Transferable Atom Equivalent Multicentered Multipole Expansion Method

C. E. WHITEHEAD,[1,*] C. M. BRENEMAN,[1] N. SUKUMAR,[1] M. D. RYAN[2]
[1]*Department of Chemistry, Rensselaer Polytechnic Institute, 110 8th St., Troy, NY 12180*
[2]*Millennium Pharmaceuticals, 75 Sidney Street, Cambridge, MA 02139*

Received 10 June 2002 ; Accepted 17 September 2002
Published online 12 February 2003 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/jcc.10240

Abstract: The transferability of atomic and functional group properties is an implicit concept in chemistry. The work presented here describes the use of Transferable Atom Equivalents (TAE) to represent molecular electrostatic potential fields through the use of integrated atomic multipole moments that are associated with each TAE atom type used in the reconstruction. TAE molecular surface distributions of electrostatic potentials are compared with analytical *ab initio* and empirical (Gasteiger) partial charge reference models for several conformations of test peptides. Surface electrostatic potential distributions computed using TAE multipole representations were found to converge at the octopole level, with incremental improvement observed when hexadecapoles were included. Molecular electrostatic potential fields that were produced using the TAE method were observed to be responsive to conformational changes and to compare well with *ab initio* reference distributions. Generation of TAE atom types and their associated multipoles does not involve fitting to sample electrostatic potential fields, but rather utilizes integrated AIM atomic electron density distributions within representative chemical environments. The RECON program was used for TAE reconstruction. RECON is capable of processing 5000 drug-sized molecules or 25 proteins per minute per 1.7 GHz P4 Linux processor.

 J Comput Chem 24: 512–529, 2003

Key words: TAE; atoms in molecules; molecular electrostatic potential; multipole moments

Introduction

Analysis and visualization of molecular electrostatic potentials (MEPs) have found widespread use in chemistry. In addition to dominating short-range intermolecular potentials, MEPs are known to be important for describing the long-range intermolecular interactions involved in the initial steps of molecular recognition. Various implementations of MEPs are used extensively in molecular mechanics, field similarity searches, and in generating QSAR/QSPR descriptors.[1–4] In earlier work, Politzer et al. pioneered the use of MEPs evaluated on molecular van der Waals surfaces to define a set of General Interaction Property Function (GIPF) descriptors that have proven useful for modeling intermolecular interactions in the condensed phase, and within biological systems.[5–7] In that work, it was found that simple correlations exist between certain of the GIPF descriptors and observed hydrogen bond acceptor/donor strengths and molecular dipole moments.[1,6] In keeping with Politzer's studies, one of the authors has used electronic properties calculated on molecular van der Waals surfaces as QSAR/QSPR descriptors to address a number of problems. These descriptors include surface histogram characterizations of electron density-derived properties calculated on the 0.002 electron/au^3 isodensity surface.[8,9] Initial use of surface histograms in combination with machine learning methods have allowed the generation of robust QSAR/QSPR models for a variety of observables, and have shown promise as the descriptors of choice for modeling ADME properties. It can be argued that descriptors based on electronic properties on the molecular van der Waals surface provide chemical information complementary to polar surface area descriptors alone.[10,11] Prior to the introduction of Transferable Atom Equivalent (TAE) methods, the use of electron density-derived properties in QSAR/QSPR studies has been limited as a result of the time required for their calculation. Although modern software and hardware permit the computation of many drug-sized molecules per minute using traditional techniques, effective use of these methods for peptide and protein-sized molecules has not yet proven to be practical. In addition to this limitation, MEPs have been shown to be sensitive to the method from which they are derived. Most applications of MEPs in QSAR/QSPR and the electrostatic models available for molecular me-

Current address: Pfizer Global Research & Development, Ann Arbor Laboratories, Pfizer Inc., 2800 Plymouth Road Building 28 G062W, Ann Arbor, MI 48105; Phone: (734) 622-2636; Fax: (734) 622-2782; e-mail: christopher.whitehead@pfizer.com

Correspondence to: C. M. Breneman; e-mail: brenec@rpi.edu

chanics/dynamics have been limited to atomic partial charge models, distributed multipoles from molecular orbital partitioning, or MEP-derived multipolar representations determined through least-squares fitting of *ab initio* MEP results on small molecules.

To appreciate the hierarchy of methods used for representing molecular electrostatic potential fields, it is helpful to review the basic features of the MEP and the approximations that have been used to represent it. The analytical expression for EP is defined in eq. (1):

$$EP(r) = \sum_{i=1}^{N} \frac{Z}{|\vec{R}_i - \vec{r}|} - \int \frac{\rho(\vec{r}')}{|\vec{r}' - \vec{r}|}\, dr' \tag{1}$$

The first term of the expression represents the repulsive potential arising from interaction of a positive test charge with the nuclear charges, and the second term arises from the attraction of the test charge to the electrons in the system. The accuracy of MEP calculations is dependent on the quality of the electron density distribution used in the calculation; so only for the smallest systems can highly accurate electron densities and electrostatic potentials be calculated. As a result, significant effort has been expended in the development of methods to approximate MEPs, particularly for larger systems. The computational chemistry community has been searching for methods that are accurate, polarizable, conformationally responsive, computationally inexpensive and transferable. Several simplifications that have been employed are reviewed below.

Due to its conceptual simplicity and ease of manipulation, the isotropic point charge model is probably the most widely used method to estimate MEPs. The quantum mechanical EP can be approximated by the classical point charge model given in eq. (2).

$$EP(r) = \sum_{i} \frac{q_i}{|R_i - \vec{r}|} \tag{2}$$

The variable q_i is the charge of the ith atom, and $|R_i - \vec{r}|$ is the distance from each atom to the point of interest. Two of the historical point charge approximations are derived from Mulliken orbital populations and MEP-derived charges. Mulliken population analysis (MPA) charges are obtained by partitioning the atomic overlap density between two atoms equally and are known to be very sensitive to basis set selection. MPA charges have been shown to poorly represent the MEP and do not reproduce molecular dipole moments.[12] On the other hand, MEP-derived atomic partial charges arise from a least-squares fit of the analytical electrostatic potential fields of representative small molecules.[13] In most cases, potential-derived point charges have been shown to provide a better representation of the MEP than provided by MPA charges. Even when not constrained to do so, the MEP charges are known to closely approximate the molecular dipole moment of the conformation used in the fit. Unfortunately, MEP charges are sensitive to sampling errors in their derivation and more seriously, are not inherently transferable from the small molecules used in their generation to other molecular environments.[14,15] Other groups have attempted to derive conformationally responsive and transferable atomic charges by using conformational analysis,[16] introducing constraints into the fitting procedure,[17] as well inclusion of charge-dipole polarization.[18] Even when conformationally sensitive[19] and transferable MEP charge models are used, the resulting atomic partial charges are not able to fully recover the anisotropic character or topologic properties of MEPs on real molecules.[20,21] On the other hand, multicentered multipole expansion methods can capture the anisotropic nature of MEPs by virtue of their ability to represent nonspherical charge density distributions.[22–27] In another work, Williams represented molecular electrostatic potentials through the use of atomic multipole and bond dipole models derived from an analytical electrostatic potential distribution. As with many methods requiring least-squares fitting of a complex function, artifacts were observed in this procedure that limited the transferability and accuracy of the charge model.[28]

More examples of multicentered multipole expansions include Sokalski's cumulative atomic multipole moments (CAMM) method,[23,29] Stone's distributed multipole analysis (DMA), and more recently, an Atoms in Molecules-based multipole moment method.[24,30] For CAMM, the atom-centered multipole moments are recursively constructed using the density matrix starting from any definition of the atomic charges. The multipole moments are built in such a way as to better describe the anisotropic nature of the MEP,[31] while DMA distributes multipole moments on atomic and bond centers. A method similar to DMA, described by Vigne-Maeder and Claverie, gives the user the option of selecting multipole moments at atomic, bond, fictitous short, and long bonds.[32] Although it is possible to obtain the exact multicenter multipolar part of the molecular charge distribution, a balance must be maintained between the accuracy of the calculated EP and the number of multipole moment sites. All three methods by Vigne-Maeder, Stone, and Sokalski decrease the convergence requirements of multipole moment expansion and have found use as methods for calculating inter- and intramolecular electrostatic interactions and as potentially transferable units for calculating MEPs for amino acid units.[26,33–35] Although CAMM and DMA-like methods are useful schemes for representing multipole moments (and by extension MEPs), multipole moments derived from Bader's theory of atoms in molecule (AIM) are well suited to calculate MEPs due to having their "roots" in quantum mechanics.[36–37] In a study by Bader and Laidig, the general expressions of the AIM atomic multipole moments were described and used to calculate molecular multipole moments. It was revealed that AIM multipole moments, like other properties defined through the application of the hypervirial theorem, are additive and transferable.[37–39] Bader, Laidig, and others have also confirmed the usefulness of the AIM approach for determining and studying the origins of atomic and molecular properties.[37,40–43] Although generally accepted, the AIM method has found limited use because of its high computational overhead. An alternative to the AIM procedure is the Transferable Atom Equivalent (TAE) method.[8–9,53] TAEs are atom-based electron density fragments obtained using the AIM approach, which are then organized into a library of atomic fragments for use in rapid molecular charge density reconstruction. The Transferable Atom Equivalent (TAE) method was developed to provide near *ab initio* quality electron densities, energies, and properties at minimal computational cost. It will be shown in this work that the TAE multipoles derived from the AIM approach can be used to approximate MEP distributions on molecular van der

Waals surfaces that are semiquantitative and conformationally responsive.

In Section 2, we present the theory of multicentered multipole expansion and discuss the CAMM and DMA approaches further. The AIM approach is presented in Section 3 and the TAE/RECON method in Section 4. Computational details are described in Section 5. Results of EP calculated from TAE multipole analysis (TAE MA EP) are shown in Section 6 and compared to analytical MEP and EP obtained from Gasteiger charges. We conclude with a brief summary in Section 7.

