Phi(\mathbf{r})$ on the total charge density $\rho(\mathbf{r})$, by substituting (4) into (1), and adding the contributions from the potential created by the nuclei of the molecule, it becomes possible to decompose the overall spatial distribution of the electrostatic potential of the molecule into additive localized components. In the next subsection, a model will be presented that allows each of these components to be modeled by a system of several point charges, the magnitude and positions of which are determined from the results of quantum chemical calculations.

2.2. Approximating localized charge density components with point charges

The proposed approximation for the potentials associated with localized components of the molecular charge density is based on the multipole expansion of the electrostatic potential which is valid at distances from the charge system that are significantly larger than the size of the system. This expansion, when truncated at quadrupole terms, is known to take the form [23]:

$$4\pi\epsilon_0\Phi(\mathbf{r}) = \frac{Q}{|\mathbf{r}-\mathbf{r}_0|} + \frac{(\mathbf{p}\,\mathbf{e}_r)}{|\mathbf{r}-\mathbf{r}_0|^2} + \frac{1}{2}\frac{(\mathbf{e}_r\,\hat{D}\mathbf{e}_r)}{|\mathbf{r}-\mathbf{r}_0|^3} + ..., \quad (5)$$

where $\Phi(\mathbf{r})$ represents the electrostatic potential, $\mathbf{r}_0$ is the radius vector of a certain point in a vicinity of the system, $\mathbf{e}_r = \frac{\mathbf{r}-\mathbf{r}_0}{|\mathbf{r}-\mathbf{r}_0|}$ is the unit vector directed toward the observation point $\mathbf{r}$, Q is the total charge of the system, $\mathbf{p}$ is the vector of its dipole moment, and $\hat{D}$ is the traceless quadrupole moment tensor. The latter can be represented as:

$$\hat{D} = 3\left(\hat{B} - \hat{1}\,\frac{1}{3}\,\mathrm{tr}\hat{B}\right), \quad (6)$$

where $\hat{1}$ is the identity matrix, and $\hat{B}$ is the tensor of second-order moments of the charge distribution. In the case of a system of discrete charges Q_j located at points with radius vectors $\mathbf{r}_j$, this tensor has the components

$$B_{ik} = \sum_j Q_j\,(\mathbf{r}_j - \mathbf{r}_0)_i\,(\mathbf{r}_j - \mathbf{r}_0)_k, \quad (7)$$

while, in the case of a system with a continuous distribution of charges, specified by the spatial electric charge density $\rho(\mathbf{r})$, the components of the tensor of moments are given by

$$B_{ik} = \int_V \rho(\mathbf{r})\,(\mathbf{r}-\mathbf{r}_0)_i(\mathbf{r}-\mathbf{r}_0)_k\,d\mathbf{r}. \quad (8)$$

The expressions for computing the system's dipole vector $\mathbf{p}$ in the two mentioned cases are

$$\mathbf{p} = \sum_j Q_j\,(\mathbf{r}_j - \mathbf{r}_0)$$

and

$$\mathbf{p} = \int \rho(\mathbf{r})\,(\mathbf{r}-\mathbf{r}_0)\,dV \quad (9)$$

respectively.

