Journal of Molecular Graphics and Modelling 38 (2012) 314–323

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Quantitative analysis of molecular surface based on improved Marching Tetrahedra algorithm

Tian Lu[a,b], Feiwu Chen[a,b,*]

[a] Department of Chemistry and Chemical Engineering, School of Chemistry and Biological Engineering, University of Science and Technology Beijing, Beijing 100083, People's Republic of China

[b] Beijing Key Laboratory for Science and Application of Functional Molecular and Crystalline Materials, Beijing 100083, People's Republic of China

ARTICLE INFO

Article history:
Accepted 19 July 2012
Available online 27 July 2012

Keywords:
Molecular surface
Marching Tetrahedra
Electrostatic potential
Average local ionization energy
Interpolation accuracy

ABSTRACT

Quantitative analysis of molecular surface is a valuable technique for analyzing non-covalent interaction, studying molecular recognition mode, predicting reactive site and reactivity. An efficient way to realize the analysis was first proposed by Bulat et al. (J. Mol. Model., 16, 1679), in which Marching Tetrahedra (MT) approach commonly used in computer graphics is employed to generate vertices on molecular surface. However, it has been found that the computations of the electrostatic potential in the MT vertices are very expensive and some artificial surface extremes will be presented due to the uneven distribution of MT vertices. In this article, we propose a simple and reliable method to eliminate these unreasonably distributed surface vertices generated in the original MT. This treatment can save more than 60% of total analysis time of electrostatic potential, yet the loss in accuracy is almost negligible. The artificial surface extremes are also largely avoided as a byproduct of this algorithm. In addition, the bisection iteration procedure has been exploited to improve accuracy of linear interpolation in MT. The most appropriate grid spacing for surface analysis has also been investigated. 0.25 and 0.20 bohr are recommended to be used for surface analysis of electrostatic potential and average local ionization energy, respectively.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

In a molecular system, the electrostatic potential (ESP) can be written as:

$$V_{\mathrm{Total}}(\mathbf{r}) = V_{\mathrm{Nuc}}(\mathbf{r}) + V_{\mathrm{Elec}}(\mathbf{r}) = \sum_{\mathrm{A}} \frac{Z_{\mathrm{A}}}{|\mathbf{r} - \mathbf{R}_{\mathrm{A}}|} - \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \qquad (1)$$

where Z and $\mathbf{R}$ denote nuclear charge and nuclear position, respectively. ESP measures the electrostatic interaction between a unit point charge placed at $\mathbf{r}$ and the system of interest. A positive (negative) value implies that current position is dominated by nuclear (electronic) charges.

ESP has been widely used for prediction of nucleophilic and electrophilic sites as well as molecular recognition mode for a long time [1–4]. The theoretical basis is that molecules always tend to approach each other in a complementary manner of ESP. Analyses of ESP are common performed on molecular van der Waals (vdW) surface. Although the definition of such a surface is somewhat arbitrary, most people prefer to follow Bader et al. [5] and choose isodensity surface with electron density $\rho = 0.001$ e/bohr^3 as vdW surface, since this definition reflects the specific features of the particular molecule, such as lone pairs and π electrons [6]. This is also the definition of vdW surface used in present paper.

Analyses of ESP on vdW surface have been further quantified to extract more information. It has been shown that strength and orientation of many non-covalent interactions, such as the hydrogen bonding, halogen bonding and π-hole bonding, can be well predicted and explained by analyzing the magnitude and positions of minima and maxima on vdW surface [7–10]. Politzer and coworkers [6,11] have defined a set of molecular descriptors based on ESP distribution on vdW surface, which are taken as independent variables of general interaction properties function (GIPF). GIPF successfully connected the distribution of ESP on vdW surface and many condensed phase properties, including boiling point, surface tension, heats of vaporization, and so on. These vdW surface descriptors are outlined below.

$V_{\mathrm{S,min}}$ and $V_{\mathrm{S,max}}$ correspond to minimum and maximum of ESP on vdW surface respectively. A_{S}, A_S^+ and A_S^- are total surface area and the surface areas where ESP is positive and negative, respectively. $\overline{V}_S^+$, $\overline{V}_S^-$ and $\overline{V}_S$ denote average of positive, negative and overall ESP on vdW surface, respectively.

$$\overline{V}_{\mathrm{S}}^+ = \left(\frac{1}{m}\right)\sum_{i=1}^{m} V(\mathbf{r}_i) \quad \overline{V}_{\mathrm{S}}^- = \left(\frac{1}{n}\right)\sum_{j=1}^{n} V(\mathbf{r}_j) \quad \overline{V}_S = \left(\frac{1}{t}\right)\sum_{k=1}^{t} V(\mathbf{r}_k) \qquad (2)$$

* Corresponding author at: Department of Chemistry and Chemical Engineering, School of Chemistry and Biological Engineering, University of Science and Technology Beijing, Bejing 100083, People's Republic of China. Tel.: +86 10 62332689.
E-mail address: chenfeiwu@ustb.edu.cn (F. Chen).

1093-3263/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2012.07.004

where i, j and k are indices of sampling points in positive, negative and entire regions respectively, t is the total number of surface vertices. Π is expressed as follows:

$$\Pi = \left(\frac{1}{t}\right)\sum_{k=1}^{t}[V(\mathbf{r}_k) - \overline{V}_S] \tag{3}$$

which can be viewed as the average deviation of ESP over the surface, an indicator of charge separation. The total ESP variance can be written as a sum of its positive and negative parts:

$$\sigma_{\text{tot}}^2 = \sigma_+^2 + \sigma_-^2 = \left(\frac{1}{m}\right)\sum_{i=1}^{m}[V(\mathbf{r}_i) - \overline{V}_S^+]^2 + \left(\frac{1}{n}\right)\sum_{j=1}^{n}[V(\mathbf{r}_j) - \overline{V}_S^-]^2 \tag{4}$$

σ_{tot}^2 reflects the variability of ESP on vdW surface. The larger the σ_+^2 and σ_-^2 are, the more tendency that the molecule interacts with other molecules by its positive and negative ESP regions, respectively.

Degree of charge balance is defined as:

$$\nu = \frac{\sigma_+^2\sigma_-^2}{(\sigma_{\text{tot}}^2)^2} \tag{5}$$

when σ_+^2 equals to σ_-^2, ν attains its maximum of 0.25. The closer ν is to 0.25, the more possible that a molecule can interact with others through its positive or negative ESP region.

The product of σ_{tot}^2 and ν is a very useful quantity too. A large value of $\nu\sigma_{\text{tot}}^2$ is indicative of a molecule that has relatively strong tendency to interact electrostatically with another molecule with its own kind.

Another noteworthy real space function is the average local ionization energy $\bar{I}$, which has attracted more and more attentions in recent years. $\bar{I}$ is defined as [12].

$$\bar{I}(\mathbf{r}) = \frac{\sum_i \rho_i(\mathbf{r})|\varepsilon_i|}{\rho(\mathbf{r})} \tag{6}$$

where ρ_i and ε_i are the electron density and orbital energy of the i-th molecular orbital, respectively. $\bar{I}$ has widespread uses [13,14] such as reproducing atomic shell structure, measuring electronegativity, quantifying local polarizability and hardness. However, its most important use maybe the prediction of reactivity according to the function value on vdW surface, denoted as $\bar{I}_S$. Lower value of $\bar{I}_S$ indicates that the electrons at that point are more weakly bounded. Therefore, the surface minima of $\bar{I}_S$ indicates the most likely site of electrophilic or radical attack.