Theory of Multicentered Multipole Expansion

To appreciate the capability of multipolar representations of charge densities to be used for electrostatic potential reconstruction, it is useful to cover some of the mathematics involved in the process.[44–47]

The interaction energy of a charge distribution with a field, is given by eq. (3),

$$u = \int \phi \rho d\upsilon \tag{3}$$

where ϕ is the field, and ρ is the electron density in volume element $d\upsilon$. The expression is integrated over the extent of the charge distribution. The field ϕ is a function of Cartesian coordinates and can be expressed as a Taylor expansion about zero as shown in eq. (4).

$$\phi = \phi(0) + \boldsymbol{r}_a\phi'_a(0) + \frac{1}{2!}\boldsymbol{r}_a\boldsymbol{r}_b\phi''_{ab}(0) + \frac{1}{3!}\boldsymbol{r}_a\boldsymbol{r}_b\boldsymbol{r}_c\phi'''_{abc}(0) + \ldots \tag{4}$$

The primes denote partial derivatives with respect to components a, b, c, and d, which are totally symmetric in all indices. Substituting the potential into the interaction energy equations yields the interaction energy with respect to moments and fields through the fourth order as shown in eq. (5). The moments and fields are defined in eq. (6).

$$u = q\phi(0) - \mu_a F_a(0) + \frac{1}{2!}Q_{ab}F'_{ab}(0) - \frac{1}{3!}R_{abc}F''_{abc}(0) + \frac{1}{4!}S_{abcd}F'''_{abcd}(0) \tag{5}$$

$$q = \int \rho d\upsilon$$

$$\mu_a = \int r_a \rho d\upsilon$$

$$Q_{ab} = \int r_a r_b \rho d\upsilon$$

$$R_{abc} = \int r_a r_b r_c \rho d\upsilon$$

$$S_{abcd} = \int r_a r_b r_c r_d \rho d\upsilon$$

$$F = \phi$$

$$F_a = -(\partial\phi/\partial r_a)$$

$$F'_{ab} = -(\partial^2\phi/\partial r_a \partial r_b)$$

$$F''_{abc} = -(\partial^3\phi/\partial r_a \partial r_b \partial r_c)$$

$$F'''_{abcd} = -(\partial^4\phi/\partial r_a \partial r_b \partial r_c \partial r_d)$$

$$\sum_a F'_{aa} = \sum_b F''_{abb} = \sum_c F'''_{abcc} = 0 \tag{6}$$

According to the Laplace equation, the fields within the charge distribution contract to zero. The corresponding multipole moments are then obtained by taking successive partial derivatives of $1/r$ with respect to coordinates x, y, and z. The equations for the dipole (μ), quadrupole (Θ), octupole (Ω), and hexadecapole (Φ) moments are defined in eq. (7).

$$\xi^n_{\alpha\beta\gamma\ldots\eta} = (-1)^n \frac{1}{n!}\int \rho d\upsilon r^{2n+1} \frac{\partial}{\partial r_\alpha}\frac{\partial}{\partial r_\beta}\frac{\partial}{\partial r_\gamma}\cdots\frac{\partial}{\partial r_\eta}\left(\frac{1}{r}\right)$$

$$q = \int \rho d\upsilon$$

$$\mu_\alpha = \int r_\alpha \rho d\upsilon$$

$$\Theta_{\alpha\beta} = \frac{1}{2}\int \rho d\upsilon (3r_\alpha r_\beta - r^2\delta_{\alpha\beta})$$

$$\Omega_{\alpha\beta\gamma} = \frac{1}{2}\int \rho d\upsilon (5r_\alpha r_\beta r_\gamma - r^2(r_\alpha\delta_{\beta\gamma} + r_b\delta_{\alpha\gamma} + r_\gamma\delta_{\alpha\beta}))$$

$$\Phi_{\alpha\beta\gamma\varepsilon} = \frac{1}{8}\int \rho d\upsilon (35r_\alpha r_\beta r_\gamma r_\varepsilon - 5r^2 r_\alpha r_\beta \delta_{\gamma\varepsilon} + r^4\delta_{\alpha\beta}\delta_{\gamma\varepsilon}) \tag{7}$$

Like the moments and the fields, the multipole moments are totally symmetric with respect to exchange of indices. In general, each of the multipole moments will have $2n + 1$ independent components, where n gives the rank of the multipole moment. For a molecule without symmetry there will be three dipole, five quadrupole, seven octupole, and nine independent hexadecapole moment components, respectively. A spherically symmetric molecular charge distribution will have vanishing higher order multipole moments, in contrast to the moments defined in eq. (6), which are nonvanishing. In this regard, higher order multipole moments can be seen as useful for describing the deviation from a

spherical charge distribution. Like the field values defined in eq. (6), the quadrupole, octopole, and hexadecapole moments, as shown in eq. (8), all contract to zero.

$$\sum_a \Theta_{aa} = \sum_b \Omega_{abb} = \sum_c \Phi_{abcc} = 0 \tag{8}$$

Although useful in describing the distribution of charge in a molecule, a multipole moment expansion about a single origin only converges at a distance further than the farthest point (radius of convergence) within the charge distribution, and converges too slowly for points that are not significantly larger than the radius of convergence.[24] Multicentered multipole moment expansion methods were borne out of the need to improve the convergence properties of single site multipole expansions. A number of multicentered multipole expansion methods have been described in the literature, including CAMM,[48] DMA,[24] and potential-fitted atomic and bond multipoles.[28]

Multipoles from the Theory of Atoms in Molecules (AIM)

The theory of Atoms in Molecules (AIM), developed by Bader and coworkers, provides a nonarbitrary scheme for partitioning of the electron density. Atomic multipole moments obtained from this theory have recently been described and applied to model molecular systems.[30,36,38] In the theory of Atoms in Molecules, the topologic features of the electron density are used to define an atom.[49,50] The topology of molecular electron densities can be concisely described by the gradient of the electron density, which has the following important properties: (1) it points in the direction of greatest increase in electron density, (2) gradient paths are noncrossing, and (3) each gradient path must originate or terminate at a critical point, where the gradient of the electron density vanishes:

$$\nabla\rho = 0 \tag{9}$$

Critical points are distinguished by their rank (the number of nonzero curvatures) and signature (the algebraic sum of the signs of the nonzero curvatures). For a nuclear (3,−3) critical point, all of the curvatures are negative and the density is a local maximum at the critical point. A bond critical point occurs at a point along the path of greatest electron density between two atoms, where the density is at a minimum along the bond path and at a maximum in the plane perpendicular to the bond path. Molecular structure can be defined as a collection of nuclear critical points connected by bond paths and their associated critical points. The nuclear critical points are attractors of the electron density, and all gradient paths of the electron density within the basin of an atom terminate at its nucleus. The topological atom is defined as the union of an attractor and its basin—the region of space traversed by the trajectories of the gradient of the electron density that terminate at the attractor. The bond critical point is the attractor of an interatomic surface not belonging to either atom involved in the interaction. The interatomic surface is the only closed surface that satisfies the zero-flux surface condition:

$$\nabla\rho(r) \cdot n(r) = 0 \text{ for all points on the surface } \Omega(r) \tag{10}$$

The interatomic surface separates atoms in molecules into disjoint basins, which are surrounded by zero-flux surfaces in the gradient of the electron density. In Bader's formulation, the zero-flux surface condition is a necessary constraint to define topological atoms as valid quantum subsystems[49] satisfying atomic virial, hypervirial, and Ehrenfest force theorems. Bader and Bedall concluded that:[51]

1. The total energy of a molecule is expressible as a sum of atomic energies.
2. The average potential energy of an atom is defined as the average of the virial of the forces exerted on it.
3. If the distribution of charge for an atom is identical in two different systems, then the atom will contribute identical amounts to the total energy in both systems.

These conclusions also hold for property densities of the electron distribution over the basin. The atomic statement of the hypervirial theorem leads to the definition of atomic property densities, given in eq. (11), as the expectation value of one-electron property density over regions of space surrounded by zero flux surface Ω.

$$A(\Omega) \equiv \langle\hat{A}\rangle_\Omega = \frac{N}{2}\int_\Omega d\tau \int d\tau'\{\Psi^*\hat{A}\Psi + (\hat{A}\Psi)^*\Psi\} \tag{11}$$

The total property A for the molecule, eq. (12), is equal to the sum of atomic contributions of that property.

$$\langle\hat{A}\rangle = \sum_\Omega A(\Omega) \tag{12}$$

Equipped with the atomic statements of the virial and hypervirial theorem, Bader and coworkers were able to show the transferability of chemically equivalent functional groups in molecules.[51,52] Laidig described the general expressions of multipole moments and demonstrated that under certain constraints they are transferable.[38] The traceless AIM multipole moments up to the hexadecapole are given in eq. (13).

$$q^\Omega = Z(\Omega) - \int_\Omega d\tau \int \rho d\tau'$$

$$\mu_\alpha^\Omega = -\int_\Omega d\tau \int r_\alpha \rho d\tau'$$

$$\Theta_{\alpha\beta}^\Omega = -\frac{1}{2}\int_\Omega d\tau \int \rho d\tau'(3r_\alpha r_\beta - r^2\delta_{\alpha\beta})$$

