

Article

Efficient Evaluation of Molecular Electrostatic Potential in Large Systems

Rafael Lopez * , Frank Martinez, Ignacio Ema and Jose Manuel Garcia de la Vega and Guillermo Ramirez

Departamento de Química Física Aplicada, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain; frank.martinez@estudiante.uam.es (F.M.); nacho.ema@uam.es (I.E.); garcia.delavega@uam.es (J.M.G.d.l.V.); guillermo.ramirez@uam.es (G.R.)

* Correspondence: rafael.lopez@uam.es

Received: 23 October 2019; Accepted: 9 November 2019; Published: 12 November 2019



Abstract: An algorithm for the efficient computation of molecular electrostatic potential is reported. It is based on the partition/expansion of density into (pseudo) atomic fragments with the method of Deformed Atoms in Molecules, which allows to compute the potential as a sum of atomic contributions. These contributions are expressed as a series of irregular spherical harmonics times effective multipole moments and inverse multipole moments, including short-range terms. The problem is split into two steps. The first one consists of the partition/expansion of density accompanied by the computation of multipole moments, and its cost depends on the size of the basis set used in the computation of electron density within the Linear Combination of Atomic Orbitals framework. The second one is the actual computation of the electrostatic potential from the quantities calculated in the first step, and its cost depends on the number of computation points. For a precision in the electrostatic potential of six decimal figures, the algorithm leads to a dramatic reduction of the computation time with respect to the calculation from electron density matrix and integrals involving basis set functions.

Keywords: molecular electrostatic potential; electron density; atoms in molecules; noncovalent interactions; chemical reactivity

1. Introduction

Despite of its undeniable success in the description of chemical structure and reactivity, the language of chemistry is largely grounded on concepts with loose theoretical foundations. Concepts like bond, atomic charges, orbitals, hybridization, resonance, delocalization, hyperconjugation and many others cannot be directly measured in physical experiments or expressed more precisely in the jargon of quantum mechanics, they are not observables.

The proposal of G.N Lewis [1], in the early twenties of last century, to explain all the previous work made on the study of the molecular structure was recognized from the very beginning as an outstanding contribution to the field. Because of the weak arguments that supported his model, his work was promptly followed by attempts to overcome its theoretical shortcomings by providing a more rigorous foundation based on the recently discovered quantum mechanics. In these attempts, many concepts were proposed to underpin Lewis ideas by pioneers like Coulson, Pauling, Hückel and Mulliken, just to mention some of the most conspicuous figures. Their early efforts to accommodate the complex and sometimes elusive chemical facts in the frame provided by the new emerging physics were driven in part by the available computational capabilities, that were very reduced at the time.

Thus, what at first was an admirable exercise of imagination in its origin has led to a burden of new concepts in an attempt to grasp the increasing knowledge on the complexity of the chemical

world. The effect of this multiplication of concepts has been keenly illustrated by P. Politzer in the case of the interpretation of a well established concept in chemistry as is hydrogen bonding. Numerous interpretations have been given depending on the circumstances and the authors' preferences. As Politzer humorously quotes, hydrogen bonding "has been dissected into classical and nonclassical, proper and improper (immoral?), blue-shifted and red-shifted, dihydrogen, anti-hydrogen (rebellious?), resonance-assisted, polarization-assisted, and more" [2].

Without wishing to understate the usefulness and achievements of the somehow imprecise concepts in which chemical language is currently built, it is worth recalling that there are two physically well grounded observables that are most useful in describing the realm of chemistry: the electron density and the electrostatic potential.

In the Born–Oppenheimer envisioning of molecular structure [3], molecular electron density (MED) can be regarded as a mathematical description of the electron cloud surrounding the nuclei in the molecule. As it is a function defined in 3D space, it can be displayed and easily connected with intuition, and faithful approximations can be obtained with current computational methods at a low or moderate cost, depending on the system's size. Although the importance of MED was realized very early, it remained long time as an auxiliary quantity for computing other properties like atom and bond charges [4] or to reinforce the analysis based on wave function and its related concepts in the Linear Combination of Atomic Orbitals (LCAO) framework, such as orbitals, hybridization and the like. The recognition of its usefulness as a powerful tool for analysis of molecular structure had to wait to R. Bader's work [5,6], who changed the focus considering MED as a very central property for chemical interpretation, as an alternative to the well established orbital analysis. Since then, the topological analysis of MED has unceasingly gained popularity among theoretical chemists, and currently, it has become a well established tool that is customarily used in the interpretation of the chemical structure and reactivity.

The case of molecular electrostatic potential (MESP) is somehow comparable to that of MED, with the added difficulty of the cost of its faithful computation in large systems. Former works on the application of MESP to chemical facts [7–9] highlighted the possibilities of MESP in the interpretation of reactivity. The analysis of MESP in molecular surfaces and the exploration of the relations between its values with thermochemical properties by Politzer and Murray [10–12] have proved to be useful tools for systematization and prediction of such properties. Besides, the introduction of the concepts of *sigma* and *pi* holes by these authors has been very valuable in the interpretation of molecular interactions between closed shell systems [13,14]. Furthermore, the outstanding contributions by S. Gadre et al [15–21] on the topological analysis of MESP have added new insight on the structure and properties of molecules and molecular aggregates. However, albeit this analysis of MESP topology is breaking through in current research, its application to large systems is being hindered by the cost of computing accurate values of MESP and its derivatives. Thus, the introduction of more efficient procedures like that proposed herein has greatly facilitated this type of studies, and several works using the new algorithms have appeared in recent literature [22–25].

In this work, we report details on the algorithms used for the efficient calculation of MESP in large systems, based on the Deformed Atoms in Molecules (DAM) partition/expansion [26] of MED. This procedure allows us to split the problem in two separate steps, one dealing with the MED partition/expansion itself, and another corresponding to the actual computation of MESP [27]. As it will be shown, the cost of the first step is independent of the number of MESP computation points, and the second is independent of the number of basis set functions in the calculation. This separation is crucial for the high performance of the method.

The article is organized as follows. In the next section, the theoretical foundations of the method are outlined, only essential equations are given and the remaining ones are collected in the supplementary file (SF) accompanying this work. The third section is devoted to the description of the algorithms' implementation, and is split in two subsections collecting the algorithms for the two steps of the procedure. Again, subordinate equations and technical details are addressed in SF.

Several results on performance and accuracy are presented in the fourth section and, finally, significant conclusions are drawn from these results.