An efficient molecular surface analysis algorithm has been proposed by Bulat et al. [13] for locating local extremes of ESP and $\bar{I}$ on vdW surface and for calculation of the surface descriptors listed above. Their algorithm is mainly based on Marching Tetrahedra approach (MT), which is well-known in the field of computer graphics. The calculation process can be summarized as steps below: (1) Using MT approach to construct the vdW surface by numerous triangular facets. (2) Calculating ESP or $\bar{I}$ at all triangular vertices on the surface. (3) Locating surface extremes by comparing function value of each vertex and its adjacent vertices. (4) Calculating surface descriptors.

During our implementation of this algorithm, we found that the biggest bottleneck of calculation speed is the step 2, owing to the large amount of calculations of ESP at numerous surface vertices. By carefully inspection of the vertices scattered on vdW surface, we found many vertices produced by MT approach are clustered together unreasonably. Therefore, if some of clustered vertices are deleted, it can be expected that the computational cost would be greatly reduced without significant loss of accuracy of the results. Our primary aim in this paper is to present a simple and efficient way to eliminate these unreasonably distributed vertices.

The rest of the article is organized as follows: Section 2 briefs basic principle of MT approach, Section 3 introduces the specific process of the molecular surface analysis algorithm implemented in our wavefunction analysis program Multiwfn [15,16]. In Section 4, we analyze the shortcoming of MT approach and propose a method to resolve it. In Section 5, effect and reliability of our method are examined by examples. Besides, we demonstrate the insufficiency of the linear interpolation commonly used in MT process and show the accuracy can be largely improved by using bisection iteration method. The optimal choice of grid spacing for surface analysis is also explored and discussed. In the final section we conclude this paper.

2. Marching Tetrahedra approach

So far, the most popular approach used to generate isodensity surface of volume data in computer graphics realm may be Marching Cubes (MC) [17]. The two main drawbacks of the MC approach [18–21] are (1) the so-called "ambiguity problem", which sometimes results in erroneous shape of isodensity surface; (2) lengthy looking-up table needed to be constructed, which make the approach difficult to be implemented. MT is an extension of MC [22]. The two shortcomings of MC are completely avoided. The basic idea of MT may be easier to be understood in two-dimension case, see Fig. 1, which displays how MT generates a specific isodensity contour of function $1/(x^2+y^2)$, which is spherically symmetric and derived from the center of the figure.

At the first step, a suitable spatial region is chosen, which should be large enough to avoid truncation of isodensity contour. Grid spacing is also suitably selected according to the expected accuracy. Smaller spacing leads to more exact isodensity contour and of course to much more computational efforts. The region is then evenly divided into many cubes, see Fig. 1a.

At the second step, the value of the function is calculated at each corner of cubes, then cubes are classified into three categories: (1) internal cubes: function values at all corners of the cube are larger than the isodensity value; (2) external cubes: function values at all corners of the cube are smaller than the isodensity value; (3) boundary cubes: function values at the cube corners are not all greater or less than the isodensity value, these cubes are rendered as yellow in Fig. 1b. The green circle corresponds to the exact isodensity contour of the function. The function value in the region inside (outside) the circle is larger (smaller) than the isodensity value.

At the third step, internal and external cubes are discarded, each boundary cubes is divided into two triangles. Through the use of already calculated function values at cube corners, the intersection points between isodensity contour and triangles can be evaluated by interpolation techniques, these points are shown by red points in Fig. 1c.

At the final step, intersection points in the same triangle are connected by line. Now the approximate isodensity contour of the function has been generated, as portrayed by blue polygon in Fig. 1d.

Extending above idea to a three-dimension case is straightforward. Once boundary cubes are identified according to the similar manner mentioned above, these cubes are broken up to five, six or more tetrahedra. Although the break-up mode is not unique [18], we prefer the one shown in Fig. 2. The advantage of this mode is that the same break-up mode can be applied to all cubes without leading to discontinuity within the isodensity surface [18] since two

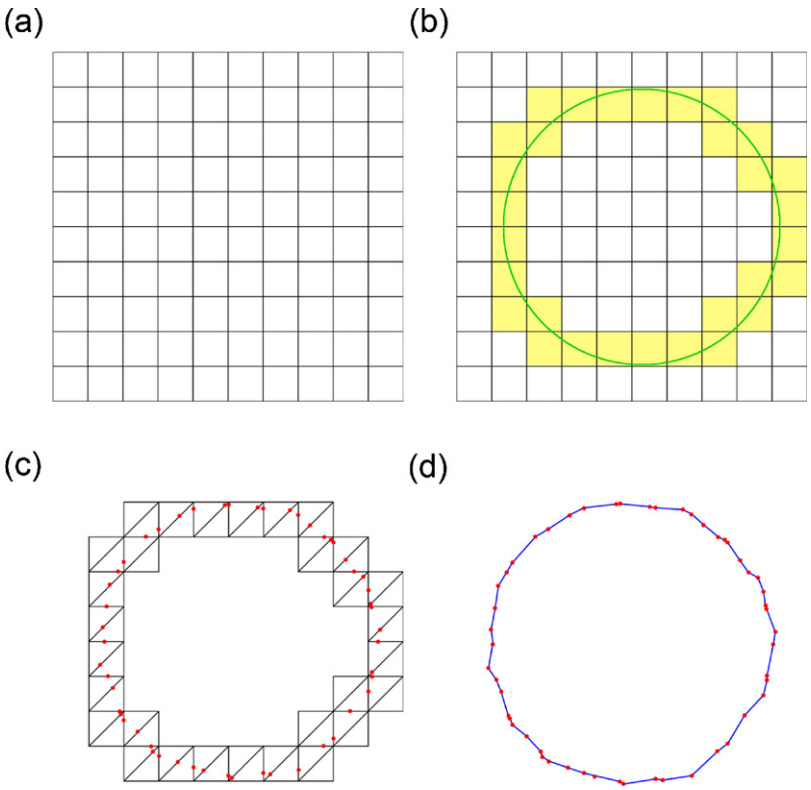
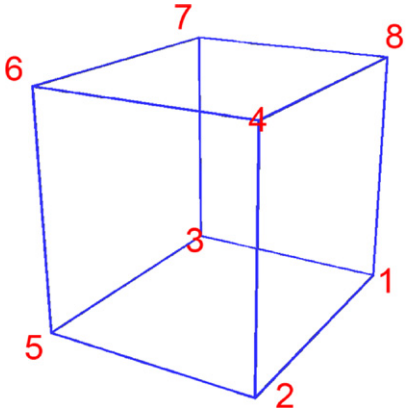


Fig. 1. Steps of using MT approach to construct a specific contour line of function $1/(x^2+y^2)$, see the description of the text. Green circle in (b) is the exact isodensity contour of the function, blue polygon in (d) is the approximate isodensity contour generated by MT approach. Red points in (c) are intersection points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)