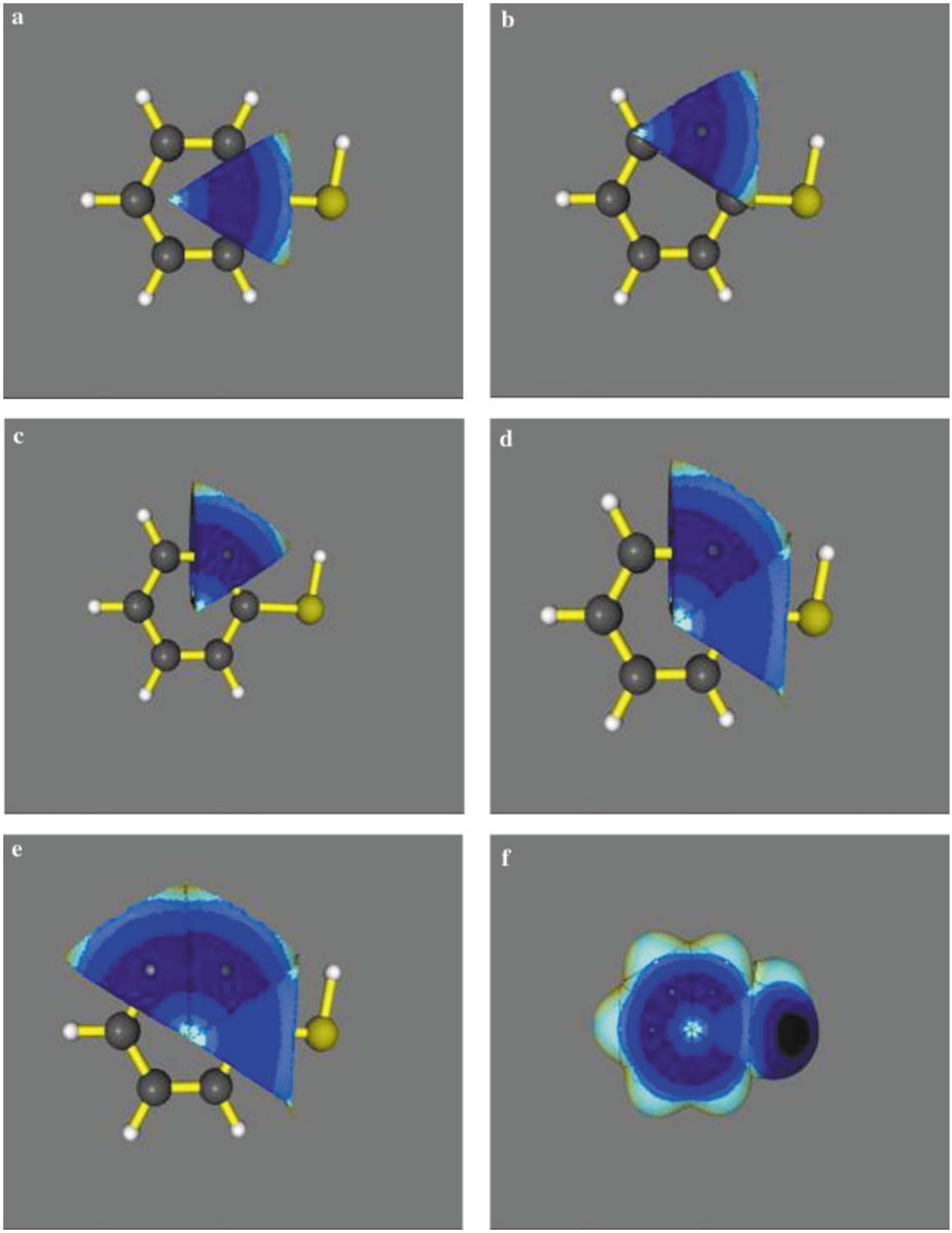

Figure 1. Reconstruction of thiophenol using the new GEOM procedure. (a) The TAE electron distribution in its native position. In the first step, the first TAE electron density fragments are translated to the molecular coordinates of the model atom, as shown in (b). The charge density distribution is then rotated using a quaternion procedure,[46] and the results are illustrated in (c). These steps are repeated for all atoms in the molecular until the entire molecular charge distribution is reconstructed, as shown in (d–f).

Table 1. Phi and Psi Angles that Define the Conformations of the Blocked Peptides (See Figure 2 for Definition).

Molecule	Phi 1	Psi 1	Phi 2	Psi 2	Phi 3	Psi 3
Ala-alpha	−57	−47				
Ala-beta	−84	159				
Ala extended	180	180				
Diglycine beta1	−60	−30	−90	0		
Diglycine beta2	−60	120	80	0		
Diglycine extended	180	180	180	180		
Triglycine alpha	−57	−47	−57	−47	−57	−47
Triglycine extended	180	180	180	180	180	180

$$\Omega^{\Omega}_{\alpha\beta\gamma} = -\frac{1}{2}\int_{\Omega} d\tau \int \rho d\tau' (5r_\alpha r_\beta r_\gamma - r^2(r_\alpha \delta_{\beta\gamma} + r_b \delta_{\alpha\gamma} + r_\gamma \delta_{\alpha\beta}))$$

$$\Phi^{\Omega}_{\alpha\beta\gamma\varepsilon} = -\frac{1}{8}\int_{\Omega} d\tau \int \rho d\tau' (35r_\alpha r_\beta r_\gamma r_\varepsilon - 5r^2 r_\alpha r_\beta \delta_{\gamma\varepsilon} + r^4 \delta_{\alpha\beta}\delta_{\gamma\varepsilon}) \quad (13)$$

The electrostatic potential at a point for multicentered multipole expansion may be then shown to be [eq. (14)]:

$$EP_{mm}(r) = \sum_{\Omega} \left[q^{\Omega} F + \mu^{\Omega}_{\alpha} F_{\alpha} + \frac{1}{3}\Theta^{\Omega}_{\alpha\beta} F'_{\alpha\beta} + \frac{1}{15}\Omega^{\Omega}_{\alpha\beta\gamma} F''_{\alpha\beta\gamma} + \frac{1}{105}\Phi^{\Omega}_{\alpha\beta\gamma\varepsilon} F'''_{\alpha\beta\gamma\varepsilon} \right] \quad (14)$$

where the field values F, defined in eq. (6), are partial derivatives of the field (ϕ) with respect to the components of r, the vector to a field point from the attractor of the atomic basin Ω.

Method of Transferable Atom Equivalents (TAE)

TAEs are atom-based electron density fragments obtained using the AIM approach. The additivity of atomic properties in Bader's theory and the transferability of topological atoms form the underlying concepts for the TAE method. The TAE method was developed to provide electron densities, energies, and properties that approach *ab initio* values at minimal computational cost. The multipole moments, described in eq. (13), are the same expressions of the multipole moments used in the TAE approach. The modeled multicentered multipole moments can be used in eq. (14) to approximate MEPs. But to be useful, the TAE multipoles must be rotated and translated from their default library orientations into their new molecular orientations. This is done through initial translation of an appropriate TAE electron density fragment to the molecular coordinates of each atom. The second step is to define a local coordinate system for each atom within the reconstructed molecule. In the final step, the TAE electron densities, in their native coordinate system, are rotated to overlay with the local coordinate system of the modeled atom. This procedure is accomplished using the TAE molecular reconstruction algorithm called RECON and is described in Figure 1. Both the TAE library construction[53] and RECON algorithm[8] have been described earlier.

During the reconstruction of molecular systems, TAE multipole moments are translated and rotated into appropriate molecular orientations. By virtue of their relationship to AIM atomic basins, TAE multipole moments are atom centered. In the reconstruction of molecules, TAEs do not lose their atom-centered nature; the multipoles are translationally invariant relative to the atomic basin, so they may be translated to their molecular orientations in bulk. The rotation of the TAE multipole moments may be viewed as a transformation from the native atom-centered coordinate system to the local coordinate system of the modeled atom. Like other tensors, the TAE multipoles transform under changes in coordinate

Table 2. EP (kj/mole) Surface Histogram Bin Values.

EP descriptor bin	EP-value
1, 2	−284.9
3, 4	−218.1
5, 6	−151.2
7, 8	−84.3
9, 10	−17.4
11, 12	49.5
13, 14	116.4
15, 16	183.3
17, 18	250.2
19, 20	317.1

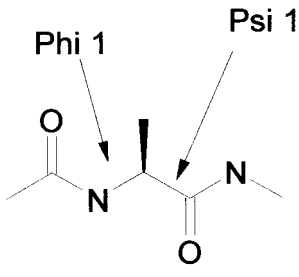

Figure 2. Acetylalanine N'methylamide.

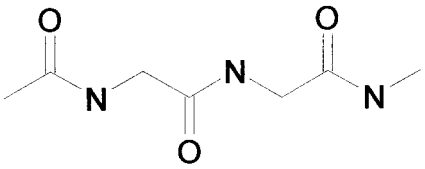

Figure 3. *N*-Acetyldiglycine *N*′-methylamide.

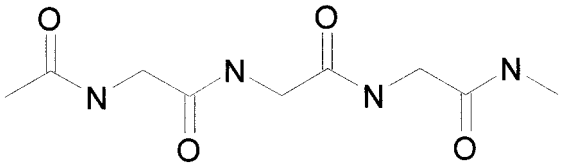

Figure 4. *N*-Acetyltriglycine *N*′-methylamide.

systems according to the law given in eq. (15) for dipoles (first order) through hexadecapoles (fourth order tensors).[54]

$$\mu'_i = \sum_k \alpha_{i'k}\mu$$

$$\Theta'_{ij} = \sum_m \sum_n \alpha_{i'm}\alpha_{j'n}\Theta_{mn}$$

$$\Omega'_{ijk} = \sum_m \sum_n \sum_o \alpha_{i'm}\alpha_{j'n}\alpha_{k'o}\Omega_{mno}$$

$$\Phi'_{ijkl} = \sum_m \sum_n \sum_o \sum_p \alpha_{i'm}\alpha_{j'n}\alpha_{k'o}\alpha_{l'p}\Phi_{mnop} \qquad (15)$$

For the quadrupole moment, Θ_{mn} and Θ'_{ij} are components of the original and rotated tensor, and $\alpha_{i\ m}$ is a component of the rotation matrix for transformation. The element $\alpha_{i\ m}$ of the rotation matrix is equal to the cosine of the angle between axis *i* of the new coordinate system and *m* of the old coordinate system. The TAE

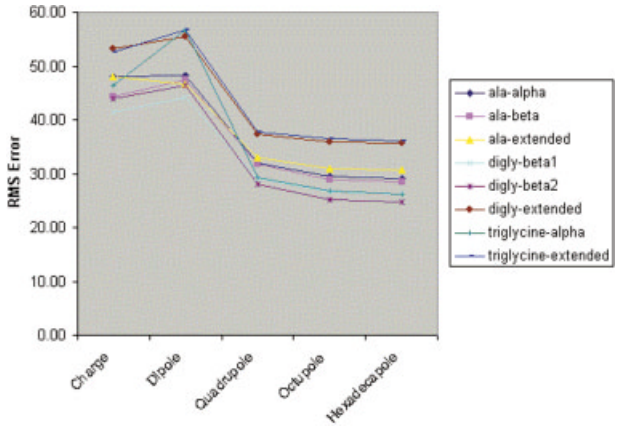

Figure 5. Plot of RMS error vs. rank of TAE multipole used to approximate the EP (kj/mole) on the 0.002 isodensity surface of the molecules in this set.