2. Method

The algorithm for the efficient evaluation of MESP reported here is based on DAM [26] partition of electron density. In this partition the molecular density is expressed as a sum of atomic contributions:

$$\rho(\mathbf{r}) = \sum_A \rho^A(\mathbf{r}_A) \tag{1}$$

where the densities of the (pseudo)atomic fragments, $\rho^A(\mathbf{r}_A)$, are obtained with a least-deformation criterion based on the fast convergence of the long-range multipole expansion of the electrostatic potential [28]. In practice, in the Linear Combination of Atomic Orbitals framework (LCAO), in which DAM is formulated, this fast convergence is achieved by assigning the one-center distributions, i.e., products of pairs of basis functions $\chi_a^A(\mathbf{r}_A)$ $\chi_{a'}^A(\mathbf{r}_A)$, to their pertaining atoms, A , and partitioning the two-center distributions, $\chi_a^A(\mathbf{r}_A)$ $\chi_b^B(\mathbf{r}_B)$, between their respective centers:

$$\chi_a^A(\mathbf{r}_A) \chi_b^B(\mathbf{r}_B) = d_{ab}^A(\mathbf{r}_A) + d_{ab}^B(\mathbf{r}_B) \tag{2}$$

In this work, we will deal with basis functions, χ , consisting of Gaussian contracted functions (CGTO), which are most used in molecular calculations.

Contracted Gaussian functions, $\chi_{LM}^{(CGTO)}$, are linear combinations of primitive Gaussian functions, $g_{LM}(\xi, \mathbf{r})$:

$$\chi_{LM}^{(CGTO)}(\mathbf{r}) = \sum_{i=1}^{N_G} c_i g_{LM,i}(\mathbf{r}) \tag{3}$$

where N_G is the number of primitive functions (length of the contraction), index i runs over the primitive functions in the contraction, and the expansion coefficients, c_i , are chosen so that the radial part of the contraction remains normalized. Primitive Gaussians on their side are defined as:

$$g_{LM}(\xi, \mathbf{r}) = \mathcal{N}_{\xi}^{r(GTO)} r^L e^{-\xi r^2} \mathcal{N}_{LM}^{\Omega} z_L^M(\theta, \phi) \tag{4}$$

and the same angular part is taken for all the primitives in a given contraction. Thus, the contracted functions can be written as:

$$\chi_{LM}^{(CGTO)}(\mathbf{r}) = \left[\sum_{i=1}^{N_G} c_i \mathcal{N}_{\xi_i}^{r(GTO)} e^{-\xi_i r^2} \right] r^L \mathcal{N}_{LM}^{\Omega} z_L^M(\theta, \phi) \tag{5}$$

where $\mathcal{N}_{\xi_i}^r$ and $\mathcal{N}_{LM}^{\Omega}$ are radial and angular normalization constants defined by:

$$\mathcal{N}_{\xi}^{r(GTO)} = 2^{L+1} \sqrt{\frac{\xi^L}{(2L+1)!!}} \left[\frac{(2\xi)^3}{\pi} \right]^{1/4} \tag{6}$$

$$\mathcal{N}_{LM}^{\Omega} = \left[\frac{(2L+1)(L-|M|)!}{2(1+\delta_{M0})\pi(L+|M|)!} \right]^{1/2} \tag{7}$$

and $z_L^M(\theta, \phi)$ are unnormalized real spherical harmonics given by:

$$z_L^M(\theta, \phi) = (-1)^M P_L^{|M|}(\cos \theta) \begin{cases} \cos M\phi & (M \geq 0) \\ \sin |M|\phi & (M < 0) \end{cases} \tag{8}$$

where $P_L^{|M|}(z)$ are the corresponding associated Legendre functions (see Equation 8.751.1 at [29]).

For two-center densities, $g_{00}^A(\zeta_A, \mathbf{r}_A)$ $g_{00}^B(\zeta_B, \mathbf{r}_B)$, consisting of pairs of spherical primitive Gaussians ($L = 0$) centered at two different points, A and B , it has been proved [28] that the best convergence in the long-range potential is achieved by assigning the whole density to the center with higher exponent ζ or, in case of equal exponents, by assigning one half to each center. This result reminds the *nearest-site algorithm* reported by A. Stone and M. Alderton [30], but without considering expansions at the bond center.

When applying this criterion to densities made of pairs of spherical contracted functions, the products of primitives become assigned to one center or another depending on their exponents, yielding a partition that can be regarded as a more realistic version of the Mulliken partition (see Figure 1).

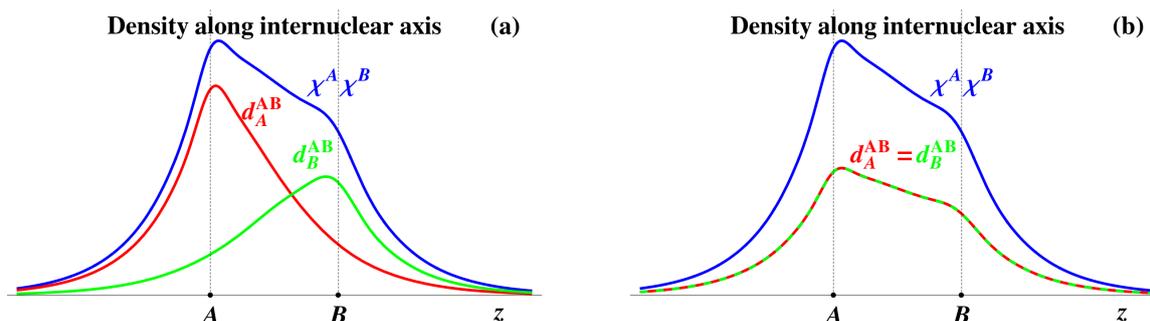


Figure 1. Partition of two-center electron densities, $\chi^A \chi^B$, into atomic fragments d_{ab}^A, d_{ab}^B : (a) DAM, (b) Mulliken.

For distributions consisting of nonspherical functions ($L \neq 0$), the partition is applied to the products of brackets of Equation (5), and the remaining terms are translated to the centers as described below.