Tetrahedron	Vertices
1	4 3 5 2
2	4 3 5 6
3	4 3 7 6
4	4 3 7 8
5	4 3 1 8
6	4 3 1 2

Fig. 2. Each boundary cube is broken up to six tetrahedra.

opposite faces of a cube share the same pattern. After that, function values at two endpoints of each edge of a tetrahedron are compared with the specified isodensity value, $\rho_{iso} = 0.001$ e/bohr^3. If the isodensity value ρ_{iso} lies between the electron density values of the two endpoints, it is clear that the edge linking these two endpoints intersects the isodensity surface. Such edge is regarded as "active edges". Intersection points on active edges are commonly evaluated by simple linear interpolation. Tetrahedra can be classified

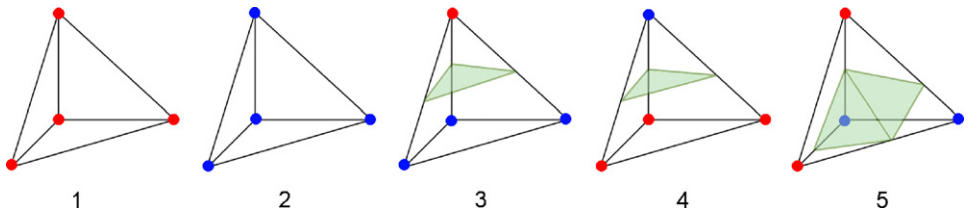


Fig. 3. Five types of tetrahedra involved in MT approach. Red and blue points denote that function values at corresponding vertices are larger and smaller than the isodensity value respectively. Transparent green triangles are newly generated facets on the isodensity surface. Types 1 and 2 correspond to internal and external tetrahedra respectively, the other three are boundary tetrahedra. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

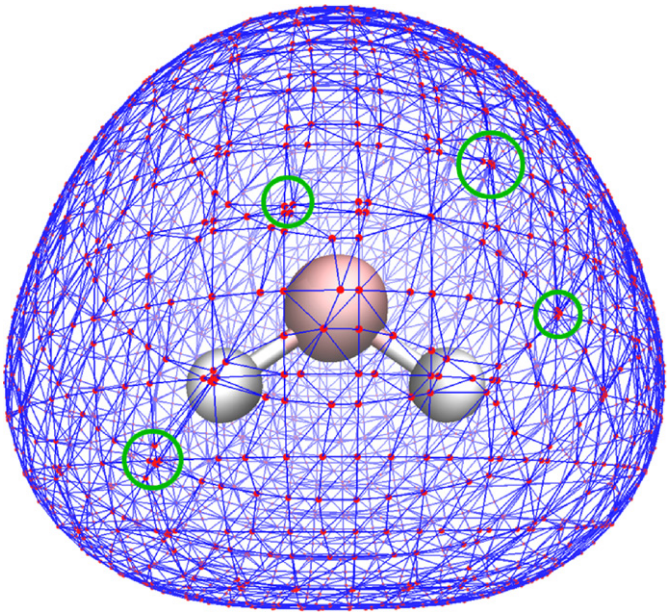


Fig. 4. Isodensity surface of $\rho = 0.001$ e/bohr^3 of water molecule generated by MT approach. Red points and blue lines correspond to vertices and edges of the facets on the isodensity surface respectively. Some regions where surface vertices heavily aggregated are highlighted by green circles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

into five types as shown in Fig. 3. By linking intersection points, each boundary tetrahedron produces one (types 3 and 4) or two (type 5) triangular facets, corresponding to the transparent green triangles in Fig. 3. All these newly generated facets collectively constitute the expected isodensity surface, see Fig. 4 for example. These intersection points will be called as surface vertices later.

It is obvious that if smaller grid spacing is used as an initial input of MT process, there will be more vertices generated. Therefore, the surface formed from these vertices will be closer to the exact isodensity surface.

3. Process of quantitative analysis of molecular surface and our implementation

The flow chart of surface analysis module in our program Multiwfn (written in Fortran90) is given in Fig. 5. In the beginning stage, one can adjust many parameters such as the spacing of grid points and the isodensity value of electron density surface in an interactive interface. The spatial range of the volume data to be calculated is automatically and adaptively determined according to molecular geometry and vdW radius of boundary atoms. Then volume data of electron density is evaluated. After that the program enters MT stage.

The MT stage in our code basically follows the algorithm described in the last section. The main process has shown in Fig. 5. Some technical details are mentioned below:

(1) Since adjacent tetrahedra have one or more shared edges, the generated intersection points may be duplicated. In order to avoid this, the correspondence between corner indices of boundary cubes and indices of generated intersection points are recorded in *surfcor2vtx* array. During each cycle of active edges, *surfcor2vtx* array is checked first to detect if the intersection point corresponding to this edge has already been generated.
(2) The connectivities of surface vertices are recorded in *vtxconn* array, which is two-dimensional integer type. For example $i = vtxconn(j,k)$ means vertex i is linked to vertex j, where k is the slot index used to record this connectivity.
(3) The linear interpolation scheme in MT process will lead to too low accuracy in surface analysis and to some artificial appearances or erroneous locations of surface extremes. Bulat et al. have noticed this problem and proposed to use the cubic interpolation instead [13]. However, realization of cubic interpolation is complicated and the gradients of electron density at two endpoints of the active edge have also to be calculated. Here we present a simple alternative scheme to treat this problem. In order to make sure that higher linear interpolation accuracy can be achieved in the scheme, a few bisection iterations are carried out first between two endpoints to shrink their interval. Then a linear interpolation is performed within the refined interval. Since electron density can be computed very quickly, this scheme is not only more efficient but also easy to be implemented. We noted that similar idea has also been used by Chan and Purisima [23]. The numerical examination will be given in Section 5.

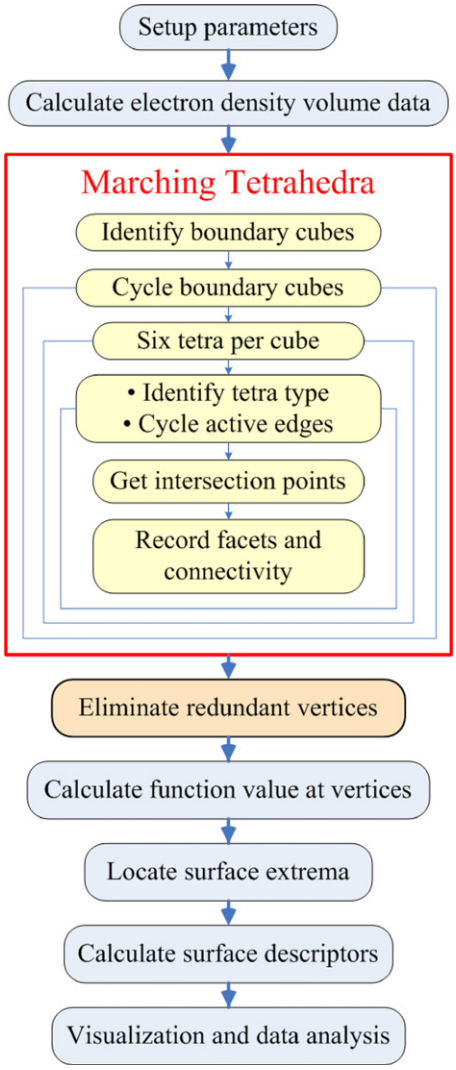


Fig. 5. Flow chart of quantitative molecular surface analysis module in Multiwfn.