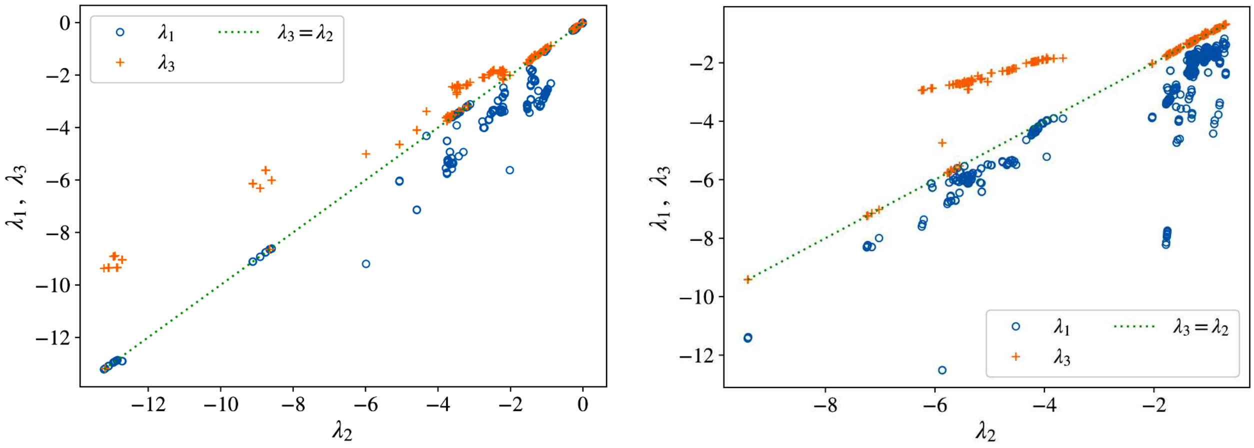
By applying expressions (6), (8), and (9) to the localized approximation (4) of the charge density of the molecule, we now obtain multipole moments corresponding to the electron pairs of each atom or covalent bond in the molecule. However, the direct use of expression (5) with such multipole moments for practical computations of the electrostatic potential of the molecule (for example, in the packages performing molecular dynamics simulations) would require the implementation of separate functions for calculating the dipole, quadrupole, and, if necessary, higher multipole moments. In contrast, each term of series (5) can further be modeled using a system of several point charges, provided that such a system of charges has the correct value of the corresponding multipole moment, and that it has zero values for all lower-order multipole moments, which are modeled by their own system of charges. The charges selected in this manner will mimic the electron pairs in terms of their contribution to the total MEP, with a particular emphasis on treating the anisotropy of such contributions properly. The ultimate approximation obtained in that way becomes readily usable in any molecular dynamics simulation package, because the functions of potentials for the interactions of point charges are usually already implemented in such software.

To minimize the computation time when using such an approximation, it is desirable to use the smallest possible total number of point charges. To minimize the number of such charges, one can choose a point $\mathbf{r}_0$ as the so-called charge center of the system, defined as

$$\mathbf{r}_c = \frac{1}{\int \rho(\mathbf{r})\,dV}\int \rho(\mathbf{r})\,\mathbf{r}\,dV. \quad (10)$$

By selecting such a reference point $\mathbf{r}_0$ coinciding with the charge center of the system, the dipole moment (9) and the corresponding term in expansion (5) become zero. Such reference point is used hereinafter.

In this case with only a single term, series (5) models the component of molecule's electrostatic potential due to the monopole and dipole terms. The modeling would then be performed with only one point charge located at the charge center of the charge component $\rho_a^{at}(\mathbf{r})$ or $\rho_j^{\mathrm{bond}}(\mathbf{r})$ found from (10). However, this approximation showed an error that was too large for practical applications, which necessitated the inclu-



a *b*

Fig. 1. The relationship between the eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ (in $e \cdot a_B^2$ units) of the tensor of charge density moments $\hat{B}$ for the components of the overall charge density of the molecule associated with its atoms (*a*) and covalent bonds (*b*)

sion of the next (quadrupole) term and a more accurate consideration of a possible anisotropy in the localized distributions of the charge components of the system.

To reduce the total number of point charges required in this case, the following reasoning was used. If the principal axes of the tensor $\hat{B}$ obtained from (8) for the continuous charge distribution with density $\rho(\mathbf{r})$ are chosen as the coordinate system's axes, then, in this chosen coordinate system, it is always possible to select a set of 7 charges, of which 6 are located at points $(a_x, 0, 0)$, $(-a_x, 0, 0)$, $(0, a_y, 0)$, $(0, -a_y, 0)$, $(0, 0, a_z)$, $(0, 0, -a_z)$, and the last one (required to ensure the zero total charge of the system) is at the origin (at the point $\mathbf{r}_0$), in such a way that their tensor of moments coincides with $\hat{B}$ exactly. Adding a term of the form $c \cdot \hat{1}$, with c being an arbitrary constant, to the tensor $\hat{B}$ will not change the tensor of quadrupole moments (6), and thus will not affect the corresponding term in the asymptotic expansion (5) of the electrostatic potential. Moreover, the orientations of the principal axes of the tensors $\hat{B}$ and $\hat{B} - c \cdot \hat{1}$ are the same.

In the important particular case, the term $c \cdot \hat{1}$ can be introduced by an additional system of 6 equal charges, each having coordinates $\pm b$ along one of the three axes of the chosen coordinate system. In this case, with an appropriate choice of the parameter b and the magnitude of the additional charges, we can compensate two of the six aforementioned charges at points $(a_x, 0, 0)$, $(-a_x, 0, 0)$, $(0, a_y, 0)$, $(0, -a_y, 0)$, $(0, 0, a_z)$, $(0, 0, -a_z)$. In other words, the traceless tensor $\hat{D}$ of quadrupole moments of the true distribution $\rho(\mathbf{r})$ can be reproduced by a system of only four charges, instead of six, located along two of the three principal axes of the tensor $\hat{B}$.