Table 3. The Extrema of the Analytically Derived EP (kj/mole) on the 0.002 Electrons au^{-3} Isodensity Surface.

	EP min.	EP max.
Alanine-alpha	−231.8	258.9
Alanine-beta	−275.5	284.7
Alanine-extended	−220.1	275.1
Diglycine-beta1	−255.5	334.8
Diglycine-beta2	−261.5	303.7
Diglycine-extended	−232.5	291.6
Triglycine-alpha	−319.9	349.0
Triglycine-extended	−233.7	296.9

multipoles are then saved in their molecular frame and are used to estimate the MEP field of the reconstructed molecule.

Methods

A molecular wave function was calculated at HF/6-31+g* level of theory for each of several conformations (Table 1) of *N*-acetyl, *N*′-methylamide blocked Alanine (Fig. 2), *N*-acetyl, *N*′-methylamide blocked diglycine (Fig. 3), *N*-acetyl, *N*′-methylamide blocked triglycine (Fig. 4) using the GAUSSIAN98 (G98) program.[55] The analytical electrostatic potential was then computed on each 0.002 electron-au^{-3} isodensity surface. These EP-encoded surfaces served as templates for calculating the electrostatic potential using

Table 4. The Extrema of the TAE-MA Derived EP (kj/mole) on the 0.002 Electrons au^{-3} Isodensity Surface.

	EP min.	EP max.	RMS
Alanine-alpha	−270.8	287.3	29.1
Alanine-beta	−330.8	244.0	28.5
Alanine-extended	−227.7	286.2	30.6
Diglycine-beta1	−277.7	330.1	26.2
Diglycine-beta2	−278.0	270.7	24.7
Diglycine-extended	−217.6	231.2	35.6
Triglycine-alpha	−397.9	345.8	26.4
Triglycine-extended	−212.8	236.0	36.1

Table 5. The Extrema and Root Mean Square Error of the GASTEIGER Charged Derived EP (kj/mole) on the 0.002 Electrons au^{-3} Isodensity Surface.

	EP min.	EP max.	RMS
Alanine-alpha	−132.0	87.1	37.5
Alanine-beta	−169.4	114.2	35.9
Alanine-extended	−110.4	88.9	38.3
Diglycine-beta1	−133.2	131.2	39.5
Diglycine-beta2	−135.1	106.6	40.5
Diglycine-extended	−112.2	95.8	41.9
Triglycine-alpha	−200.9	138.3	39.6
Triglycine-extended	−110.2	96.5	41.5

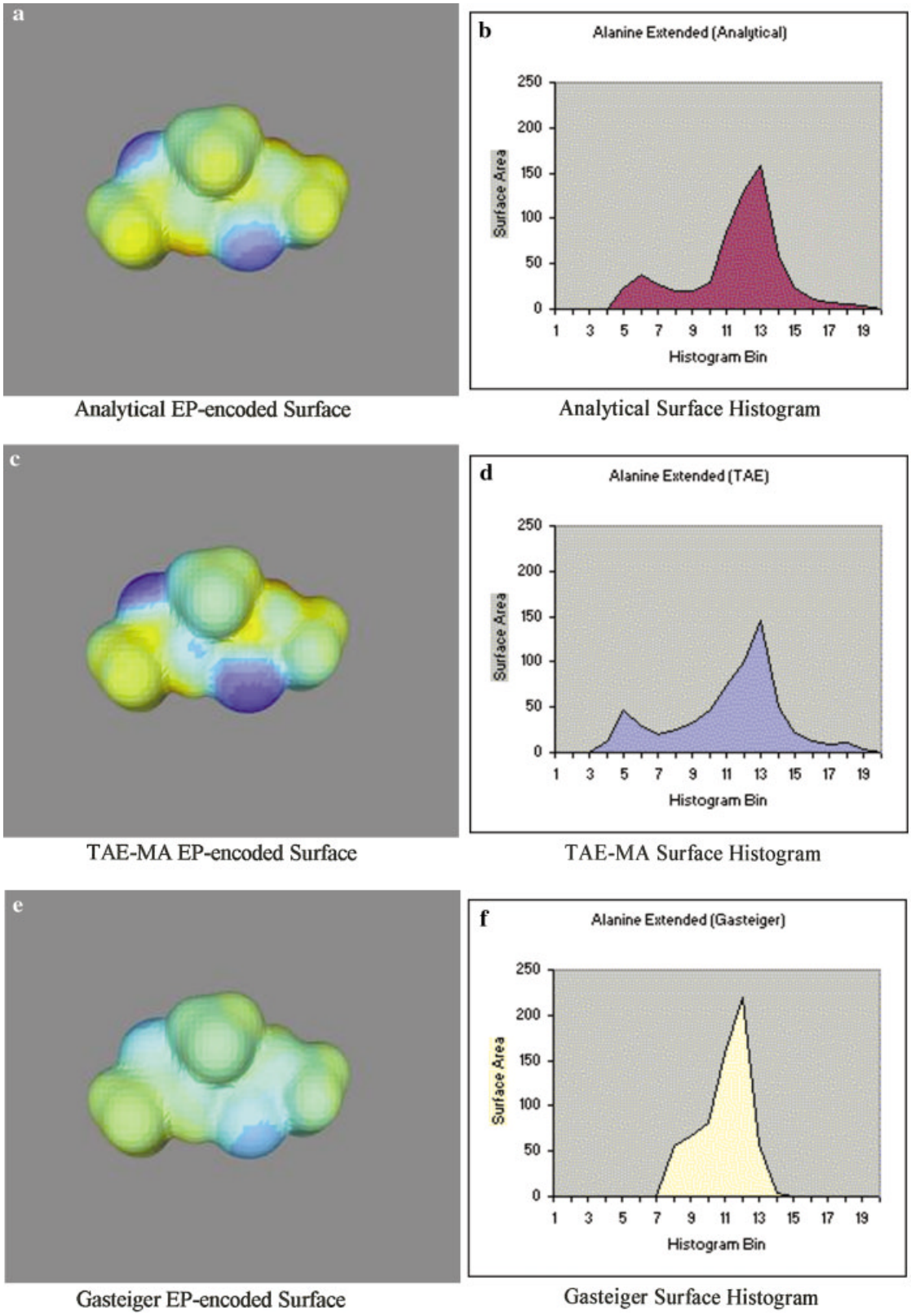

Figure 6. EP-encoded isodensity surface and surface histogram plots of the analytical, TAE-MA, and Gasteiger-derived EPs for the extended conformation of alanine.

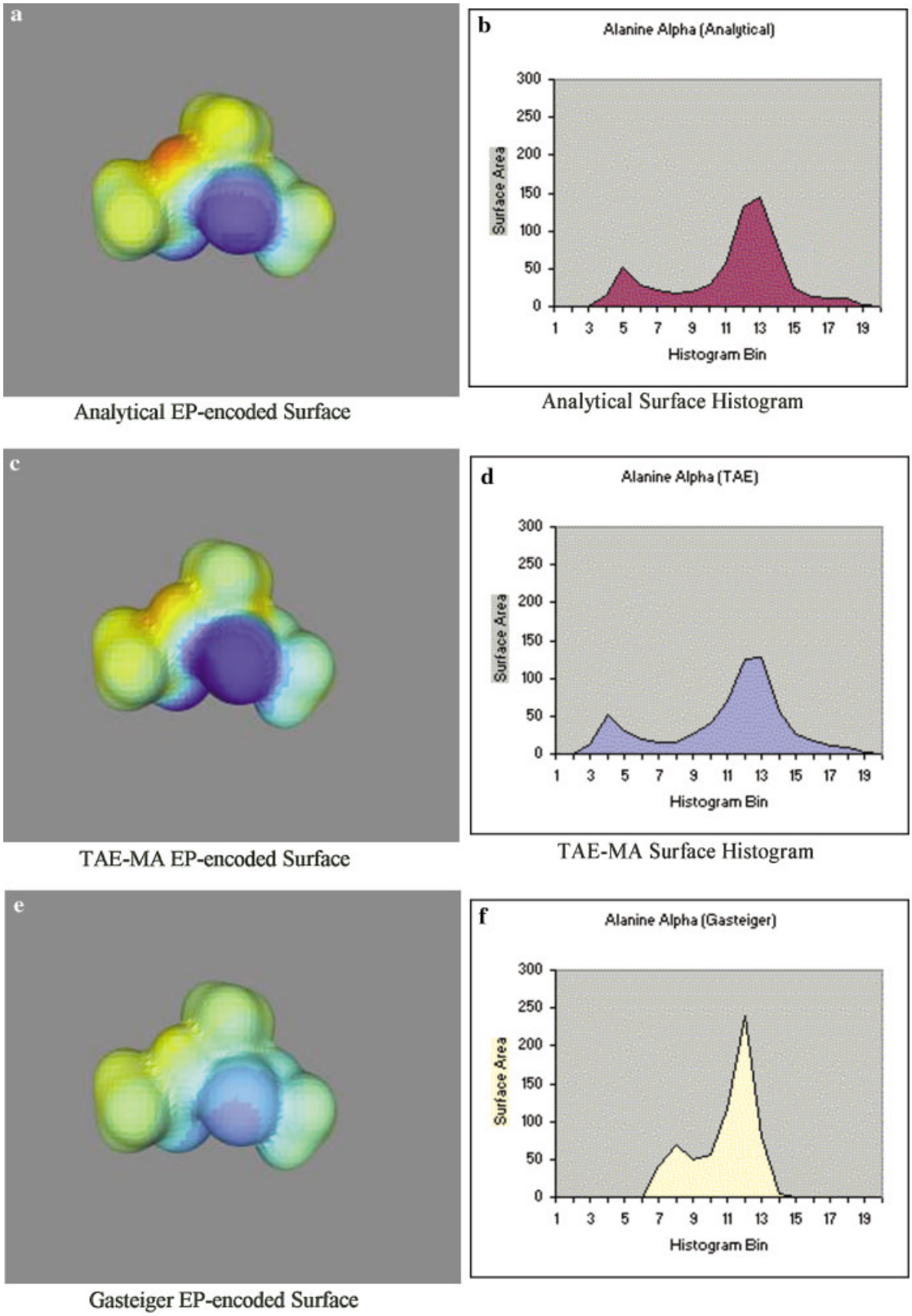

Figure 7. EP-encoded isodensity surface and surface histogram plots of the analytical, TAE-MA, and Gasteiger-derived EPs for the alpha-conformation of alanine.