After MT stage, elimination of some of unreasonably distributed vertices is carried out. Then the function value (ESP or $\bar{I}$) at vertices are computed and stored. Calculation of ESP is very time-consuming. Fortunately, the calculation can be readily

parallelized, since ESP values at these vertices are independent of each other.

Next, we follow the 2-neighborhood method proposed by Bulat et al. [13] to locate surface extremes. If function value of a vertex is simultaneously larger (smaller) than that of its first-shell neighbors and second-shell neighbors defined by connectivity, then the vertex will be recognized as local maximum point (local minimum point). Due to the aforementioned rule of recording connectivity in *vtxconn* array, scanning all of first-shell and second-shell neighbors of a vertex is quite easy and fast.

In the calculation of surface descriptors stage, we first calculate area A and function value of each facet, the latter could be simply obtained by averaging function values at its three vertices. Then evaluation of the quantities shown in Eqs. (2)–(5) is straightforward, for example $\overline{V}_S^+$ and σ_+^2 can be calculated as:

$$\overline{V}_S^+ = \left(\frac{1}{A_{tot}^+}\right)\sum_{i=1}^{m} V_i A_i \quad A_{tot}^+ = \sum_{i=1}^{m} A_i$$
$$\sigma_+^2 = \left(\frac{1}{A_{tot}^+}\right)\sum_{i=1}^{m} [(V_i - \overline{V}_S^+)]^2 A_i \tag{7}$$

where i is the index of the facets of which the function value is positive. Facet areas play the role of weighting function.

In the final stage, positions and function values of surface extrema as well as all surface descriptors are outputted. The built-in graphical module in Multiwfn allows surface extrema to be visualized directly in an interactive environment. Users can also choose to modify or export the calculated data.

4. Eliminating redundant vertices

Through many visual inspections of the surface vertices produced by standard MT approach, we found a large portion of vertices often heavily aggregated into many tiny regions. In Fig. 4 some of such regions are highlighted by green circles. Aggregated vertices mainly present where isodensity surface just intersecting corners of boundary cubes. We have noted that different break-up modes of boundary cubes result in different distribution of surface vertices, however no mode is free of this phenomenon. Because of the very close distances among the aggregated vertices, the areas of the facets formed from these vertices are very small and almost have no contribution to qualities of surface descriptors. Since these vertices are nearly meaningless for molecular surface analysis, they will be referred as "redundant vertices" later. In practical applications, almost all of computational time in molecular surface analysis task is used to calculate ESP at surface vertices. It can be expected that if redundant vertices are properly eliminated, a great deal of time would be saved without detectable loss of accuracy, which will facilitate greatly the application of quantitative molecular surface analysis on larger system or virtual screening task.

Some methods are proposed to avoid redundant vertices, for example, the regularized Marching Tetrahedra algorithm [24] and a class of mesh simplification methods [25–27]. However, these methods are either complicated or not directly applicable to quantitative molecular surface analysis. Here we present an effective method to eliminate redundant vertices, which may be viewed as a special case of mesh simplification methods. The basic idea is simple: Each pair of connected vertices are checked in turn, if distance between two vertices (referred as a and b) is found to be smaller than a specific criteria, then vertex b is moved to their average position, and then vertex a is merged into vertex b. Meantime vertex connectivity list as well as vertex indices of related facets are updated correspondingly. Finally, any facet containing two vertices with identical index will be discarded. After many numerical tests, 0.4 times of grid spacing is chosen to be the distance criteria.

Our elimination method has three advantages: (1) very fast; (2) very easy to be implemented; (3) the elimination step can be directly inserted after MT stage without modifying original MT code. The pseudo-code for realizing the method in our program is given below, the complete code is merely about 70 lines and can be found in surfana.f90 file in source code package, which can be freely download [16].

```
do vtx1 = 1, all vertices
  If vtx1 has been eliminated then skip
  do vtx2 = vtx1 + 1, all vertices
    If vtx2 has been eliminated then skip
    if (distance between vtx1 and vtx2 < criteria) then
      do inei = 1, neighbors of vtx2
        If inei is eliminated or inei == vtx1 then skip
        If inei has not connected to vtx1 then make them connected by
  updating vtxconn array
      end do
      Marking vtx2 as eliminated
      Set coordinate of vtx1 to the average position of vtx1 and vtx2
      Recording the process "vtx1 has been merged to vtx2" in lookup
  table mergerelat
    end if
  end do
end do
do itri = 1, all facets
  do ivtx = 1, the three vertices in facet itri
    if (ivtx has been eliminated) then
Setting index of ivtx to which vertex that ivtx is finally merged to by
  looking up mergerelat
    end if
  end do
end do
do itri = 1, all facets
  if (indices of any two vertices in itri are identical) then
    Marking itri as eliminated
  end if
end do
```

Finally, one should pay attention to some unreasonable vertices and related facets after the elimination process. The vertices with only two neighbors (green point in pattern 3 of Fig. 6) and related facets (green triangle in pattern 3 of Fig. 6) are needed to be searched and eliminated. Otherwise, the calculated surface descriptors will be inaccurate. Such vertices and facets are occurred during the vertices merging process as illustrated by patterns 1 and 2 in Fig. 6. In practice, the number of such vertices is so small (about 2%), hence the effect of eliminating them is negligible.

5. Results and discussion

In this section we will try to give answers to three questions by examining practical systems: (1) If the accuracy of linear interpolation for intersection points could be remarkably improved by the bisection algorithm? (2) If the redundant vertices elimination method proposed above is reliable and efficient? (3) How large of grid spacing is the best choice?

The wavefunctions of the molecules used in following analyses were generated by Gaussian 03 program [28] at B3PW91/6-31(d,p) level [29–31]. It was shown that this level is quite satisfactory for computing ESP and $\bar{I}$ [2,14]. Molecular geometries were optimized at the same level. All test data were produced by Multiwfn 2.3.2, of which the parallel mode was enabled. All calculations were done with single QuadCore Intel i7-2630QM CPU.