Let us now consider some peculiarities associated with the orientation of the principal axes and the eigenvalues of the tensor $\hat{B}$ for the components of the overall electron charge density of the molecule, which correspond to the electron pairs of its atoms or covalent bonds. In particular, it follows from the analysis of the data obtained for a sample of molecules from the GMTKN55 database (*vide infra* for further details) and presented in Fig. 1 that, for these molecules, two out of the three eigenvalues (expressed in $e \cdot a_{\mathrm{B}}^2$ units, where $a_{\mathrm{B}} \approx 0.529$ Å denotes the Bohr radius) of the tensor $\hat{B}$ are close to each other, meaning that the underlying charge density distributions are predominantly “uniaxial”.

This means that if one of the eigenvalues which are approximately equal one to another is chosen as the coefficient c, and $c \cdot \hat{1}$ is then subtracted from the tensor $\hat{B}$, then it becomes possible, in most cases, to model the contribution of the charge density distribution of an atom or bond to the quadrupole term in expansion (5) using a system of only three point charges lying on a straight line.

Such peculiarity can be observed on the distributions shown in Fig. 2. In particular, for the charge density distributions corresponding to covalent bonds, the principal axis of the tensor $\hat{B}$ corre-

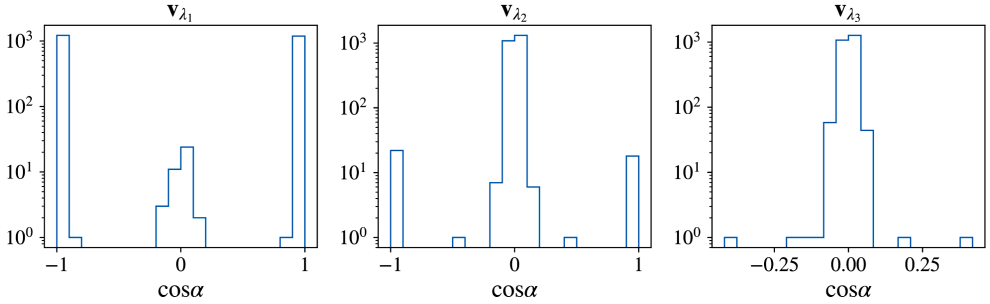


Fig. 2. Distribution of the cosine of the angle between the direction of the covalent bond and the eigenvectors $\mathbf{v}_{\lambda_1}$, $\mathbf{v}_{\lambda_2}$, $\mathbf{v}_{\lambda_3}$ of the tensor of charge density moments $\hat{B}$. For each molecule, the eigenvectors were ordered in the ascending order of their corresponding eigenvalues ($\lambda_1 \leq \lambda_2 \leq \lambda_3$)

sponding to the *smallest* (and negative) eigenvalue of this tensor is usually aligned in the direction of the bond. Indeed, in this case, the values close to ±1 dominate in the distributions of the cosine of the angle between the corresponding vectors. Thus, if the *largest* eigenvalue of the tensor $\hat{B}$ is taken as the coefficient c in the term $c \cdot \hat{1}$, then subtracting $c \cdot \hat{1}$ from $\hat{B}$ will leave only one of the diagonal elements of the charge moment tensor significantly different from zero, and this non-zero element will correspond to the eigenvector aligned in the direction of the bond. In such case, the contribution of the quadrupole terms can be modeled using point charges located on a line parallel to the bond direction.

Moreover, when the coefficient c in the term $c \cdot \hat{1}$ is selected to be equal to the *largest* eigenvalue of the tensor $\hat{B}$, the diagonal components (expressed in the principal axis of the tensor $\hat{B}$) of the tensor $\hat{B} - c \cdot \hat{1}$ will always have a negative sign. Consequently, when creating a system of point charges, whose tensor of moments (7) coincides with $\hat{B} - c \cdot \hat{1}$, only negative charges are sufficient to be used. If the magnitude of these charges is further chosen so that their sum equals the total electronic population of the original charge distribution $\rho(\mathbf{r})$, then, even in the cases where all three eigenvalues of the tensor $\hat{B}$ are noticeably different, the desired system can be formed with 4 (instead of 5) point charges. This is so, because there is no need for an additional charge at the origin to ensure the overall charge neutrality of the constructed system.

With the above considerations, it is now possible to summarize the steps needed to find the magnitude and location for the point charges needed to model the electrostatic potential of the given molecule.