The next step in the procedure is to expand the densities of the atomic fragments thus obtained as a series of radial factors times unnormalized real regular solid harmonics (hereafter *regular harmonics*, for short), namely:

$$\rho^A(\mathbf{r}_A) = \sum_{l=0}^{\infty} \sum_{m=-l}^l z_l^m(\mathbf{r}_A) \rho_{lm}^A(r_A) \tag{9}$$

where the regular harmonics, $z_l^m(\mathbf{r})$, are related to unnormalized real spherical harmonics by:

$$z_l^m(\mathbf{r}) = r^l z_l^m(\theta, \phi) \tag{10}$$

The partition of density given in Equation (1) combined with the expansion (9) allows us to write the electrostatic potential as[27]:

$$V(\mathbf{r}) = \sum_{A=1}^N \left\{ \frac{\zeta_A}{r_A} - \sum_{l=0}^{\infty} \sum_{m=-l}^l z_l^m(\mathbf{r}_A) \left[\frac{Q_{lm}^A(r_A)}{r_A^{2l+1}} + q_{lm}^A(r_A) \right] \right\} \tag{11}$$

in terms of atomic nuclei charges, ζ_A , effective multipoles, $Q_{lm}(r)$, and inverse multipoles, $q_{lm}(r)$:

$$Q_{lm}^A(r) = \frac{(l - |m|)!}{(l + |m|)!} \int_{r' < r} d\mathbf{r}' z_l^m(\mathbf{r}') \rho^A(\mathbf{r}') = \frac{4 \pi}{2l + 1} \int_0^r dr' r'^{2l+2} \rho_{lm}^A(r') \tag{12}$$

$$q_{lm}^A(r) = \frac{(l - |m|)!}{(l + |m|)!} \int_{r' > r} d\mathbf{r}' \frac{z_l^m(\mathbf{r}')}{r'^{2l+1}} \rho^A(\mathbf{r}') = \frac{4 \pi}{2l + 1} \int_r^{\infty} dr' r' \rho_{lm}^A(r') \tag{13}$$

In this way, the molecular electrostatic potential results in a sum of atomic contributions and the short/long-range separation can be carried out at the atomic level. Thus in the long-range region, the effective multipoles can be accurately replaced by point multipoles:

$$Q_{lm}^A = \lim_{r \rightarrow \infty} Q_{lm}^A(r) = \frac{4\pi}{2l+1} \int_0^\infty dr' r'^{2l+2} \rho_{lm}^A(r') \tag{14}$$

and the inverse multipoles do vanish: $\lim_{r \rightarrow \infty} q_{lm}^A(r) = 0$.

As it will be shown below, this is a most useful feature when dealing with large systems because a huge fraction of the atomic contributions to MESP can be computed from only the long range terms *even in the regions where molecular short-range potential is necessary*. This apparent paradox comes from the fact that, in these systems, the molecular short-range at a given point usually involves a reduced number of atomic short-range contributions coming from atoms in the neighborhood of the point, the remaining contributions being of long-range type.

3. The Algorithm

The algorithm for the evaluation of the electrostatic potential in large molecular systems, according to the method described in the previous section, consists of two main steps, which are executed in sequence. First, a partition of the molecular density must be carried out with the DAM procedure followed by the expansion of the atomic fragments and the computation of effective and inverse multipoles. Second, once the partition/expansion has been made, the electrostatic potential is computed in the desired points with the aid of Equation (11).

It is worth noticing that the first step depends on the size of the basis set used in the computation of molecular density, but it is independent of the number of points where the MESP has to be computed. On the other hand, the second step depends on the number of points for computation, but it is independent of the basis set size.

This decoupling of processes depending on the system size (number of basis functions) from those depending on the number of MESP computation points, combined with the short/long-range separation at atomic level, makes the procedure reported here most efficient to deal with large systems, provided that the MESP is to be evaluated in a not too reduced number of points.

3.1. Algorithm for Molecular Density Partition/Expansion

According to DAM partition criterion, the density of atomic fragment at *A* is given by:

$$\rho^A(\mathbf{r}) = \sum_{a,a' \in A} \chi_a^A(\mathbf{r}_A) \chi_{a'}^A(\mathbf{r}_A) + \sum_{B \neq A} \sum_{a \in a} \sum_{b \in B} d_{ab}^A(\mathbf{r}_A) \tag{15}$$

where, for two-center CGTO distributions:

$$d_{ab}^A = \sum_i^{NG_a} \sum_j^{NG_b} \Theta(\xi_i^a - \xi_j^b) c_i c_j g_i^a(\xi_i^a, \mathbf{r}_A) g_j^b(\xi_j^b, \mathbf{r}_B) \tag{16}$$

$\Theta(x)$ being the step function:

$$\Theta(x) = \begin{cases} 0 & x < 0 \\ 1/2 & x = 0 \\ 1 & x > 0 \end{cases} \tag{17}$$

As our purpose is to express the fragments d_{ab}^A in coordinates referred to center *A* and the second primitive in each term of Equation (16) is centered at *B*, it is necessary to translate the functions g_j^b to center *A*. The translation of the exponential factor in g_j^b can be made in terms of Bessel *I* functions as proposed by Kaufmann and Baumeister [31]. Working in an *aligned frame*, with *A* placed at the origin and *B* lying on the *z* axis, i.e., $\mathbf{R}_A = (0, 0, 0)$, $\mathbf{R}_B = (0, 0, R_B)$, the translation formula reads:

$$e^{-\xi_j^b |\mathbf{r} - \mathbf{R}_B|^2} = e^{-\xi_j^b (r^2 + R_B^2)} \sum_{l=0}^{\infty} (2l+1) \sqrt{\frac{\pi}{4 \xi_j^b r R_B}} I_{l+1/2}(2 \xi_j^b r R_B) P_l(\cos \theta) \tag{18}$$

where $R_B \equiv |\mathbf{R}_B|$, $I_{l+1/2}(z)$ are the corresponding Bessel functions (see [29] 8.467) and $P_l(z)$ are the Legendre polynomials (ibid 8.91).

On the other hand, the remaining factor, which is in essence the regular harmonic $z_{L_B}^{M_B}(\mathbf{r})$ (apart from a constant factor), can be translated to center A by well known formulas [32]. In the aligned frame, the formula reads:

$$z_l^m(\mathbf{r} - \mathbf{R}) = \sum_{k=|m|}^l \binom{l+|m|}{k+|m|} (-R)^{l-k} z_k^m(\mathbf{r}) \tag{19}$$

Once the functions are referred to center A , the one-center products of regular harmonics, appearing both in the one-center and in the translated two-center distributions, are decomposed in terms of regular harmonics as described in section, *Decomposition of products of regular harmonics*, in SF. The final radial factors are identified as the quantities that multiply the corresponding regular harmonics resultant from the decomposition.