If not stated explicitly in the text, three times of bisection iterations were performed before linear interpolation during generating intersection points. Grid spacing is set to 0.25 and 0.20 bohr respectively for analyses of ESP and of $\bar{I}$ on molecular surface. The time costs referred below are for analyses of ESP. Relative to the ESP

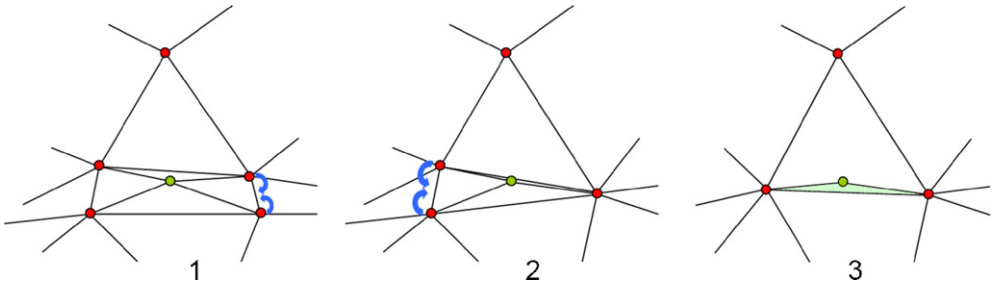


Fig. 6. Vertices merging process. Blue arrows represent merging direction. (1) Two vertices on the right hand side of the pattern 1 are merged into one. (2) The two vertices one the left hand side of the pattern 2 are merged into one. (3) The green vertex and the light green facet in pattern 3 should be finally eliminated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

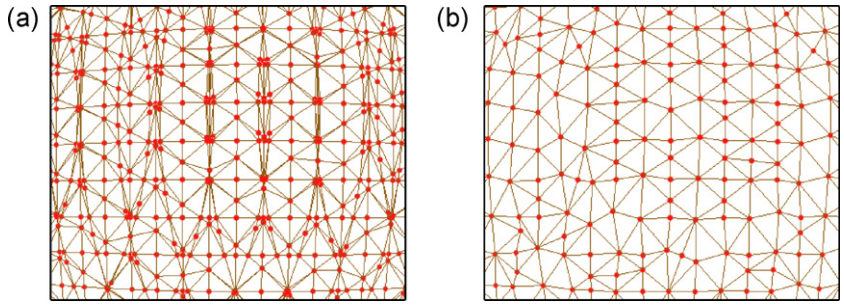


Fig. 7. Surface vertices of a local region of NH_2COH before elimination (a) and after elimination (b) of redundant vertices.

time, the time consumed in analyses of $\bar{I}$ can always be ignored and will not be concerned.

5.1. Interpolation accuracy of intersection points

As mentioned earlier, the accuracy of simple linear interpolation in MT process is not high enough to be used to evaluate the positions of surface vertices. We thus propose to perform a few bisection iterations first to shrink the interval between the two endpoints before linear interpolation. The linear interpolation accuracy is then significantly improved and the generated isodensity surface is much closer to the exact one. We selected two representative systems to illustrate this point, see Table 1, in which RMSD (ρ) is defined to measure overall interpolation accuracy [13]:

$$\mathrm{RMSD}\,(\rho) = \sqrt{\sum_{i}^{N} \frac{(\rho_i - \rho_{\mathrm{iso}})^2}{N}}$$

where N is total number of surface vertices, ρ_i is the electron density at vertex i, ρ_{iso} equals to 0.001 in present study. MaxDev (ρ) will denote the maximum deviation of electron density at surface vertices to ρ_{iso}.

From Table 1 one can see that all listed quantities (except for time spent) gradually converged with the increase of the number of bisection iterations. Directly performing linear interpolation overestimated volume and surface area, and underestimated σ^2_{tot}. Further, the inaccuracy of the direct linear interpolation would result in occurrence of some artificial surface extrema. Through numerical experiments, we found that two bisection iterations are enough to avoid any artificial extrema and three iterations are enough for obtaining accurate surface vertices. Though performing bisection iterations needs some additional computational cost, yet the total time spent is still negligible. The times for adenine and serine are only 1.64 and 1.25 s, respectively.

We noted that in Ref. [13], the authors showed that cubic interpolation reduced RMSD (ρ) one order of magnitude relative to

Table 1
The relationship between the number of bisection iterations performed before linear interpolation and interpolation accuracy, number of found surface extrema of ESP, some molecular properties as well as wall clock time spent.[a]

	Iter.	RMSD (ρ)	MaxDev (ρ)	N_{min}	N_{max}	A_S	Volume	σ^2_{tot}	Time
Adenine ($H_5C_5N_5$)	0	3.93E−05	1.43E−04	5	6	156.968	150.159	531.556	0.1
	1	7.41E−06	3.37E−05	3	4	155.833	149.570	536.251	0.5
	2	1.97E−06	1.01E−05	3	3	155.716	149.444	537.294	1.0
	3	5.51E−07	4.97E−06	3	3	155.691	149.413	537.553	1.6
	4	4.21E−07	4.79E−06	3	3	155.685	149.405	537.625	2.0
	5	4.38E−07	4.74E−06	3	3	155.683	149.404	537.659	2.4
Serine ($H_7C_3NO_3$)	0	4.60E−05	1.75E−04	4	6	137.338	123.989	217.889	0.1
	1	8.60E−06	3.57E−05	4	4	136.114	123.445	221.051	0.5
	2	2.33E−06	1.17E−05	4	4	135.999	123.334	221.669	0.9
	3	6.62E−07	6.18E−06	4	4	135.973	123.307	221.842	1.2
	4	5.08E−07	6.15E−06	4	4	135.967	123.300	221.881	1.4
	5	5.27E−07	6.13E−06	4	4	135.966	123.298	221.892	1.7

[a] Iter. is the number of bisection iterations performed before doing linear interpolation, 0 means only doing linear interpolation; N_{min} and N_{max} are the number of found surface minima and maxima of ESP respectively; A_S is in Å^2; vdW volume is in Å^3; $\overline{V}_S$ is in kcal/mol (1 kcal/mol = 4186 J/mol). The wall clock time consumed in MT stage is in s.

Table 2
Number of surface vertices, consumed wall clock time and various surface descriptors calculated before and after redundant vertices elimination for six representative molecules.[a]

	N_{vtx}	Time	A_S^+	A_S^-	$\overline{V}_S^+$	$\overline{V}_S^-$	σ_+^2	σ_-^2	Π	$V_{S,min}$	$V_{S,max}$
Adenine ($H_5C_5N_5$)											
Before	39,930	240	70.28	85.45	24.93	−19.95	335.22	202.66	22.23	−50.41	66.59
After	15,154	93	70.25	85.44	24.93	−19.95	335.00	202.56	22.23	−50.41	66.59
Relat. diff.	−62.0%	−61.3%	−0.04%	−0.01%	−0.01%	−0.02%	−0.07%	−0.05%	−0.02%	−0.01%	0.00%
DMF (H_7C_3NO)											
Before	30,526	73	84.49	35.30	11.89	−22.15	25.10	218.04	14.20	−43.30	21.82
After	11,731	28	84.44	35.32	11.89	−22.13	25.07	218.08	14.20	−43.30	21.83
Relat. diff.	−61.6%	−61.6%	−0.06%	0.04%	0.02%	−0.09%	−0.14%	0.01%	−0.01%	0.01%	0.03%
Dopamine ($H_{11}C_8NO_2$)											
Before	50,758	392	101.50	95.43	13.86	−15.09	132.21	113.83	14.46	−43.03	52.02
After	19,562	153	101.52	95.35	13.84	−15.09	132.11	113.68	14.46	−43.02	52.02
Relat. diff.	−61.5%	−61.0%	0.03%	−0.09%	−0.09%	0.02%	−0.07%	−0.13%	−0.04%	−0.02%	0.00%
NH_2COH											
Before	19,660	16	43.78	33.84	19.91	−23.22	169.01	198.21	21.23	−42.32	48.63
After	7675	6	43.79	33.80	19.89	−23.24	169.06	197.87	21.23	−42.31	48.63
Relat. diff.	−61.0%	−62.5%	0.04%	−0.12%	−0.09%	0.07%	0.03%	−0.17%	−0.02%	−0.02%	0.00%
Serine ($H_7C_3NO_3$)											
Before	34,630	128	68.00	68.02	17.55	−18.00	129.48	92.56	17.78	−37.32	49.03
After	13,440	51	67.97	68.00	17.54	−17.99	129.32	92.52	17.77	−37.31	49.03
Relat. diff.	−61.2%	−60.2%	−0.05%	−0.02%	−0.03%	−0.06%	−0.12%	−0.04%	−0.05%	−0.05%	0.00%
Naphthalene (H_8C_{10})											
Before	43,660	305	89.03	82.61	8.04	−9.77	15.12	26.45	8.90	−16.92	14.32
After	16,665	112	89.12	82.49	8.03	−9.78	15.17	26.35	8.90	−16.92	14.31
Relat. diff.	−61.8%	−63.3%	0.10%	−0.15%	−0.17%	0.10%	0.35%	−0.39%	−0.03%	0.00%	−0.01%
MURD[b]	61.5%	61.6%	0.05%	0.07%	0.07%	0.06%	0.13%	0.13%	0.03%	0.02%	0.01%