2.3. Selection of the magnitudes and locations for point charges

As indicated in the previous subsection, the proposed model consists of the following steps. First, the structurally additive components of the overall spatial charge density of the molecule are found according to (4) using the CLPO method. Then, each of these components is associated with one, two, or four negative point charges which model its contribution to the overall electrostatic potential while accounting for its anisotropy. Finally, the obtained set of charges, mimicking the properties of the electron distribution, is complemented with positive point charges corresponding to the nuclei of all atoms in the system. In this process, the magnitude and position of the point charges should be obtained for each of the terms in (4) as following.

For the chosen term in (4), the charge center (10) of the electron charge density given by this term is taken as the origin, and all components of the tensor $\hat{B}$ are determined according to (8), for arbitrary orientations of the coordinate axes. By solving the eigenvalue-eigenvector problem for the tensor $\hat{B}$, the principal axes of the tensor $\hat{B}$ are found, and the largest eigenvalue is subtracted from the diagonal elements of $\hat{B}$. Denote the resulting non-zero diagonal elements as $\tilde{\lambda}_1 = \lambda_1 - \lambda_3$ and $\tilde{\lambda}_2 = \lambda_2 - \lambda_3$ (where $\lambda_1 \leq \lambda_2 \leq \lambda_3$ are the eigenvalues of $\hat{B}$). Depending on whether $|\tilde{\lambda}_1|$ or $|\tilde{\lambda}_2|$ exceeds a certain threshold value $\lambda_{\text{thresh}} \geq 0$, a system of four (if $\tilde{\lambda}_1 \leq \tilde{\lambda}_2 < -\lambda_{\text{thresh}} \leq 0$), two (if $\tilde{\lambda}_1 < -\lambda_{\text{thresh}} \leq \tilde{\lambda}_2 \leq 0$), or one (if $-\lambda_{\text{thresh}} \leq \tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq 0$) point charges is constructed. Such system will model the contribution of the considered term (4) to the overall electrostatic potential of the molecule. In this process, λ_{thresh} can even be set to zero, but any non-zero value of this parameter allows reducing the overall number of used point charges without any excessive increase in the error of the obtained approximation model.

In the case where $\lambda_{\text{thresh}} \neq 0$, but the condition $-\lambda_{\text{thresh}} \leq \tilde{\lambda}_1 \leq \tilde{\lambda}_2 \leq 0$ is satisfied, the single point charge for the system being constructed is placed at the origin (which is at the charge center of the considered term of (4)). In other cases, two (when $\tilde{\lambda}_1 < -\lambda_{\text{thresh}} \leq \tilde{\lambda}_2 \leq 0$) or four (when $\tilde{\lambda}_1 \leq \tilde{\lambda}_2 < -\lambda_{\text{thresh}}$) point charges are placed on the coordinate axes which were chosen according for the eigenvectors corresponding to the eigenvalues λ_1 and λ_2. In all cases, the magnitudes of all introduced charges are chosen to be equal to the same value, and this value is chosen to ensure that the sum of all charges equals the total electronic charge of the considered term from (4).

In choosing the position of the charges along the coordinate axes, the following considerations are used. In the cases where $\tilde{\lambda}_1 < -\lambda_{\text{thresh}}$ and $\tilde{\lambda}_2 < -\lambda_{\text{thresh}}$, the distances a_i from the origin to the corresponding point charges q_i are determined from the condition

$$2q_i \cdot a_i^2 = \tilde{\lambda}_i, \tag{11}$$

where i is 1 or 2. By satisfying condition (11), the constructed system of charges will have the tensor of moments that coincides with $\hat{B} - c \cdot \hat{1} = \hat{B} - \lambda_3 \cdot \hat{1}$. Therefore, such system will have the same tensor of quadrupole moments (6) as the original electron charge density term in expression (4). Additionally, setting all placed charges q_i to be equal to each other will ensure proportionality $|a_i| \sqrt{-\tilde{\lambda}_i}$, allowing the obtained positions of the charges q_i to be further used for the visual comparison of the quadrupole moments of the considered charge density component depending on their orientation.

It is worth stressing that all the input data needed for the proposed model to produce the ultimate distribution of point charges can be obtained from quantum-chemical calculations, so that the ultimate system of charges can thus be referred to as non-empirical and requires no additional calibration of any adjustable parameters (except for λ_{thresh} which only influences computational efficiency).