In practice, the algorithm proceeds as follows:

1. The interval $0 \leq r \leq 20$ bohr is partitioned in n_{interv} subintervals with boundaries corresponding to previously selected values of r that will be noted as λ_i , $i = 0, 1, \dots, n_{interv}$ (currently $n_{interv} = 30$, see SF for details).
2. For each interval, the variable r is mapped onto the interval $[-1, 1]$ according to:

$$t \equiv 2 \frac{r - \lambda_{i-1}}{\lambda_i - \lambda_{i-1}} - 1 \tag{20}$$

and a set of values of t is chosen as the zeroes of the Chebyshev T polynomial of order n (currently $n = 30$) given by (see [33] 22.16.4):

$$t_k = -\cos[\pi (k - 1/2)/n] \quad k = 1, \dots, n \tag{21}$$

3. For each center, A , of the system, one-center distributions are expanded as:

$$\chi_a^A(\mathbf{r}_A) \chi_{a'}^A(\mathbf{r}_A) = \sum_l \sum_m f_{lm}^{L_a, M_a; L_{a'}, M_{a'}}(r) z_l^m(\mathbf{r}) \tag{22}$$

As described in section, *Expansion of one-center distributions*, of SF. The radial factors are evaluated in the tabulation points $r_{j,i} = (t_j + 1) (\lambda_i - \lambda_{i-1})/2 + \lambda_{i-1}$, multiplied by the $\rho_{aa'}$ element of the density matrix, and accumulated.

4. Likewise, for each center, A , a loop over all the remaining centers $B \neq A$ is performed. In this loop, for each center B , all the fragments d_{ab}^A coming from two-center distributions with one function at A and the other at B , and attributed to center A , are expanded in an aligned frame as a series of regular solid harmonics times radial factors

$$d_{ab}^A = \sum_l \sum_m f_{lm}^{L_a, M_a; L_b, M_b}(r) z_l^m(\mathbf{r}) \tag{23}$$

as described in section, *One-center expansion of two-center fragments*, of SF. The radial factors are evaluated in the tabulation points $r_{j,i} = (t_j + 1) (\lambda_i - \lambda_{i-1})/2 + \lambda_{i-1}$, multiplied by the $\tilde{\rho}_{ab}$ element of the density matrix, which has been previously rotated to the aligned frame (that is what the tilde means) and accumulated in the aligned frame. Next, the locally accumulated radial factors (i.e., for fixed B) are rotated back to the molecular frame, and the resultant radial factors are further accumulated together with those coming from the one-center distributions and with the radial factors of other pairs of centers to yield the full radial factor $\rho_{lm}^A(\mathbf{r}_A)$ of Equation (9). Details on rotations of both density matrix and radial factors are given in section, *Rotations*, in SF.

- The tabulations of the $\rho_{lm}^A(\mathbf{r}_A)$ radial factors are used to decide whether they are negligible or not and, for non-negligible factors, to carry out a numerical projection on Chebyshev T polynomials of variable t in each interval $r \in [\lambda_{i-1}, \lambda_i]$. This projection yields the corresponding piecewise expansion of $\rho_{lm}^A(\mathbf{r}_A)$. Details of this expansion are given in section *Expansion of atomic radial factors* in SF. Thus, the final expansion in the i -th interval takes the form:

$$\rho_{lm}^{i,A}(r_A) \simeq e^{-\zeta_i r_A} \sum_{k=0}^{n_i} c_k^{(i)}(l, m) T_k(t), \quad r_A \in [\lambda_{i-1}, \lambda_i] \quad (24)$$

where t is a function of r_A , as defined in (20), and the exponential factor is introduced when $\rho_{00}^A(r_A)$ (leading term in expansion (9)) decays steeply in the interval (see SF for details); otherwise, $\zeta_i = 0$ is taken. The number of polynomials taken in the expansion at the i -th interval, n_i , is determined on the fly by analyzing the convergence of the projections. The expansion coefficients, $c_k^{(i)}(l, m)$, of non-negligible factors are stored in a buffer. An array with a set of suitable pointers to address the coefficients is also generated and stored.

- Once the radial factors of expansion have been piecewise expanded, they are used to compute the auxiliary partial integrals:

$$Q_{lm}^A(\lambda_{i-1}, r_{j,i}) \equiv \frac{4 \pi}{2l + 1} \int_{\lambda_{i-1}}^{r_{j,i}} dr' r'^{2l+2} \rho_{lm}(r'), \quad r_{j,i} \in [\lambda_{i-1}, \lambda_i] \quad (25)$$

$$q_{lm}^A(\lambda_i, r_{j,i}) \equiv \frac{4 \pi}{2l + 1} \int_{r_{j,i}}^{\lambda_i} dr' r' \rho_{lm}(r'), \quad r_{j,i} \in [\lambda_{i-1}, \lambda_i] \quad (26)$$

in the same tabulation points, $r_{j,i}$, as used for the density, as well as the auxiliary constants:

$$Q_{lm}^A(\lambda_i) \equiv \frac{4 \pi}{2l + 1} \int_{\lambda_{i-1}}^{\lambda_i} dr' r'^{2l+2} \rho_{lm}(r') \quad (27)$$

and

$$q_{lm}^A(\lambda_i) \equiv \frac{4 \pi}{2l + 1} \int_{\lambda_{i-1}}^{\lambda_i} dr' r' \rho_{lm}(r') \quad (28)$$

Details are given in section, *Effective multipoles from density expansion*, in SF.

- The tabulations of $Q_{lm}^A(r_{j,i})$ and $q_{lm}^A(r_{j,i})$ are used to project these partial integrals onto Chebyshev T polynomials in the same intervals as used for the radial factors of density. In this case, no exponential factor is necessary:

$$Q_{lm}^A(\lambda_i, r) \simeq \sum_{k=0}^{n_i} b_k^{(i)}(l, m) T_k(t), \quad r \in [\lambda_{i-1}, \lambda_i] \quad (29)$$

$$q_{lm}^A(\lambda_i, r) \simeq \sum_{k=0}^{n_i} d_k^{(i)}(l, m) T_k(t), \quad r \in [\lambda_{i-1}, \lambda_i] \quad (30)$$

The numbers of polynomials in the intervals, n_i , are the same as in the corresponding radial factors. In this way, the pointers defined for addressing density expansion coefficients, $c_k^{(i)}(l, m)$, can be used also for coefficients $b_k^{(i)}(l, m)$ and $d_k^{(i)}(l, m)$ of Equations (29) and (30).