[a] N_{vtx} is the number of surface vertices; Wall clock time is in s; A_S^+ and A_S^- are in Å^2; $\overline{V}_S^+$, $\overline{V}_S^-$, $V_{S,min}$, $V_{S,max}$ and Π are in kcal/mol; σ_+^2 and σ_-^2 are in (kcal/mol)^2, similarly hereinafter.
[b] Mean unsigned relative deviation.

linear interpolation. While in our test, three bisection iterations reduced this quantity as many as two orders. MaxDev (ρ) was also lowered to the same level. Besides, implementation of bisection algorithm is much easier than cubic interpolation. So we believe that carrying out three bisection iterations before linear interpolation is a better choice than using cubic interpolation in MT stage of quantitative molecular surface analysis.

5.2. The effect and reliability of redundant vertices elimination method

We arbitrarily selected a local region of NH_2COH molecule to exhibit the distribution of surface vertices before and after redundant vertices elimination, see Fig. 7. This is also a typical pattern in other regions and in other systems. It is evident that before the elimination, the distribution is rather uneven; many tiny regions are unnecessarily described by too many aggregated vertices. After elimination, the uniformity has a significant improvement, no cluster of vertices has been found. This visual observation is directly correlated with the frequency count of facet areas shown in Fig. 8. It can be seen from Fig. 8a that the molecular surface model constructed by standard MT approach is dominated by the facets whose areas are almost zero. It is shown in Fig. 8b that these too small facets are eliminated with our algorithm.

The number of surface vertices, the wall clock time consumed and various surface descriptors calculated before and after redundant vertices elimination are listed in Table 2. Six different molecules are selected as a test set. As pointed out earlier, evaluation of ESP at surface vertices is a bottleneck of quantitative molecular surface analysis, our elimination method cuts down 61.5% of surface vertices in average and saves as high as 61.6% of the total computational time. The data in Table 2 show that the time saving is not at the expense of sacrificing the accuracy of the results, the maximal differences between the surface descriptors calculated before and after vertices elimination are less than four

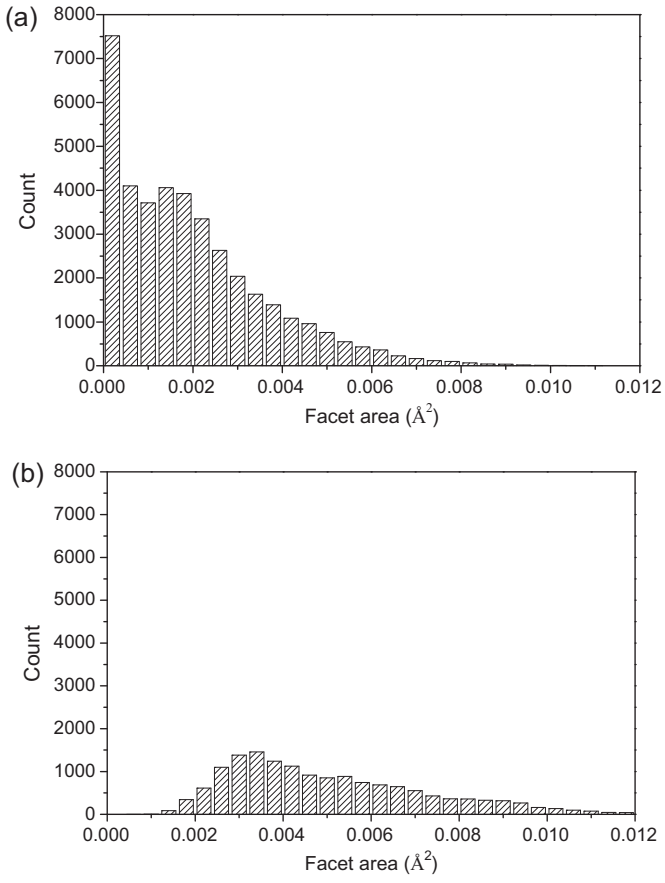


Fig. 8. Frequency count of areas of surface facets. (a) Before redundant vertices elimination. (b) After redundant vertices elimination.

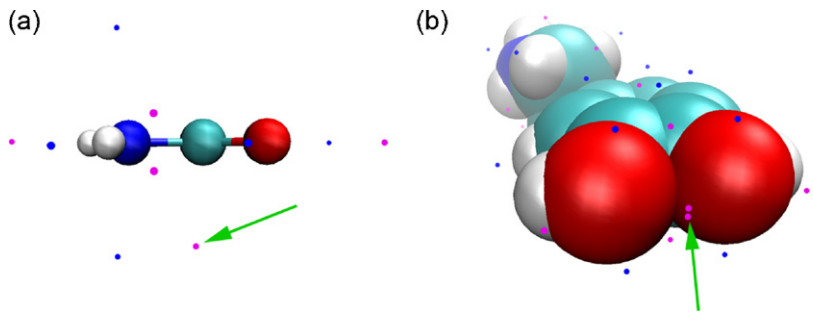


Fig. 9. Surface extrema of $\bar{I}$ of NH_2COH (a) and dopamine (b). Pink and blue points correspond to surface maxima and minima respectively. The maximum pointed by arrow in (a) has the lowest value compared to other maxima. The maximum pointed by arrow in (b) is very close to another one. If vertices elimination is employed or further decreasing grid spacing to refine the result, the two artificial maxima will disappear. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

thousandth. For most cases the differences are even lower than one thousandth, and thereby can be completely ignored.

We also examined the effect of the vertices elimination on locating surface extrema of ESP. For the six molecules involved in Table 2, the numbers of both surface minima and maxima found before and after the elimination are identical. The maximum displacement of extrema perturbed by the elimination method is 0.069 Å, such small displacement nearly cannot be detected through visual inspection. The maximal relative change of ESP value at surface extrema due to the elimination is merely 0.05%.