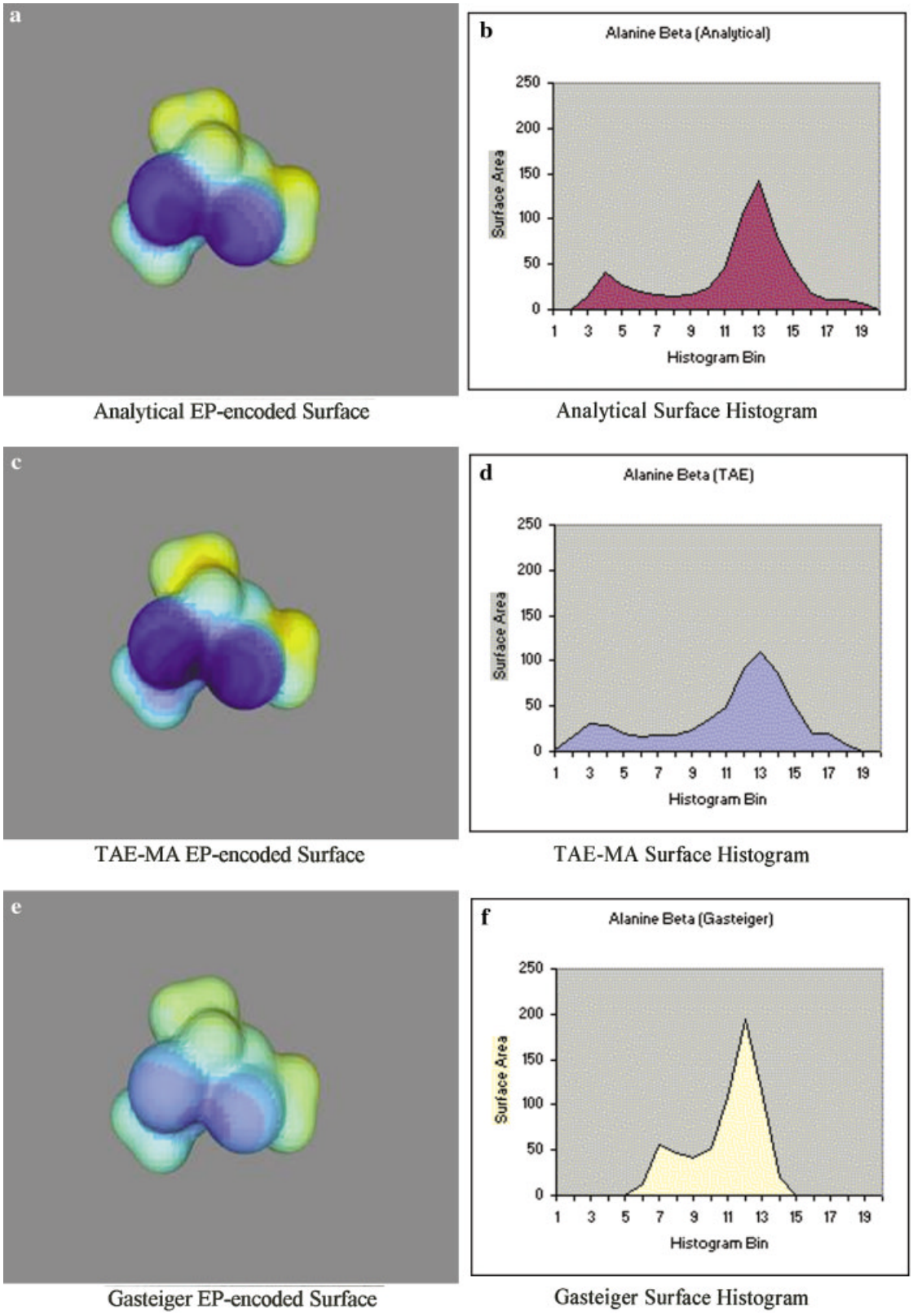

Figure 8. EP-encoded isodensity surface and surface histogram plots of the analytical, TAE-MA, and Gasteiger-derived EPs for the beta-conformation of alanine.

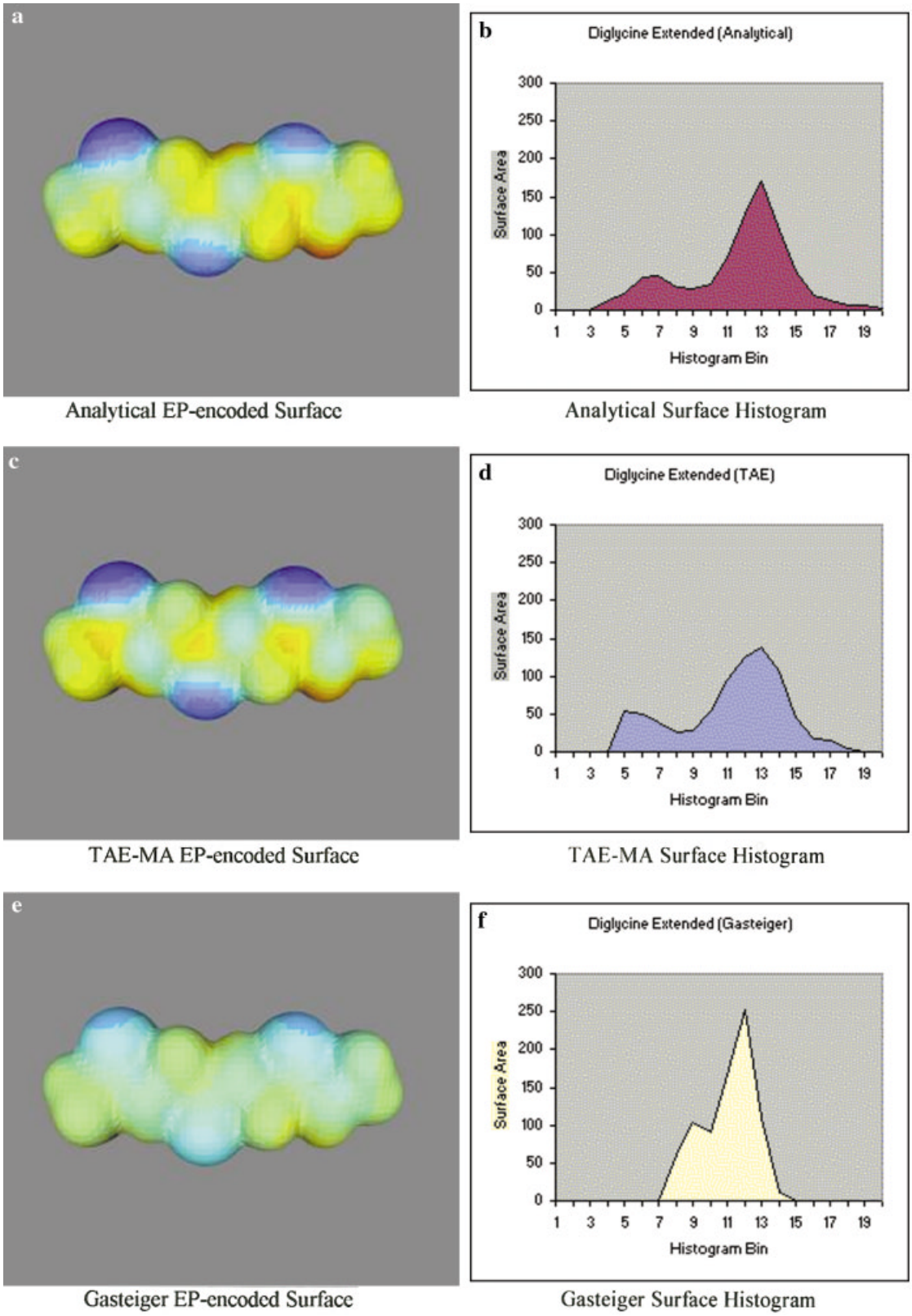

Figure 9. EP-encoded isodensity surface and surface histogram plots of the analytical, TAE-MA, and Gasteiger-derived EPs for the extended confirmation of diglycine.