Another essential feature of the proposed model is that it essentially operates on the partition of the molecular electron charge density performed in a Hilbert space (which is typical of the population-based models of atomic charges [24]), but provides charges suitable for approximating the components of the intermolecular interaction energy (which is typical of the schemes based on fitting the charges to the electrostatic potential in the real space [25,26]). It has been shown previously that the population-based descriptors of the electron charge distrubution are less dependent on the conformation of a molecule [27, 28] (and, thus, should be more transferrable) than those derived from the distribution of the overal molecular electrostatic potential in the real space.

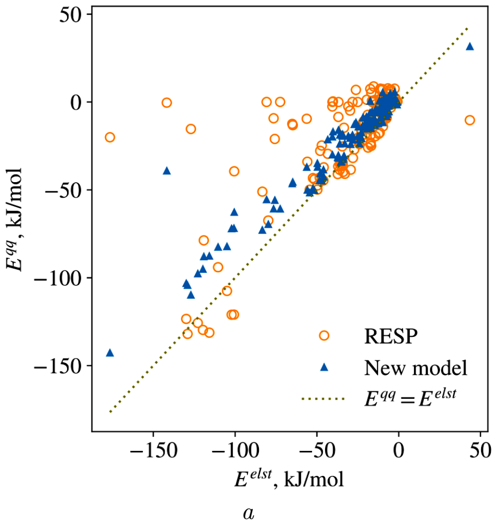
In the following section, the proposed model is applied to the approximate calculation of the electrostatic component of the intermolecular interaction energy. The accuracy of such approximation is compared to the accuracy of a similar approximation based on the RESP point charge model (see [26] and references therein), which is used [13] in the widely applied force fields designed for molecular dynamics simulations.

3. Computation Details

To evaluate the errors related to the approximation of the molecular electrostatic potential with the sum of additive localized components constructed from point charges and parametrized using the proposed physical model, a sample of 145 molecular systems was created with the use of the GMTKN55 database [29]. This database was chosen, because it contains non-covalently bonded complexes of small molecules, which were selected by its authors to provide a balanced representation of the main practically important types of non-covalent interactions.

3.1. Quantum-chemical methods

For each molecule from this sample, the second-order Møller–Plesset perturbation theory (MP2) with Def2-TZVPP basis set was used to obtain *ab initio* one-particle density matrices, with the orbital relaxation taken into account. The quantum-chemical calculations were performed using the PSI4 package (version 1.2a1.dev781) [30]. The modules of the same software package were also used to compute the matrix elements of the dipole and quadrupole potentials between the atomic basis functions for all the studied molecules. The results are available via



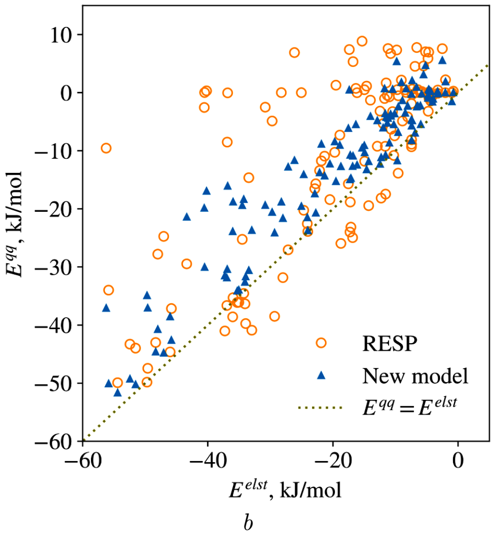


Fig. 3. Comparison of the electrostatic energy component of the intermolecular interaction E^{elst} with its value E^{qq} approximated within the RESP point charge model and the new proposed model in 145 complexes from the GMTKN55 database [29]. The results for all complexes are shown in panel (*a*), while panel (*b*) shows the subset of these complexes with electrostatic energy in the interval from –60 kJ/mol to 0 kJ/mol

DOI: 10.6084/m9.figshare.23804184. These computations were carried out using a specially developed driver program. The localized orbitals were found according to the CLPO approach, by using the JANPA program (version 2.03) [18].

To implement the calculations in accordance with the proposed model for finding the magnitude and positions of point charges, a separate Python3 program was created. The threshold value of $\lambda_{\text{thresh}} \geq 0$ was set to $0.1ea_{\text{B}}^2$, where e is the electron charge, and $a_{\text{B}} \approx 0.529$ Å is the Bohr radius.

The RESP point charges needed for a comparative analysis were obtained using the Multiwfn package (version 3.7(dev)/2019-Jul-18) [31].