Although the time cost in evaluating $\bar{I}$ is trivial and hence it is not necessary to eliminate redundant vertices to save time for surface analyses of $\bar{I}$, it is still of interest to investigate the impact on its surface extrema due to the vertices elimination. For the six molecules in Table 2, the maximum relative change of $\bar{I}$ value at global surface minima and maxima are only 0.06% and 0.03% respectively. The positions of most surface extrema are basically unperturbed by vertices elimination. However, this is not always true for all states. We found a few surface maxima located by 2-neighborhood method are artificial. If surface vertices are directly generated by MT approach, these maxima often either have very low $\bar{I}$ value or are very close to other maxima, see Fig. 9 for example. The reason for the artificial maxima may be that 2-neighborhood method is not very suitable for the case when surface vertices aggregated severely, especially for $\bar{I}$, whose variance on vdW surface is often more complex than that of ESP. Fortunately, after implementing our vertices elimination method to regularized the distribution of vertices, these artificial maxima disappeared. This is a surprising byproduct of our algorithm.

5.3. Grid spacing dependency of quantitative molecular surface analysis

Grid spacing of volume data of electron density directly determines the number of generated surface vertices and hence the accuracy of surface analysis. Thus, finding out the optimal spacing is of great significance. We calculated some surface properties for dopamine and naphthalene and listed the result in Table 3. From the table it can be seen that the proportions of eliminated vertices by our method are about 62%. This ratio is almost unchanged with grid spacing, because the distance criteria we used is a relative quantity (0.4 times of grid spacing) rather than an absolute number. Since a smaller grid spacing corresponds to shorter length of active edges and thus better interpolation accuracy for surface vertices, RMSD (ρ) becomes small with decrease of grid spacing.

As expected, all listed surface descriptors listed in Table 3 gradually converge with decreasing grid spacing. We also found that the volume, $\overline{V}_S$, σ^2_{tot} and Π do not depend on grid spacing so strongly. For example, the deviation of the volume, $\overline{V}_S$, σ^2_{tot} and Π calculated with grid spacing of 0.6 bohr with respect to the ones with

Table 3
Grid spacing dependency of some surface properties of dopamine and naphthalene.

Spacing[a]	N_{vtx}	N_{elim}[b]	RMSD (ρ)	Volume	A_S	$\overline{V}_S$	σ^2_{tot}	Π	$V_{S,\min}$	$V_{S,\max}$	N_{ESP}[c]	$N_{\bar{I}}$[d]
Dopamine ($H_{11}C_8NO_2$)												
1.00	1226	60.6%	9.70E−06	194.03	195.42	−0.182	235.46	14.21	−42.72	51.72	5/5	18/17
0.75	2148	61.4%	5.33E−06	194.40	196.10	−0.189	240.08	14.32	−42.75	51.81	4/5	19/18
0.60	3332	62.1%	3.34E−06	194.39	196.39	−0.175	242.79	14.38	−42.48	51.98	4/6	19/18
0.45	5935	62.0%	1.90E−06	194.70	196.66	−0.174	244.20	14.42	−42.89	51.87	4/6	21/18
0.35	9954	61.5%	1.18E−06	194.91	196.78	−0.173	245.34	14.44	−42.92	52.06	5/6	21/19
0.25	19,562	61.5%	5.85E−07	195.08	196.88	−0.172	245.79	14.46	−43.02	52.02	5/6	21/18
0.20	30,146	62.0%	3.80E−07	195.17	196.91	−0.171	246.08	14.46	−43.00	52.01	5/6	21/18
0.15	53,943	61.8%	2.13E−07	195.25	196.94	−0.171	246.28	14.47	−43.03	52.02	5/6	21/18
0.10	122,286	61.5%	9.45E−08	195.33	196.96	−0.171	246.40	14.47	−43.03	52.03	5/6	21/18
Naphthalene (H_8C_{10})												
1.00	974	64.9%	7.48E−06	170.91	170.53	−0.55	40.71	8.75	−16.90	14.35	4/8	15/14
0.75	1826	62.4%	3.95E−06	171.59	171.01	−0.55	40.85	8.82	−16.92	14.34	6/8	16/15
0.60	2755	64.0%	2.72E−06	171.81	171.29	−0.54	41.13	8.85	−16.91	14.33	8/10	16/14
0.45	5079	62.6%	1.49E−06	172.13	171.45	−0.53	41.43	8.87	−16.91	14.32	8/12	16/14
0.35	8143	63.7%	9.00E−07	172.34	171.54	−0.53	41.53	8.89	−16.92	14.31	10/12	16/14
0.25	16,665	61.8%	4.61E−07	172.50	171.61	−0.53	41.52	8.90	−16.92	14.31	10/12	20/14
0.20	26,123	61.8%	2.93E−07	172.58	171.63	−0.53	41.61	8.90	−16.92	14.31	10/12	20/14
0.15	46,657	61.6%	1.65E−07	172.66	171.65	−0.53	41.59	8.90	−16.92	14.32	10/12	20/14
0.10	105,158	61.5%	7.35E−08	172.74	171.67	−0.53	41.62	8.90	−16.92	14.32	10/12	20/14

[a] Grid spacing is in bohr.
[b] The proportion of eliminated vertices in all vertices produced by standard MT approach.
[c] The number of surface minima and maxima of ESP, separated by slash.
[d] The number of surface minima and maxima of $\bar{I}$, separated by slash.

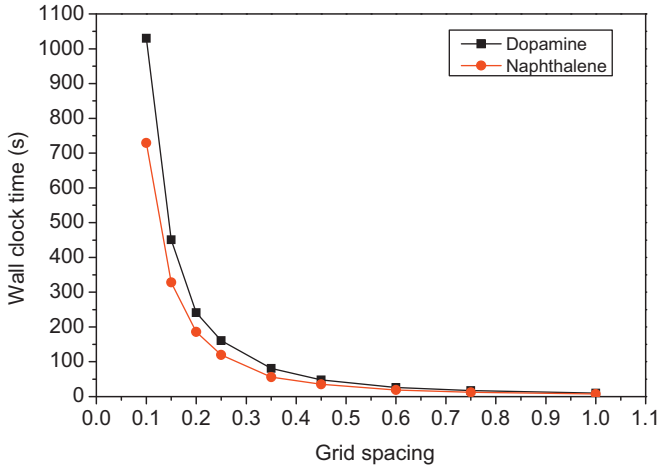


Fig. 10. Variance of wall clock time spent with grid spacing.

spacing of 0.1 bohr for dopamine are just −0.48%, −0.29%, 2.5% and −1.47%, respectively. The results produced with 1.0 bohr spacing may be still usable for qualitative analysis. However, analyses of surface extrema have obviously stringent requirement on grid spacing. From Table 3 we can see that in order to locate all surface extrema of ESP ($\bar{I}$), grid spacing of at least 0.35 (0.25) bohr must be used. Since the requirement of grid spacing is different from system to system, on the consideration of safety, the grid spacing suggested to be used is 0.25 bohr for ESP and 0.20 bohr for $\bar{I}$.

The times spent on calculating the surface properties for dopamine and naphthalene with respect to the grid spacing are plotted in Fig. 10. It can be found in the figure that there is a notable turning point around grid spacing of 0.25 bohr. The time cost will increase sharply if the spacing is further lowered from this point. Therefore, the grid spacing of 0.25 bohr is also optimal choice from the computational cost point of view.