- Atomic point multipoles of Equation (14) are obtained by:

$$Q_{lm}^A = \sum_{i=1}^{n_{inter}} Q_{lm}^A(\lambda_i) \tag{31}$$

where the sum runs over the intervals.

9. Molecular geometry and data corresponding to the tabulation of radial factors are stored in an external binary file with extension `damqt`, ready to be used for computation of DAM expansion of density. In particular, the following information is stored: number of atoms, number of basis functions and number of shell functions, atomic number and Cartesian coordinates of nuclei, basis set, length of expansion (9) (l_{max}), and for each center A : pointers to expansion coefficients of radial factors, fitting exponents, ζ_i , and expansion coefficients, $c_k^{(i)}(l, m)$.
10. Atomic multipole moments Q_{lm}^A , auxiliary quantities $Q_{lm}^A(\lambda_i)$ and $q_{lm}^A(\lambda_i)$, and expansion coefficients $b_k^{(i)}(l, m)$ and $d_k^{(i)}(l, m)$ are stored in another external binary file with extension `dmqtv`. Since the pointers to b_k and d_k are the same as those used for a_k by construction, they do not require to be stored again.

3.2. Algorithm for Electrostatic Potential Expansion

Once the files containing the information on MED partition/expansion and the auxiliary quantities for MESP have been generated, MESP can be computed at any desired points using this information. This step is completely independent of the first one, so that the computation can be made as many times as necessary and in different sets of points without requiring repetition of the partition/expansion process.

To compute the MESP, the following algorithm is employed:

1. MED partition/expansion data stored in file `damqt` are read and stored in memory.
2. MESP auxiliary data are read from `dmqtv`, stored in memory and used for computing further auxiliary quantities. In particular, partial accumulated sums:

$$Q_{lm}^{i,A} = \sum_{k=0}^i Q_{lm}^A(\lambda_k) \quad i = 1, 2, \dots, n_{inter} \tag{32}$$

and

$$q_{lm}^{i,A} = \sum_{k=i}^{n_{inter}} q_{lm}^A(\lambda_k) \quad i = 1, 2, \dots, n_{inter} \tag{33}$$

are computed and stored too.

3. A double loop over atoms (outer) and tabulation intervals (inner) is performed to determine the length of expansion (11) in each interval and the long-range radius for the atom. This radius is chosen as the lower limit of the interval i , λ_{i-1} , for which $q_{lm}^{i,A}$ is lower than a user defined *long-range threshold*.
4. Next, MESP is computed, running over the atoms, with Equation (11). For points placed in the *long-range* region, lr , of atom A , the contribution to MESP, $V_{lr}^A(\mathbf{r}_A)$, is computed in terms of the corresponding atomic point multipoles Q_{lm}^A as:

$$V_{lr}^A(\mathbf{r}_A) = \frac{\zeta_A}{r_A} - \sum_{l=0}^{\infty} \sum_{m=-l}^l z_l^m(\mathbf{r}_A) \frac{Q_{lm}^A}{r_A^{2l+1}} \tag{34}$$

For points placed in the *short-range* region, sr , the contribution is computed by means of:

$$V_{sr}^A(\mathbf{r}_A) = \frac{\zeta_A}{r_A} - \sum_{l=0}^{\infty} \sum_{m=-l}^l z_l^m(\mathbf{r}_A) \left[\frac{Q_{lm}^A(r_A)}{r_A^{2l+1}} + q_{lm}^A(r_A) \right] \tag{35}$$

and the quantities $Q_{lm}^A(r_A)$ and $q_{lm}^A(r_A)$ are obtained in terms of $Q_{lm}^{i,A}$ and $q_{lm}^{i+1,A}$ of Equations (32) and (33), i being the index of the interval such that $\lambda_i \leq r_A < \lambda_{i+1}$, plus the expansions (29) and (30) for the integrals in the interval $[\lambda_i, r_A]$ and $[r_A, \lambda_{i+1}]$, respectively.

In all cases, the regular solid harmonics are fast and accurately computed by recursion, as described in section, *Recurrence relations of regular solid harmonics*, of SF. In the short-range case, eqs (29) and (30) are evaluated with the coefficients b_k and d_k previously retrieved from file `dmqt.v` and stored in memory, and with the Chebyshev polynomials computed by recursion, as shown in section, *Recurrence relations of Chebyshev polynomials*, of SF.

5. If MESP derivatives are wanted, they can be computed together with MESP and using the same auxiliary quantities[34]. The procedure is quoted in section, *Computing MESP derivatives*, in SF.
6. Data on basis set and density matrix are only necessary if computation of MESP in terms of nuclear attraction integrals and density matrix, without DAM partition/expansion, is required. As this is an expensive procedure, its usage should be restricted to those cases in which a reference is necessary for testing the accuracy of the algorithm reported here.

4. Results

To test the performance of the method, we have started by analyzing the accuracy of the results attained with the current algorithm. For this purpose, we have computed MESP values for benzene molecule in a set of equally spaced points corresponding to a $129 \times 129 \times 129$ grid in the octant defined by: $x, y, z \in [0, 20]$ (length in bohr). These results have been compared with the MESP *exact values*, V^{exact} , computed using the electron density matrix and the integrals involving basis set functions:

$$V^{exact}(\mathbf{r}_C) = \sum_{r=1}^{nbasis} \sum_{s=1}^r (2 - \delta_{rs}) \rho_{rs} \int d\mathbf{r} \frac{\chi_r(\mathbf{r}) \chi_s(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_C|} \tag{36}$$

where ρ_{rs} are the elements of the density matrix, and $nbasis$ stands for the number of basis functions, χ_r , in the LCAO calculation.

The results of this comparison are reported in Table 1 for four different lengths in expansion (9), namely: $l_{max} = 5, 10, 15, 20$. The highest absolute error, Δ_{max} , and the root mean square error, $rmse$, in the grid points significantly decrease with the length of multipole expansion, and they suggest that an expansion with $l_{max} = 10$ is sufficiently precise for most practical applications. Nevertheless, higher precision can be attained for more demanding applications like topological analysis at a very moderate cost.