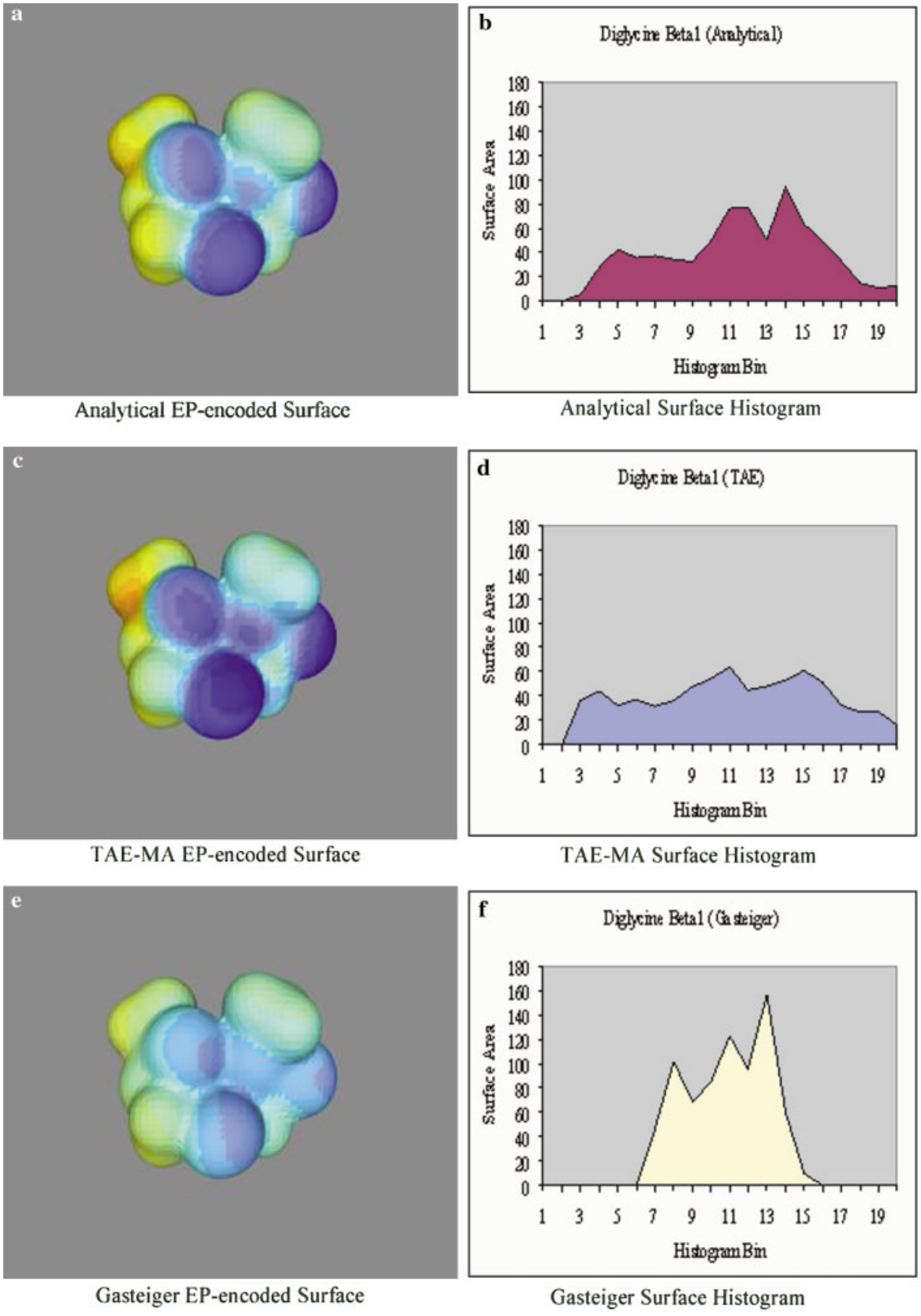

Figure 10. EP-encoded isodensity surface and surface histogram plots of the analytical, TAE-MA, and Gasteiger-derived EPs for the beta 1 conformation of diglycine.

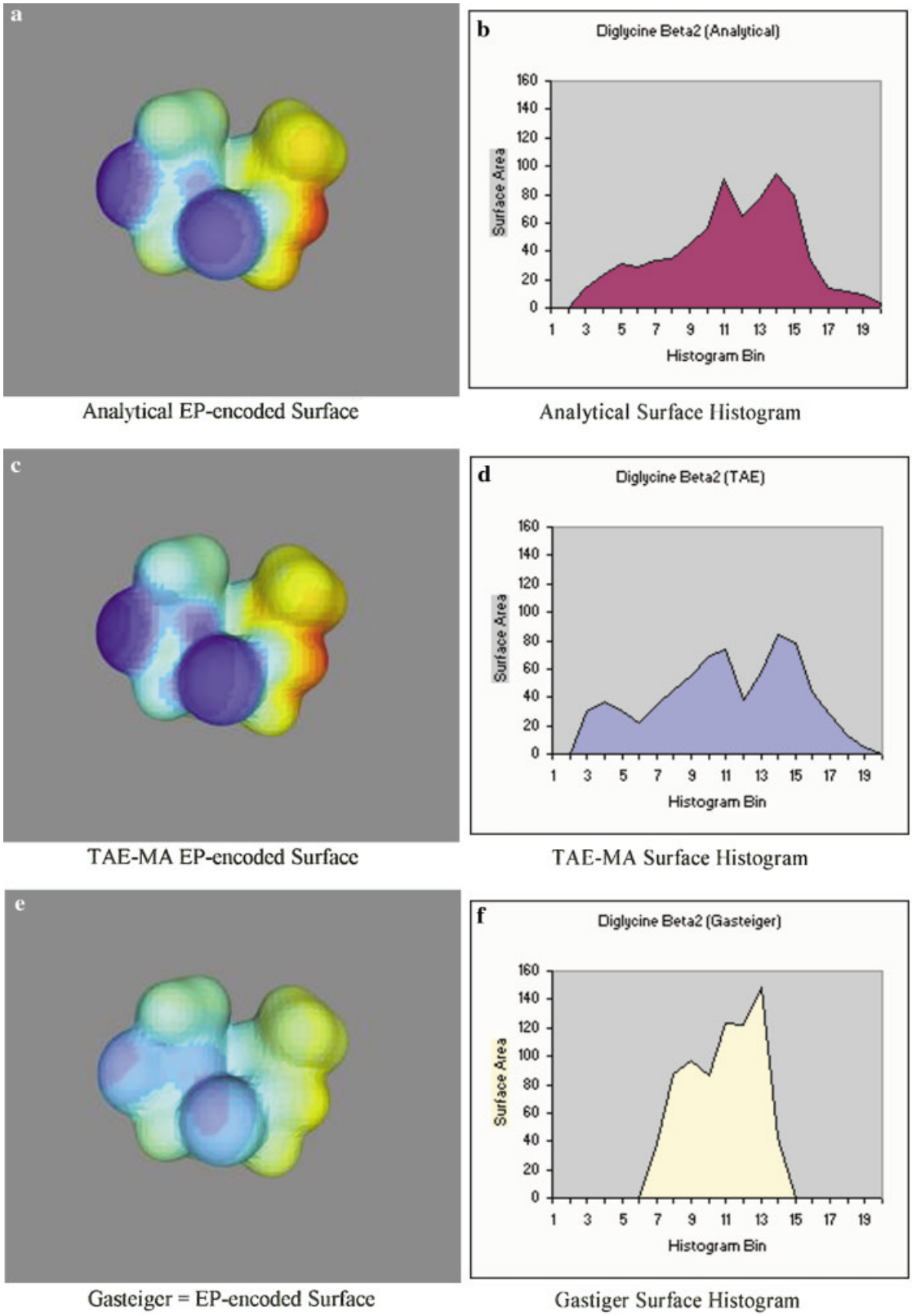

Figure 11. EP-encoded isodensity surface and surface histogram plots of the analytical, TAE-MA, and Gasteiger-derived EPs for the beta 2 conformation of diglycine.

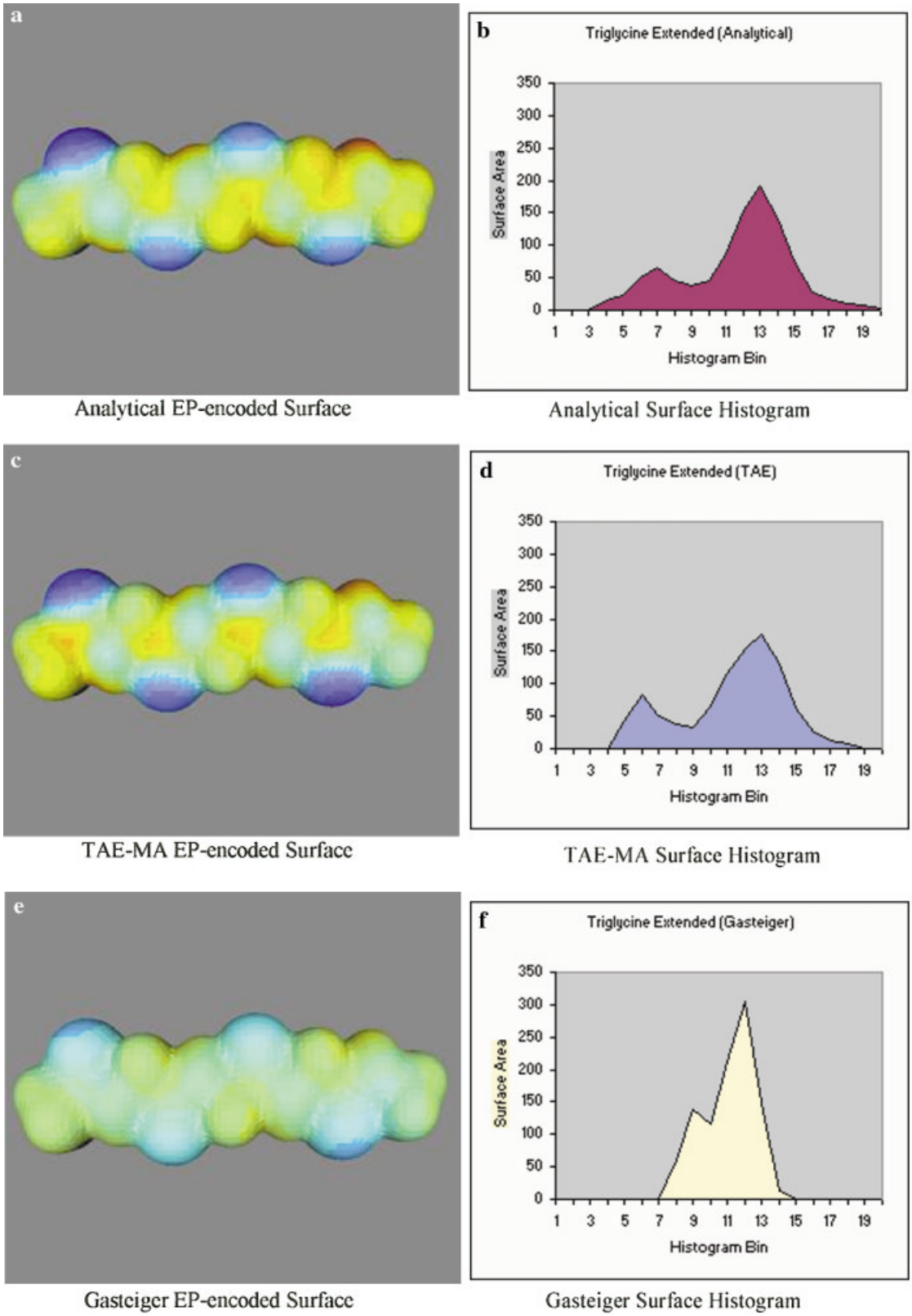

Figure 12. EP-encoded surface and surface histogram plots of the analytical, TAE-MA, and Gasteiger-derived EPs for the extended conformation of triglycine.