6. Conclusions

In this paper, many numerical aspects of quantitative molecular surface analysis have been discussed in detail. Our primary purpose is to present a method to eliminate unreasonably distributed vertices produced by standard MT process to reduce computational time of ESP on molecular surface. It has been demonstrated that our method is capable to reduce as many as about 62% total time spent in molecular surface analysis for ESP. Eliminating these vertices not only almost did not deteriorate the accuracy of the result, but also avoided the presence of some artificial surface extrema of $\bar{I}$. It has also been shown that the accuracy of the direct linear interpolation commonly used in MT process is unsatisfactory for surface analysis and can be improved significantly if three times of bisection iterations are performed before the linear interpolation. The extra time spent on bisection iterations is however negligible. Finally, we explored the optimal value of grid spacing for surface analysis, 0.25 and 0.20 bohr are recommended to be used for analysis of ESP and $\bar{I}$, respectively. Generally this setting gives accurate enough results with reasonable computational cost. However, if one does not need information of surface extrema, the grid spacing can be appropriately enlarged.

Acknowledgements

The authors thank the National Natural Science Foundation of China (Project Nos. 20773011 and 21173020) for the financial support.

References

[1] J.S. Murray, P. Politzer, Electrostatic potentials: chemical applications, in: P.v.R. Schleyer (Ed.), Encyclopedia of Computational Chemistry, vol. 2, John Wiley & Sons, West Sussex, 1998, pp. 912–920.
[2] J.S. Murray, P. Politzer, The electrostatic potential: an overview, WIREs Computational Molecular Science 1 (2011) 153–163.
[3] P. Politzer, P.R. Laurence, K. Jayasuriya, Molecular electrostatic potentials: an effective tool for the elucidation of biochemical phenomena, Environmental Health Perspectives 61 (1985) 191–202.
[4] P. Politzer, J.S. Murray, Molecular electrostatic potentials and chemical reactivity, in: K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 2, John Wiley & Sons, Inc., New York, 1991, pp. 273–312.
[5] R.F.W. Bader, M.T. Carroll, J.R. Cheeseman, C. Chang, Properties of atoms in molecules: atomic volumes, Journal of the American Chemical Society 109 (1987) 7968–7979.
[6] P. Politzer, J.S. Murray, The fundamental nature and role of the electrostatic potential in atoms and molecules, Theoretical Chemistry Accounts 108 (2002) 134–142.
[7] P. Politzer, J.S. Murray, P. Lane, σ-Hole bonding and hydrogen bonding: competitive interactions, International Journal of Quantum Chemistry 107 (2007) 3046–3052.
[8] P. Politzer, P. Lane, M. Concha, Y. Ma, J. Murray, An overview of halogen bonding, Journal of Molecular Modeling 13 (2007) 305–311.
[9] T. Clark, M. Hennemann, J. Murray, P. Politzer, Halogen bonding: the σ-hole, Journal of Molecular Modeling 13 (2007) 291–296.
[10] J. Murray, P. Lane, T. Clark, K. Riley, P. Politzer, σ-Holes, π-holes and electrostatically-driven interactions, Journal of Molecular Modeling 18 (2012) 541–548.
[11] J.S. Murray, T. Brinck, P. Lane, K. Paulsen, P. Politzer, Statistically-based interaction indices derived from molecular surface electrostatic potentials: a general interaction properties function (GIPF), Journal of Molecular Structure (THEOCHEM) 307 (1994) 55–64.
[12] P. Sjoberg, J.S. Murray, T. Brinck, P. Politzer, Average local ionization energies on the molecular surfaces of aromatic systems as guides to chemical reactivity, Canadian Journal of Chemistry 68 (1990) 1440–1443.
[13] F. Bulat, A. Toro-Labbé, T. Brinck, J. Murray, P. Politzer, Quantitative analysis of molecular surfaces: areas volumes, electrostatic potentials and average local ionization energies, Journal of Molecular Modeling 16 (2010) 1679–1691.
[14] P. Politzer, J. Murray, F. Bulat, Average local ionization energy: a review, Journal of Molecular Modeling 16 (2010) 1731–1742.
[15] T. Lu, F. Chen, Multiwfn: A multifunctional wavefunction analyzer, Journal of Computational Chemistry 33 (2012) 580–592.
[16] http://Multiwfn.codeplex.com (accessed 22.03.12).
[17] W.E. Lorensen, H.E. Cline, Marching cubes: a high resolution 3D surface construction algorithm, in: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, ACM, 1987, pp. 163–169.
[18] D.A. Rajon, W.E. Bolch, Marching cube algorithm: review and trilinear interpolation adaptation for image-based dosimetric models, Computerized Medical Imaging and Graphics 27 (2003) 411–435.
[19] T.T. Elvins, A survey of algorithms for volume visualization, ACM SIGGRAPH Computer Graphics 26 (1992) 194–201.
[20] T.S. Newman, H. Yi, A survey of the marching cubes algorithm, Computers & Graphics 30 (2006) 854–879.
[21] L. Thomas, L. Hélio, V. Antônio Wílson, T. Geovan, Efficient implementation of Marching Cubes' cases with topological guarantees, Journal of Graphics Tools 8 (2003) 1–16.
[22] S.L. Chan, E.O. Purisima, A new tetrahedral tesselation scheme for isosurface generation, Computers & Graphics 22 (1998) 83–90.
[23] S.L. Chan, E.O. Purisima, Molecular surface generation using marching tetrahedra, Journal of Computational Chemistry 19 (1998) 1268–1277.
[24] G.M. Treece, R.W. Prager, A.H. Gee, Regularised marching tetrahedra: improved iso-surface extraction, Computers & Graphics 23 (1999) 583–598.
[25] W.J. Schroeder, J.A. Zarge, W.E. Lorensen, Decimation of triangle meshes, in: Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques, ACM, 1992, pp. 65–70.
[26] P. Cignoni, C. Montani, R. Scopigno, A comparison of mesh simplification algorithms, Computers & Graphics 22 (1998) 37–54.
[27] K. Low, T. Tan, Model simplification using vertex-clustering, in: Proceedings of the 1997 Symposium on Interactive 3D Graphics, ACM, Providence, Rhode Island, United States, 1997, pp. 75–81.
[28] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vrevon, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Menucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, C. Adamo, J. Jaramillo, R. Gomparts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.-Y. Peng, A. Namayakkara, M. Challacombe, P.M.W. Gill, B. Johnson,

W. Chen, M.-W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, E.01, Gaussian, Inc., Wallingford, CT, 2004.
[29] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, Journal of Chemical Physics 98 (1993) 5648–5652.
[30] W.J. Hehre, R. Ditchfield, J.A. Pople, Self – consistent molecular orbital methods. XII. Further extensions of Gaussian – type basis sets for use in molecular orbital studies of organic molecules, Journal of Chemical Physics 56 (1972) 2257–2261.
[31] P.C. Hariharan, J.A. Pople, The influence of polarization functions on molecular orbital hydrogenation energies, Theoretical Chemistry Accounts 28 (1973) 213–222.