Table 1. Precision of MESP computed with the current algorithm.

l_{max}	5	10	15	20
Δ_{max}^a	0.189×10^{-2}	0.305×10^{-4}	0.954×10^{-6}	0.119×10^{-6}
$rmse^b$	0.379×10^{-4}	0.264×10^{-6}	0.274×10^{-8}	0.164×10^{-9}
High prec ^c	660	1,239,142	1,996,569	2,097,390

^a Highest absolute error. ^b Root mean square error. ^c Number of grid points with more than eleven accurate decimal figures.

Remarkably, for expansions with $l_{max} = 10$ or higher, a great amount of points are computed with a high number of accurate decimal figures (twelve or more) as is shown in the last row of Table 1. This can be explained because short-range contributions in most points of the grid (those lying outside the molecular volume) are very small or even negligible, and precision is greatly determined by the

convergence of the long-range expansion, i.e., the number of terms in (9), the accuracy in the radial factors playing a minor role.

In this regard, it is interesting to check how the algorithm performs in those points in which short-range terms are important. For this purpose, the number of points with a number of decimal accurate figures is plotted in Figure 2 for the four expansions. As it can be seen, the number of points addressed in the curve corresponding to $l_{max} = 5$ is significantly greater than in the remaining curves. This is due to the insufficient convergence of the long-range MESP in this case, as mentioned before. In the remaining curves, the convergence in the long-range MESP has been achieved to a great extent, and the precision is mainly determined by the short-range terms. As the curves show, also in these cases the precision readily increases with the length of the expansion in two senses: a steady augment of the number of correct decimal figures in the least precise points is observed, and a significant decrease in the number of points for a given precision occurs. Raw data used for the figure can be found in section, *Precision of MESP calculation*, in SF.

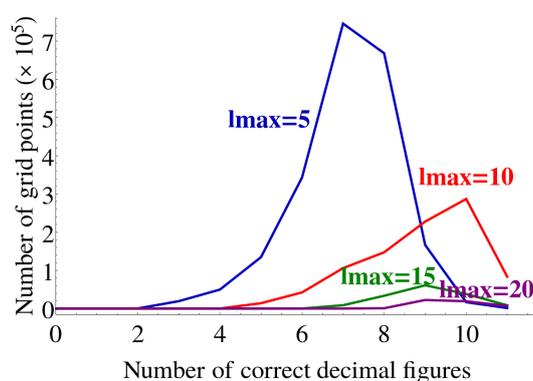


Figure 2. MESP precision.

The computation wall-clock time with Equation (36) was 5600 seconds, and 2.9, 6.6, 16.2 and 28.0 s for the respective expansions. On the other hand, DAM partition/expansion step with $l_{max} = 20$ took only 2.2 s for benzene density computed with Dunning's cc-pVDZ basis set [35]. These times were measured on a laptop with processor Intel(R) Core(TM) i7-6700HQ CPU @ 2.60 GHz, running an MPI parallel version of the codes compiled with gfortran, and using 4 processors. Although the calculation was made at HF level, it is important to stress that the computational cost of the full algorithm is independent of the computational level of the molecular density. This is so because the partition/expansion step only depends on the number of elements of the density matrix (i.e., the square of the basis set size), and the MESP computation only depends on the length of expansion (9) and the number of atoms in the system.

It is fairly evident that the algorithm based on DAM yields results that can be sufficient for most practical purposes at a computational time that, in this case, is between two and three orders of magnitude lower than the time required for computation from density matrix and integrals. This is so in a rather small system in which the short-range atomic contributions in the selected grid are about 20% of the total contributions, for a long-range threshold equal to 10^{-7} taken in cases of $l_{max} = 5, 10$, and about 40% for a threshold of 10^{-8} taken in cases of $l_{max} = 15, 20$. As it will be shown below, for really large systems, the fraction of short-range contributions in equivalent grids is much smaller, and the gain in the computational cost increases with the system size with respect to the computation by means of Equation (36). A further test on MESP surface extrema on a density isosurface has been included in the supplementary files.

Once the validity of the results has been established from the point of view of accuracy, we have analyzed the performance of the algorithm in large systems. In Table 2 we collect the results of MESP calculations in a set of molecules ranging from a small system like benzene (12 atoms, 222 basis functions) to two large ones: CC-MMIM BF_4 , consisting of a three circumcoronene slices with two

pairs of MMIM BF₄ ionic liquid molecules embedded between the circumcoronene sheets (617 atoms) and a DNA fragment with 750 atoms. In the last two, the electron densities correspond to PM7 calculations [36] (only basis valence functions).

Table 2. Computational cost in seconds of MESP computation with the current algorithm ^a. Percentage of long-range contributions to MESP for a threshold of 10⁻⁸.

Molecule	Benzene	Liposidomycin	DBCOx2 ^b	CC-MMIM BF ₄ ^c	DNA Fragment
n. atoms	12	71	360	617	750
n. basis ^d	222	1157	7560	1715	2250
partition/expansion time ^e	2.2	61	780	1.1 ^g	2.0 ^h
box size ^f	40	84	76	112	156
time MESP ($l_{max} = 10$)	7	36	230	195 ^g	254 ^h
time MESP ($l_{max} = 15$)	16	65	435	—	—
time MESP ($l_{max} = 20$)	43	105	617	—	—
% long-range ^h	60	95	93	99.5	99.99

^a MESP computed on a 129 × 129 × 192 grid (2,146,689 points) of a cubic box. ^b Dimer of polyhydroxylated circumcoronene. ^c Three sheets of circumcoronene with two pairs of MMIM BF₄ molecules embedded between the sheets. PM7 calculation. ^d Number of basis functions. ^e Time in DAM partition/expansion with $l_{max} = 20$. ^f Length in bohr of the cubic box edges. ^g Due to ZDO approximation, only terms up to $l_{max} = 2$ do not vanish. ^h Due to ZDO approximation, only terms up to $l_{max} = 4$ do not vanish.

With these results, we have analyzed the dependence of the computational cost of the two steps involved. Three different expansions have been used ($l_{max} = 10, 15, 20$), except in CC-MMIM BF₄ and DNA fragment. In these cases, due to the ZDO approximation involved in PM7 method, only valence one-center distributions contribute to densities. Consequently, the partition/expansion is a very fast process and terms in MED and MESP expansions with l_{max} higher than twice the highest L value in the basis set are zero.

As expected, the computational time of the partition/expansion step in small systems increases with the square of the number of basis functions, and at a lower rate when dealing with large systems, in which cases the dependence tends to be linear. The increment of the cost with respect to the length of expansion (11) is also smaller than the predicted l_{max}^2 dependence. In the supplementary files, a little movie showing the MESP of the DNA fragment over its molecular surface defined as MED isosurface with $\rho = 0.001$ au is included. Red color means positive MESP values, blue color, negative values.