multipole moments derived from TAE reconstruction of each conformer of the peptides. The locally developed program RECON_SURF_EP was then used to calculate EP values on each molecular isodensity surface. The EP surface files generated in this manner can be visually compared to the analytical EP. This program also generates data files that can be used to compare the surface EPs derived from analytical and TAE multipole analysis (TAE-MA). Included in the output data files are values for the root-mean-square error [eq. (16)], the minimum and maximum values of the EP and molecular surface EP histogram data arising from these comparisons.

$$RMS = \sqrt{\frac{\sum_i (EP_{\text{analytical}}(i) - EP_{\text{calculated}}(i))^2}{N}} \qquad (16)$$

The molecular surface EP histogram data are generated by finding the total surface area that falls between a set of predetermined values of the EP on the van der Waals surface (shown in Table 2). As a useful comparison, the same output is generated by calculating molecular electrostatic potential distributions on molecular isodensity surfaces using Gasteiger charges.

Results

As part of the process of comparing surface EP distributions with *ab initio* results for the test molecules in several conformations, it is useful to examine the convergence properties of the atom-centered multipole expansions used in this study. Figure 5 displays the relationship between the order of the moment and the RMS error on the van der Waals surface relative to *ab initio* EP distributions. For all of the molecules studied in this set it is apparent that the MEP converges at the quadrupole moment since including the octupole and hexadecapole moment only gave marginal increases in accuracy (as determined by RMS error). This is in contrast to observations made by Kosov and Popelier,[30] in which they observed suitable convergence of the AEP occurs at the hexadecapole moment. For the sake of completeness, the MEP results reported in the remainder of this work will be those derived from multipole moments up to the hexadecapole moment.

As mentioned above, the isodensity surface EP distributions computed using the Gasteiger point charge model and TAE multicentered multipole analysis were compared to the surface EP distribution obtained from analytical evaluation of each conformer wave function at the HF/6-31+G* level of theory. Table 3 gives the values of the analytical minimum and maximum EP on the isodensity surface for the eight molecules/conformations studied in this article. The minimum and maximum EP for the TAE-MA and the Gasteiger charges are shown in Tables 4 and 5, together with RMS error values. The TAE-MA method was found to give a better representation of the EP distribution on the isodensity surface than the Gasteiger charges. The average RMS error for the eight peptide structures was 29 kJ mol^{-1} for the TAE MA vs. the 39 kJ mol^{-1} for the Gasteiger charges. In addition, the surface EP derived using Gasteiger charges does not cover the complete range of EP spanned by the analytical EP, whereas the TAE-MA EP covers the same range albeit slightly more negative. In Table 4 it can also be seen that the RMS error for the TAE-MA EP is highest for the extended conformation for all of the molecules in this set. The RMS error alone does not accurately reflect the differences between the approximate and analytical EP distributions. These dissimilarities can be more adequately quantified by analyzing the distribution of the EP values on the isodensity surface through the use of surface area histograms. Figures 6–13 show the EP-encoded surface as well as the surface histogram plots for all of the conformations of the small peptides studied in this article. We will focus on the EP distribution for the extended and alpha conformations of triglycine (the analysis of the conformations of alanine and diglycine show similar trends and results).

Table 6. Surface Area (au^2) in Each Histogram Bin for the Extended Conformation of Triglycine for the EP Derived Using Analytical, TAE-MA, and Gasteiger Charges.

	Analytical triglycine extended	TAE-MA triglycine extended	Gasteiger triglycine extended
Bin_1	0.0	0.0	0.0
Bin_2	0.0	0.0	0.0
Bin_3	0.0	0.0	0.0
Bin_4	14.4	0.0	0.0
Bin_5	21.6	43.0	0.0
Bin_6	51.1	82.9	0.0
Bin_7	65.0	50.6	0.0
Bin_8	45.0	36.6	60.1
Bin_9	38.6	32.1	138.1
Bin_10	46.0	65.5	116.6
Bin_11	85.4	115.8	212.7
Bin_12	153.5	154.0	305.6
Bin_13	191.3	175.6	147.9
Bin_14	141.9	130.5	12.7
Bin_15	75.8	62.1	0.0
Bin_16	28.1	24.6	0.0
Bin_17	18.5	13.8	0.0
Bin_18	8.9	6.7	0.0
Bin_19	6.3	0.0	0.0
Bin_20	2.5	0.0	0.0

Analytical MEP vs. Approximated MEP for Triglycine in the Extended Conformation

The analytical EP on the isodensity surface of triglycine in the extended conformation is compared to the EP derived from TAE-MA and Gasteiger charges. Figure 12a, c, and e displays the EP-encoded isodensity surfaces for analytical, TAE-MA, and Gasteiger methods, respectively, while Figure 12b, d, and f displays the corresponding histogram plots. The values used to generate the plots are shown in Table 6. As seen in the table and the figures, the Gasteiger charges do not reproduce the analytical EP on the isodensity surface to the degree seen in the TAE multipole results. The minima and maxima of the EP on the isodensity surface for the Gasteiger charge-derived EP are also compressed: -110.2 and 97 kJ mol^{-1}, respectively. This range is much less than the range of -234 to 297 kJ mol^{-1} observed in the analyti-

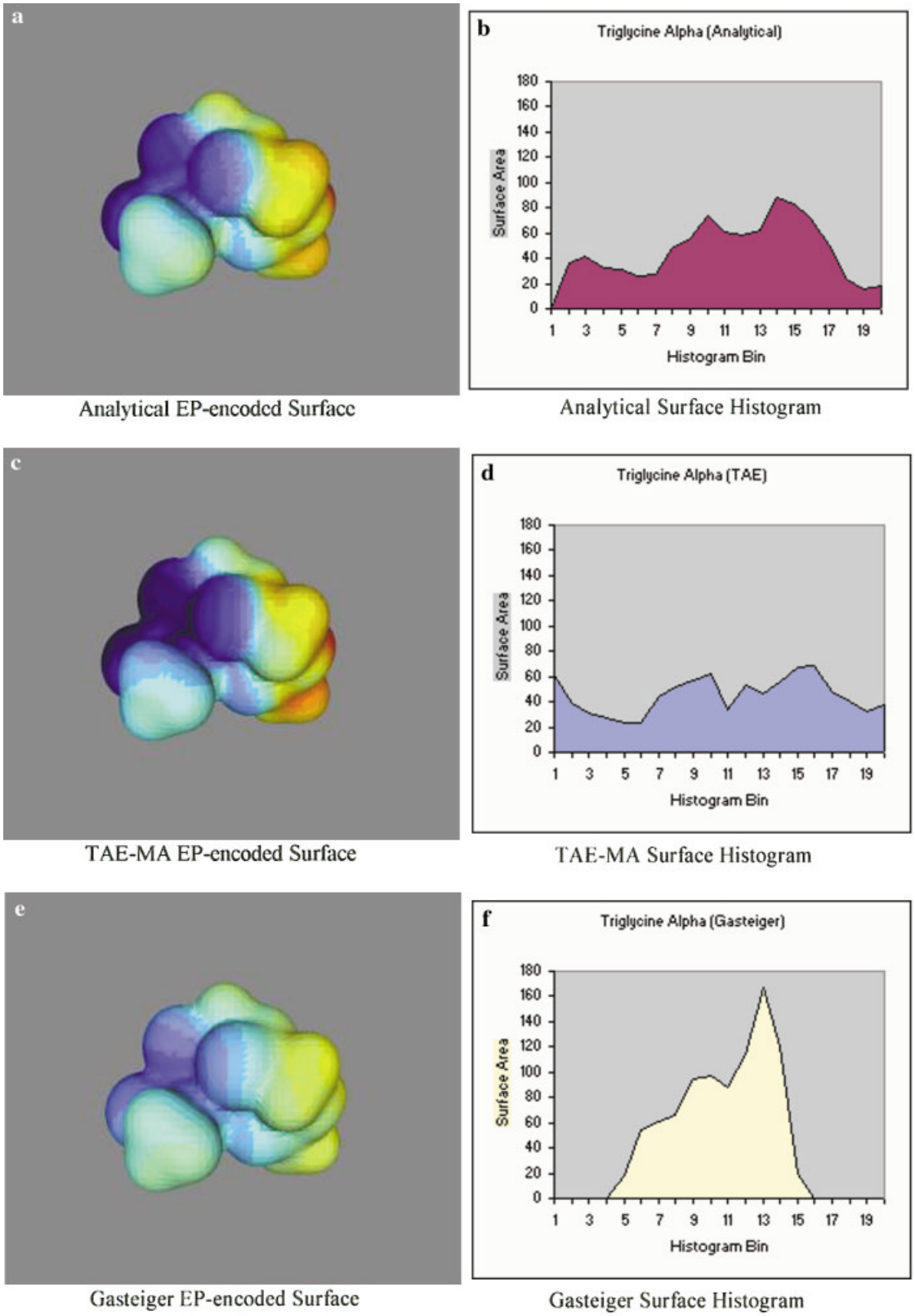

Figure 13. EP-encoded isodensity surface and surface histogram plots of the analytical, TAE-MA, and Gasteiger-derived EPs for the alpha conformation of triglycine.

cally-derived EP, or for the TAE-MA-derived EP extrema that range from −212 to 236 kJ mol^{-1}. Analyzing the histogram plots in Figure 12b, d, and f reveal that the TAE-MA almost quantitatively recovers the analytical EP distribution.