3.2. Accuracy metrics

To evaluate the accuracy of the proposed model, the electrostatic component of the intermolecular interaction energy was used as the key metric. This choice is related to the important role of this quantity in the force fields used in molecular dynamics simulations, as well as is motivated by the fact that the net value of the electrostatic component results from a delicate balance between repulsive and attractive terms. Typically, such net value is much smaller in absolute value than the sum of attractive or repulsive terms alone. It is thus expected that the ability to correctly approximate the electrostatic component of the intermolecular interaction energy should be a more challenging test as compared to, e.g., assessing the errors in the absolute values of the electrostatic potential of a single molecule on some grid of points in the space.

Accordingly, the electrostatic component E^{qq} of the intermolecular interaction energy was obtained as

$$E^{qq} = \sum_{i\in A}\sum_{j\in B} \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|} \qquad (12)$$

from the point charges derived with the proposed model. In (12), the summation was performed over all charges q_i and q_j, which were obtained separately for each of the molecules (A and B) in the complex by applying the proposed model. To characterize the approximation errors of the proposed model, the energy E^{qq} predicted by this model was compared with the true values of the corresponding component calculated as

$$E^{elst} = \int \frac{\rho_A(\mathbf{r})\,\rho_B(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}\, d\mathbf{r}' \qquad (13)$$

based on the spatial charge distributions $\rho_A(\mathbf{r})$ and $\rho_B(\mathbf{r}')$ (which include the contributions from atomic nuclei) obtained for isolated molecules A and B from *ab initio* quantum-chemical computations.

Expression (12) was also used to compute an approximation E^{qq} to E^{elst} based on the RESP point-charge method, in which the point atomic charges were placed at the locations of the atomic nuclei. In these computations, the same charge density distribution was used to derive both the RESP charges and the charges suggested by the proposed model.

4. Results and Discussion

The values of the electrostatic component of the intermolecular interaction energy were obtained for the studied complexes within the proposed model, as well as with the RESP method. The obtained values are compared in Fig. 3 with the true values of this component obtained using formula (13) with the true distributions of the charge densities found from *ab initio* quantum-chemical computations.

To quantify the agreement between the true and approximated values of the electrostatic contribution to the intermolecular interaction energy, the root-mean-square deviation (RMSD) was calculated as

$$\delta E^{\mathrm{RMSD}} = \sqrt{\frac{1}{N_{\mathrm{compl}}} \sum_{k=1}^{N_{\mathrm{compl}}} \left(E_k^{qq} - E_k^{elst}\right)^2},$$

where the summation index k enumerates the investigated complexes, and N_{compl} corresponds to their total number. The obtained values of RMSD values are $\delta E^{\mathrm{RMSD}} = 25.1$ kJ/mol when using the point charges obtained according to the RESP model, and $\delta E^{\mathrm{RMSD}} = 11.3$ kJ/mol (2.7 kcal/mol) when using the point charges obtained according to the proposed model.

Such results and the generally better agreement between the values E^{qq} and E^{elst} observed in Fig. 3 in the case of the proposed model indicate a more pronounced correctness in approximating the contributions of covalent bonds and atomic lone electron pairs to the overall electrostatic potential of the molecule.

As an example of the spatial arrangement of the point charges obtained from the proposed physical model relatively to the atoms of the investigated system, Fig. 4 presents the case of two water molecules in the geometry of their hydrogen-bonded dimer.

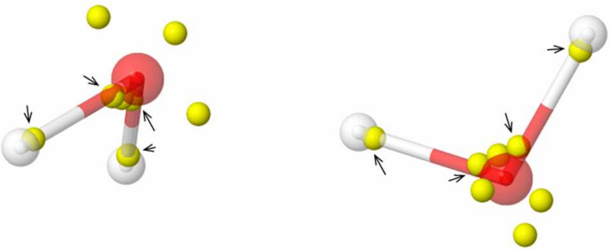


Fig. 4. Arrangement of point charges (depicted as small, non-transparent spheres) for modeling the electrostatic potential of molecules in a hydrogen-bonded dimer of water, as determined from the proposed model. Arrows indicate the charges whose properties are determined by the characteristics of O–H covalent bonds

In particular, the shown arrangement of point charges reflects the anisotropy in the distribution of the electronic charge density on the O atom, which is related to the existence of its two lone electron pairs. These pairs are known to be essential for the hydrogen bonding.

5. Conclusions

A physical model is proposed for approximating the electrostatic potential of a molecule with a sum of localized components corresponding to the individual atoms or covalent bonds, each component being constructed from one, two, or four point charges. Within the framework of the proposed model, the magnitudes and spatial positions of such charges can be determined from the data found from non-empirical quantum chemical methods, without involving additional adjustable parameters. Moreover, the model allows one to gradually vary the number of additional point charges depending on the required accuracy of the ultimate approximation of the overall electrostatic potential of the molecule.