Finally, in Table 3 we analyze the performance of MPI parallelization of the codes corresponding to both steps of the algorithm. Calculations have been carried out on a polyhydroxylated circumcoronene system with 360 centers, with a basis set consisting of 7560 contracted GTOs. The expansion of the atomic fragments has been made up to $l_{max} = 20$, and the MESP has been computed with this expansion on a 129 × 129 × 129 cubic grid (2,146,689 points) within an interval $x, y, z \in [-38, 38]$. The wall clock time, the average time per processor and the standard deviation are provided for the DAM partition/expansion and for the MESP tabulation.

In both steps, the computational time scales very well with the number of processors. In the partition/expansion step, a linear fit of the clock wall time vs. the reciprocal of the number of processors gives a regression parameter of $R^2 = 0.9998$, and the same value is obtained for the fit of the average time. In case of MESP tabulation, the scaling is likewise, with $R^2 = 0.9982$ and $R^2 = 0.9990$, respectively.

Table 3. Performance of the algorithm parallelization. All times in seconds.

N. Procs ^a	Partition/Expansion			MESP Tabulation ^b		
	Wall Clock ^c	Average ^d	Std. Dev. ^e	Wall Clock ^c	Average ^d	Std. Dev. ^e
1	2165	—	—	2048	2048	—
2	1125	1124	0.2	992	989	4
4	670	669	0.8	552	520	26
6	531	530	0.8	406	366	32
8	466	465	0.5	335	293	24
10	402	400	0.9	272	237	21

^a Number of MPI processes on a system with an Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz with 10 cores.

^b Computation on a $129 \times 129 \times 129$ grid (2,146,689 points). ^c Highest time among processors. ^d Average time per processor. ^e Standard deviation.

The standard deviations show that the time is more evenly shared among processors in the partition/expansion than in the tabulation, but in both cases the time distribution is satisfactory. Furthermore, in the second step the standard deviation is affected by the fact that the main processor spends more time than the remaining ones, due to the workload associated to the gathering of tabulations accomplished by ancillary processors, which are stored in external files, and to tidying tasks.

5. Conclusions

A method for the efficient computation of molecular electrostatic potential has been reported. The method splits the problem in two steps. The first one consists of the partition of the molecular electron density in terms of (pseudo)atomic densities, which are further expanded as a truncated series of spherical harmonics times radial factors. The second step corresponds to the computation of the electrostatic potential from these expansions. The computational cost of the first step depends on the size of the basis set used in the LCAO computation and the length of the expansions, and it is independent of the number of points in which the potential is computed. The cost of the second step is independent of the basis set size and only depends on the expansions length and the number of tabulation points. The result is an algorithm that reduces the computational time around two orders of magnitude when compared with the calculation from electron density matrix and integrals involving the basis set, for a precision of 6 decimal figures. The algorithm can be very efficiently parallelized with MPI, showing a good sharing of computational cost among processors in both steps. The algorithm has been implemented in DAMQT package for the analysis of molecular electron density and related properties. A recent version of the package can be downloaded freely from: <https://data.mendeley.com/datasets/7mwfftd2x4/2>.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/2079-3197/7/4/64/s1>.

Author Contributions: Conceptualization: R.L., I.E., J.M.G.d.l.V., G.R.; Methodology: R.L., G.R., I.E.; Software: R.L., I.E., F.M., G.R.; Supervision: J.M.G.d.l.V.; Validation: F.M.; Writing: R.L., J.M.G.d.l.V.

Funding: This research received no external funding.

Acknowledgments: We wish to thank Noel Ferro for supplying us the Catechin data for testing MESP on a density isosurface.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CGTO	Contracted Gaussian type orbital
CPU	Central processing unit
GTO	Gaussian type orbital (primitive)
DAM	Method of deformed atoms in molecules
DAMQT	Package for the analysis of electron density and related properties
LCAO	Linear combination of atomic orbitals
MED	Molecular electron density
MESP	Molecular electrostatic potential
MMIM	Dimethyl imidazolium
MPI	Message passing interface
ZDO	Zero differential overlap

References

1. Lewis, G.N. The atom and the molecule. *J. Am. Chem. Soc.* **1916**, *38*, 762–785. [[CrossRef](#)]
2. Politzer, P.; Murray, J.S. An Overview of Strengths and Directionalities of Noncovalent Interactions: σ -Holes and π -Holes. *Crystals* **2019**, *9*, 165. [[CrossRef](#)]
3. Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **1927**, *389*, 457–484. [[CrossRef](#)]
4. Mulliken, R.S. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. *J. Chem. Phys.* **1955**, *23*, 1833–1840. [[CrossRef](#)]
5. Bader, R.F.W.; Henneker, W.H. Molecular Charge Distributions and Chemical Binding. *J. Chem. Phys.* **1967**, *46*, 3341–3363. [[CrossRef](#)]
6. Bader, R.F.W. *Atoms in Molecules: A Quantum Theory*; Clarendon Press: Oxford, UK, 1990.
7. Scrocco, E.; Tomasi, J. The electrostatic molecular potential as a tool for the interpretation of molecular properties. In *New Concepts II*; Springer: Berlin/Heidelberg, Germany, 1973; Volume 42, pp. 95–170, doi:10.1007/3-540-06399-4_6. [[CrossRef](#)]
8. Scrocco, E.; Tomasi, J. Electronic molecular structure, reactivity and intermolecular forces: An heuristic interpretation by means of electrostatic molecular potentials. *Advan. Quantum Chem.* **1978**, *11*, 116–193.
9. Sjöberg, P.; Politzer, P. Use of the Electrostatic Potential at the Molecular Surface to Interpret and Predict Nucleophilic Processes. *J. Phys. Chem.* **1990**, *94*, 3959–3961. [[CrossRef](#)]
10. Jane S, M.; Lane, P.; Brinck, T.; Politzer, P. Relationships between Computed Molecular Properties and Solute-Solvent Interactions in Supercritical Solutions. *J. Phys. Chem.* **1993**, *97*, 5144–5148.
11. Murray, J.S.; Lane, P.; Brinck, T.; Paulsen, K.; Grice, M.E.; Politzer, P. Relationships of Critical Constants and Boiling Points to Computed Molecular Surface Properties. *J. Phys. Chem.* **1993**, *97*, 9369–9373. [[CrossRef](#)]
12. Murray, J.S.; Peralta-Inga, Z.; Politzer, P.; Ekanayake, K.; Lebreton, P. Computational Characterization of Nucleotide Bases: Molecular Surface Electrostatic Potentials and Local Ionization Energies, and Local Polarization Energies. *Int. J. Quantum Chem.* **2001**, *83*, 245–254. [[CrossRef](#)]
13. Murray, J.S.; Paulsen, K.; Politzer, P. Molecular surface electrostatic potentials in the analysis of non-hydrogen-bonding noncovalent interactions. *Proc. Indian Acad. Sci.* **1994**, *106*, 267–275.
14. Clark, T.; Hennemann, M.; Murray, J.; Politzer, P. Halogen bonding: The σ -hole. *J. Mol. Model.* **2007**, *13*, 291–296. doi:10.1007/s00894-006-0130-2. [[CrossRef](#)] [[PubMed](#)]
15. Pathak, R.K.; Gadre, S.R. Maximal and minimal characteristics of molecular electrostatic potentials. *J. Chem. Phys.* **1990**, *93*, 1770–1773. [[CrossRef](#)]
16. Gadre, S.R.; Pathak, R.K. Nonexistence of local maxima in molecular electrostatic potential maps. *Proc. Ind. Acad. Sci. (Chem. Sci.)* **1990**, *102*, 189–192.
17. Gadre, S.R.; Shrivastava, I.H. Shapes and sizes of molecular anions via topographical of electrostatic potential analysis. *J. Chem. Phys.* **1991**, *94*, 4384–4390. [[CrossRef](#)]
18. Gadre, S.R.; Kulkarni, S.A.; Suresh, C.H.; Shrivastava, I.H. Basis set dependence of the molecular electrostatic potential topography. A case study of substituted benzenes. *Chem. Phys. Lett.* **1995**, *239*, 273–281. [[CrossRef](#)]
19. Roy, D.; Balanarayan, P.; Gadre, S.R. An appraisal of Poincaré–Hopf relation and application to topography of molecular electrostatic potentials. *J. Chem. Phys.* **2008**, *129*, 174103. [[CrossRef](#)] [[PubMed](#)]