Using the set of 145 molecular complexes from the GMTKN55 database, it has been shown that the system of point charges obtained according to the proposed model allows modeling the electrostatic component of the intermolecular interaction energy with a root-mean-square deviation of 11.3 kJ/mol, which is 2.2 times lower than the corresponding deviation obtained when using point charges found from the RESP model. The obtained results along with the possibility of representing the proposed approximation of the electrostatic potential with the functions already available in the existing molecular dynamics software packages pose the proposed model as the promising one for creating the more reliable sets of interatomic potentials for investigating the molecular systems using computer modeling methods.

1. J.A. Lemkul. Chapter one – pairwise-additive and polarizable atomistic force fields for molecular dynamics simulations of proteins. In: *Progress in Molecular Biology and Translational Science*, *Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly*. Edited by B. Strodel, B. Barz (Academic Press, 2020).
2. L. Monticelli, D.P. Tieleman. Force fields for classical molecular dynamics. In: *Biomolecular Simulations: Methods and Protocols*, *Methods in Molecular Biology*. Edited by L. Monticelli, E. Salonen (Humana Press, 2013).
3. T.Y. Nikolaienko. Interaction of anticancer drug doxorubicin with sodium oleate bilayer: Insights from molecular dynamics simulations. *J. Mol. Liq.* **235**, 31 (2017).
4. G.A. Cisneros, M. Karttunen, P. Ren, C. Sagui. Classical electrostatics for biomolecular simulations. *Chem. Rev.* **114** (1), 779 (2014).
5. G.M.A. Junqueira, M.Y. Ballester, M.A.C. Nascimento. Reactivity properties of the hoso and hso$_2$ isomers in liquid medium: a sequential monte carlo/quantum mechanics study. *J. Mol. Model.* **29** (6), 189 (2023).
6. M. Liu, S. Wang. Mcdock: A monte carlo simulation approach to the molecular docking problem. *J. Comput. Aid. Mol. Des.* **13** (5), 435 (1999).
7. K. Coutinho, H.C. Georg, T.L. Fonseca, V. Ludwig, S. Canuto. An efficient statistically converged average configuration for solvent effects. *Chem. Phys. Lett.* **437** (1), 148 (2007).
8. A.A. Adeniyi, M.E. S. Soliman. Implementing qm in docking calculations: Is it a waste of computational time? *Drug. Discov. Today* **22** (8), 1216 (2017).
9. A.I. Samtsevich, L.A. Bulavin, L.F. Sukhodub, T.Y. Nikolaienko. Interaction of dna nucleotide bases with anticancer drug thiotepa: Molecular docking and quantum-mechanical analysis. *Ukr. Biochem. J.* **86** (2), 50 (2014).
10. G.M. Morris, M. Lim-Wilby. Molecular docking. In: *Molecular Modeling of Proteins*, *Methods Molecular Biology*. Edited by A. Kukol (Humana Press, 2008).
11. M. Huix-Rotllant, N. Ferre. Analytic energy, gradient, and hessian of electrostatic embedding QM/MM based on electrostatic potential-fitted atomic charges scaling linearly with the MM subsystem size. *J. Chem. Theory Comput.* **17** (1), 538 (2021).
12. J.F. Gonthier, S.N. Steinmann, M.D. Wodrich, C. Corminboeuf. Quantification of "fuzzy" chemical concepts: A computational perspective. *Chem. Soc. Rev.* **41** (13), 4671 (2012).
13. S. Riniker. Fixed-charge atomistic force fields for molecular dynamics simulations in the condensed phase: An overview. *J. Chem. Inf. Model.* **58** (3), 565 (2018).
14. C. Oostenbrink, A. Villa, A.E. Mark, W.F. Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25** (13), 1656 (2004).
15. G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati, G. Terraneo. The halogen bond. *Chem. Rev.* **116** (4), 2478 (2016).
16. M.R. Scholfield, C.M.V. Zanden, M. Carter, P.S. Ho. Halogen bonding (X-bonding): A biological perspective. *Protein Sci.* **22** (2), 139 (2013).
17. P. Politzer, J.S. Murray, M.C. Concha. σ-hole bonding between like atoms; a fallacy of atomic charges. *J. Mol. Model.* **14** (8), 659 (2008).
18. T.Y. Nikolaienko, L.A. Bulavin. Localized orbitals for optimal decomposition of molecular properties. *Int. J. Quantum Chem.* **119** (3), e25798 (2019).
19. T.Y. Nikolaienko. The maximum occupancy condition for the localized property-optimized orbitals. *Phys. Chem. Chem. Phys.* **21** (9), 5285 (2019).
20. E.R. Davidson. *Reduced Density Matrices in Quantum Chemistry* (Academic Press, 1976) [ISBN: 9780122058509].
21. A.J. Thakkar, A.C. Tanner, V.H. Smith. Inter-relationships between various representations of one-matrices and related densities: A road map and an example. In: *Density Matrices and Density Functionals*. Edited by R. Erdahl, V.H. Smith (Springer Netherlands, 1987).
22. I. Mayer. *Bond Orders and Energy Components: Extracting Chemical Information from Molecular Wave Functions* (CRC Press, Taylor & Francis, 2017) [ISBN: 9781420090116].
23. L.D. Landau, E.M. Lifshitz. *The Classical Theory of Fields* (Pergamon Press, 1975) [ISBN: 9780080181769].
24. E.R. Davidson, A.E. Clark. A viewpoint on population analyses. *Int. J. Quantum Chem.* **122** (8) (2022).
25. J.S. Murray, P. Politzer. The electrostatic potential: An overview. *WIRES Comput. Mol. Sci.* **1** (2), 153 (2011).
26. A. Alenaizan, L.A. Burns, C.D. Sherrill. Python implementation of the restrained electrostatic potential charge model. *Int. J. Quantum Chem.* **120** (2), e26035 (2020).
27. T. Nikolaienko, L. Bulavin, D. Hovorun. Effective atomic charges of canonical 2′-deoxyribonucleotides and their conformational dependences. *Uk.r J. Phys.* **57** (10), 1024.
28. T.Y. Nikolaienko, L.A. Bulavin, D.M. Hovorun. Can we treat ab initio atomic charges and bond orders as conformation-independent electronic structure descriptors? *RSC Adv.* **6** (78), 74785 (2016).
29. L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, S. Grimme. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **19** (48), 32184 (2017).
30. R.M. Parrish, L.A. Burns, D.G.A. Smith, A.C. Simmonett, A.E. DePrince, E.G. Hohenstein, U. Bozkaya, A.Y. Sokolov, R. Di Remigio, R.M. Richard, J.F. Gonthier, A.M. James, H.R. McAlexander, A. Kumar, M. Saitow, *et al.* Psi4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and in-