20. Kumar, A.; Gadre, S.R. Exploring the Gradient Paths and Zero Flux Surfaces of Molecular Electrostatic Potential. *J. Chem. Theory Comput.* **2016**, *12*, 1705–1713. [[CrossRef](#)] [[PubMed](#)]
21. Kumar, A.; Gadre, S.R. Molecular Electrostatic Potential Based Atoms in Molecules: Shielding Effects and Reactivity Patterns. *Aust. J. Chem.* **2016**, *69*, 975–982. [[CrossRef](#)]
22. Alkorta, I.; Martín-Fernández, C.; Montero-Campillo, M.M.; Elguero, J. Hydrogen-Bonding Acceptor Character of Be₃, the Beryllium Three- Membered Ring. *J. Phys. Chem.* **2018**, *122*, 1472–1478. [[CrossRef](#)] [[PubMed](#)]
23. Alkorta, I.; Elguero, J.; Bene, J.E.D. Complexes of O=C=S with Nitrogen Bases: Chalcogen Bonds, Tetreln Bonds, and Other Secondary Interactions. *Chem. Phys. Chem.* **2018**, *19*, 1886–1894. [[CrossRef](#)] [[PubMed](#)]
24. Tahir, M.N.; Ashfaq, M.; de la Torre, A.F.; Caballero, J.; W.Hernández-Rodríguez, E.; Ali, A. Rationalizing the stability and interactions of 2,4-diamino-5-(4-chlorophenyl)-6-ethylpyrimidin-1-ium 2-hydroxy-3, 5-dinitrobenzoate salt. *J. Mol. Struct.* **2019**, *1193*, 185–194. [[CrossRef](#)]
25. Kote, S.R.; Albuquerque, J.V.; Shirsat, R.N.; Phadke, R.V.; Kohapare, J.T.; Dhoble, S.S. TLC Detection and Theoretical Structure Elucidation of Nitrogen Containing Compounds with Cobalt Thiocyanate. *Anal. Chem. Lett.* **2019**, *9*, 453–462. [[CrossRef](#)]
26. Rico, J.F.; López, R.; Ema, I.; Ramírez, G.; Ludeña, E. Analytical method for the representation of atoms-in-molecules densities. *J. Comput. Chem.* **2004**, *25*, 1355–1363. [[CrossRef](#)] [[PubMed](#)]
27. Rico, J.F.; López, R.; Ema, I.; Ramírez, G. Electrostatic potentials and fields from density expansions of deformed atoms in molecules. *J. Comput. Chem.* **2004**, *25*, 1347–1354. [[CrossRef](#)] [[PubMed](#)]
28. Rico, J.F.; López, R.; Ramírez, G. Analysis of the molecular density. *J. Chem. Phys.* **1999**, *110*, 4213–4220. [[CrossRef](#)]
29. Gradshteyn, I.S.; Ryzhik, I.M. *Table of Integrals, Series and Products*, 4th ed.; Academic Press: New York, NY, USA, 1980.
30. Stone, A.; Alderton, M. Distributed Multipole Analysis. *Mol. Phys.* **1985**, *56*, 1047–1064. [[CrossRef](#)]
31. Kaufmann, K.; Baumeister, W. Single-centre expansion of Gaussian basis functions and the angular decomposition of their overlap integrals. *J. Phys. B At. Mol. Opt. Phys.* **1989**, *22*, 1–12. [[CrossRef](#)]
32. Rico, J.F.; López, R.; Ema, I.; Ramírez, G. Translation of Real Solid Spherical Harmonics. *Int. J. Quantum Chem.* **2013**, *113*, 1544–1548. [[CrossRef](#)]
33. Abramowitz, M.; Stegun, I. *Handbook of Mathematical Functions*, 9th printing ed.; Dover: Long Island, NY, USA, 1970.
34. López, R.; Rico, J.F.; Ramírez, G.; Ema, I.; Zorrilla, D.; Kumar, A.; Yeole, S.D.; Gadre, S.R. Topology of molecular electron density and electrostatic potential with DAMQT. *Comput. Phys. Commun.* **2017**, *214*, 207–215. [[CrossRef](#)]
35. Dunning, T.H., Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023. [[CrossRef](#)]
36. Stewart, J.J.P. Optimization of parameters for semiempirical methods VI: More modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1–32. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).