teroperability. *J. Chem. Theory Comput.* **13** (7), 3185 (2017).
31. T. Lu, F. Chen. Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* **33** (5), 580 (2012).

Received 29.07.23

Т.Ю. Ніколаєнко, Л.А. Булавін

АПРОКСИМАЦІЯ ЕЛЕКТРОСТАТИЧНОГО ПОТЕНЦІАЛУ МОЛЕКУЛИ ТОЧКОВИМИ ЗАРЯДАМИ, ЩО ІМІТУЮТЬ ЇЇ ЕЛЕКТРОННІ ПАРИ

У традиційних потенціалах міжатомної взаємодії (“силових полях”) їхня електростатична складова істотно впливає на коректність моделювання ними водневих зв’язків та інших нековалентних взаємодій, притаманних біомолекулярним системам. У роботі запропоновано фізичну модель для апроксимації електростатичного потенціалу молекули (МЕП), засновану на розкладанні її густини заряду, отриманої з перших принципів, на локалізовані складові. На відміну від традиційних схем, в яких для апроксимації МЕП зазвичай використовують точкові заряди, розміщені на атомах, у запропонованому підході такі заряди розміщуються у точках, вибраних так, щоб зімітувати анізотропію розподілу електронної густини, зумовлену електронними парами атомів або ковалентних зв’язків. Ця особливість приводить до більш точного представлення загального електростатичного потенціалу молекули, що підтверджується застосуванням запропонованої моделі для апроксимації електростатичної складової міжмолекулярної взаємодії в 145 нековалентно-зв’язаних молекулярних комплексах із бази GMTKN55. У результаті такого застосування було знайдено, що середньоквадратичне відхилення апроксимованих значень електростатичної складової енергії міжмолекулярної взаємодії від справжніх для запропонованої моделі становило 2,7 ккал/моль, що у 2,2 рази менше порівняно з традиційним методом, заснованим на використанні RESP-зарядів на атомах.

Ключові слова: електростатична складова енергії взаємодії, густина електронного заряду, електростатичний потенціал молекули, заряди атомів.