Analytical vs. Approximated EP for Triglycine in Alpha Conformation

The MEP derived from TAE multipole expansions and Gasteiger point charges were compared to the analytically derived MEP on the isodensity surface of triglycine in the alpha conformation. The surface area histograms, for the three methods, are shown in Figure 13b, d, and f, while Figure 13a, c, and e shows the EP encoded on the isodensity surface. The corresponding numerical values for the plots are shown in Table 7. As demonstrated in Figure 13e–f, The Gasteiger charges do not accurately reproduce the electrostatic potential on the isodensity surface. The range for the Gasteiger EP on the surface is −201 to 138 kJ mol^{-1} (Table 5), which is much less than the analytical EP range of −320 to 349 kJ mol^{-1} (Table 3). The TAE-MA EP more closely approximates the analytical range with values from −398 to 346 kJ mol^{-1} (Table 4). This observation may be reinforced further by analyzing the distribution of the EP values on the isodensity surface shown in the histogram plot in Figure 13a, c, and e. Analysis of these histogram plots reveal that the surface distribution of the TAE-MA EP is qualitatively similar to that of the analytical EP. The morphology of the Gasteiger EP distribution on the isodensity surface is quite different, and of limited range when compared to the analytical EP distribution. An apparent artifact of the TAE-MA multipole method is observed in this case, where a tail is seen near the most negative potentials. This observation is highlighted in Table 8, where it is shown that the TAE-MA EP distribution has 48 au^2 more surface area with EP values of less than −218.8 kJ mol^{-1}.

Table 7. Surface Area (au^2) in Each Histogram Bin for the Alpha Conformation of Triglycine for the EP Derived Using Analytical, TAE-MA, and Gasteiger Charges.

	Analytical triglycine alpha	TAE-MA triglycine alpha	Gasteiger triglycine alpha
Bin_1	0.0	60.6	0.0
Bin_2	35.9	38.8	0.0
Bin_3	41.2	31.6	0.0
Bin_4	32.9	27.0	0.0
Bin_5	30.8	23.9	18.6
Bin_6	25.6	23.5	54.7
Bin_7	27.4	44.7	61.2
Bin_8	48.3	51.4	66.6
Bin_9	55.6	56.8	94.6
Bin_10	73.4	62.7	97.5
Bin_11	61.1	33.9	88.7
Bin_12	57.8	53.0	114.1
Bin_13	61.9	47.1	167.2
Bin_14	87.9	55.7	120.4
Bin_15	83.4	66.9	19.9
Bin_16	71.8	68.2	0.0
Bin_17	51.1	47.8	0.0
Bin_18	23.2	40.7	0.0
Bin_19	15.5	31.7	0.0
Bin_20	18.4	37.2	0.0

Table 8. Amount of Surface (au^2) Broken into Very Negative, Negative, Positive, and Very Positive for the EP on the Isodensity Surface of Triglycine in the Extended Conformation for the EP Derived Using Analytical, TAE-MA, and Gasteiger.

	Analytical triglycine alpha	TAE-MA triglycine alpha	Gasteiger triglycine alpha
Bins 1–4	110.0	158.0	0.0
Bins 5–10	261.2	263.1	393.2
Bins 11–15	352.1	256.6	510.2
Bins 16–20	180.0	225.7	0.0

The TAE-MA multipolar reproductions were found to generate EPs on the isodensity surface that are more polar than the analytical EP. There can be at least two reasons for this error: the first is that the assumption of transferability without polarization of the multipole moments may be inadequate. In the alpha conformation of triglycine there is an internal hydrogen bond between a carbonyl and the amide nitrogen of the peptide linkage. In the real molecule, the formation of the hydrogen bond is accompanied by polarization of the electron density. Besides rigid translation and rotation of the components of the TAE multipole, they are static in this implementation—meaning that the TAE multipole moments do not "feel" the effects of the field of other multipole moments upon rotation. We are experimenting with polarizing the TAE multipoles using the polarizability moments defined by Laidig.[37,39] The other source of error may be that the 0.002 electrons au^{-3} isodensity surface used to calculate the electrostatic potential is within the radius of convergence of the multipoles. The minimum distance between an atom and a point on the isodensity surface of triglycine in the alpha conformation is 1.98 au (1.04 Angstroms). The reason for selecting the 0.002 electrons au^{-3} isodensity surface is that electronic properties calculated on the 0.002 isodensity surface have been used extensively as QSAR/QSPR descriptors by our group and others.[56] Among the TAE descriptors, the surface histogram of the EP on the isodensity surface has served as a useful alternative to the polar surface area descriptors used in ADMET modeling. Calculating the MEPs on the solvent-accessible surface should enhance the quantitative agreement between the TAE and analytical MEPs.

Conclusions

Molecular reconstruction using multicentered multipole expansions of TAE atomic electron density distributions provides a scheme for estimating MEP that is accurate, conformationally responsive, computationally inexpensive, and transferable. In contrast to other multipole expansion schemes that rely on arbitrary

partitioning of the molecular electron density, or schemes that employ empirically determined point charges, the TAE-MA scheme employs a nonarbitrary partitioning of the molecular electron density into constituent atoms based on the theory of Atoms in Molecules. Atoms defined in AIM theory [using the constraint in eq. (10)] are known to satisfy the virial, hypervirial, and Ehrenfest force theorems and are transferable to other molecules with atoms in similar environments. The TAE-MA method provides a semiquantitative picture of the MEP, obtained at a fraction of the computational cost of a full-fledged *ab initio* calculation, and one that is better than other approximate methods in common use.

The TAE-MA scheme thus constitutes a very fast and reasonably accurate method of calculating the electrostatic potential on an isodensity surface. When compared to the analytical EP at HF/6-31+G* level of theory on the isodensity surface, the TAE multipoles produced surface histograms with similar morphology. Molecular EP distributions derived using Gasteiger point charges were found to have larger errors when compared to ab initio reference values.

QSAR/QSPR descriptors computed from the TAE RECON method have proven successful in describing a variety of chemical and biochemical phenomena. Most of these descriptors are, however, insensitive to molecular conformation. Supplementing these descriptors with TAE-MA derived EP could provide a means to generate conformationally sensitive electrostatic descriptors, which should have a wide range of applicability.

References

1. Breneman, C.M.; Martinov, M. In Molecular Electrostatic Potential: Concept and Applications; Murray, J. S.; Sen, K. Eds.; Elsevier: Amsterdam, 1996; p 143.
2. Brinck, T.; Murray, J. S.; Politzer, P. J Org Chem 1993, 58, 7070.
3. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A. J Am Chem Soc 1993, 115, 9620.
4. Tasi, G.; Palinko, I. Using Molecular Electrostatic Potential Maps for Similiarity Studies. Topics in Current Chemistry 1995, 174, 45.
5. Politzer, P.; Murray, J. S.; Peralta-Inga, Z. Int J Quantum Chem 2001, 85, 676.
6. Gross, K. C.; Seybold, P. G.; Peralta-Inga, Z.; Murray, J. S.; Politzer, P. J Org Chem 2001, 66, 6919.
7. Murray, J. S.; Politzer, P. Theochemistry 1998, 425, 107.
8. Breneman, C. M.; Rhem, M. J Comput Chem 1997, 18, 182.
9. Mazza, C. B.; Sukumar, N.; Breneman, C. M.; Cramer, S. M. Anal Chem 2001, 73, 5457.
10. Liu, R.; Sun, H.; So, S.-S. J Chem Inf Comput Sci 2001, 41, 1623.
11. Stenberg, P.; Luthman, K.; Artursson, P. Pharmaceut Res 1999, 16, 205.
12. Wiberg, K. B.; Rablen, P. R. J Comput Chem 1993, 14, 1504.
13. Momany, F. A. J Phys Chem 1978, 82, 592.
14. Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. J Comput Chem 1996, 17, 367.
15. Francl, M. M.; Chirlian, L. E. Rev Comput Chem 2000, 14, 1.
16. Reynolds, C. A.; Essex, J. W.; Richards, W. G. J Am Chem Soc 1992, 114, 9075.
17. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. J Phys Chem 1993, 97, 10269.
18. Colonna, F.; Evleth, E. M. Chem Phys Lett 1993, 212, 665.
19. Angyan, J. G.; Chipot, C. Int J Quantum Chem 1994, 52, 17.
20. Gadre, S. R.; Pundlik, S. S.; Shrivastava, I. H. Proc Indian Acad Sci (Chem Sci) 1994, 106, 303.
21. Aleman, C.; Orozco, M.; Luque, F. J. Chem Phys 1994, 189, 573.
22. Koch, U.; Egert, E. J Comput Chem 1995, 16, 937.
23. Sokalski, W. A.; Poirier, R. A. Chem Phys Lett 1983, 98, 86.
24. Stone, A. J. Chem Phys Lett 1981, 83, 233.
25. Stone, A. J.; Alderton, M. Mol Phys 1985, 56, 1047.
26. Strasburger, K.; Sokalski, W. A. Chem Phys Lett 1994, 221, 129.
27. Koch, U.; Stone, A. J. J Chem Soc Faraday Trans 1996, 92, 1701.
28. Williams, D. E. J Comput Chem 1988, 9, 745.
29. Sokalski, W. A.; Sawaryn, A. J Chem Phys 1987, 87, 526.
30. Kosov, D. S.; Popelier, P. L. A. J Phys Chem A 2000, 104, 7339.
31. Kedzierski, P.; Sokalski, W. A. J Comput Chem 2001, 22, 1082.
32. Vigne-Maeder, F.; Claverie, P. J Chem Phys 1988, 88, 4934.
33. Faerman, C. H.; Price, S. L. J Am Chem Soc 1990, 112, 4915.
34. Price, S. L.; Stone, A. J. J Chem Soc Faraday Trans 1992, 88, 1755.
35. Pluta, T. S. Electronic Multipole Moments and Polarizabilities and Their Application to Weak Intermolecular Interactions, in Chemistry; University of Florida: Gainesville, 1990; p 113.
36. Kosov, D. S.; Popelier, P. L. A. J Chem Phys 2000, 113, 3969.
37. Laidig, K. E.; Bader, R. F. W. J Chem Phys 1990, 93, 7213.
38. Laidig, K. E. J Phys Chem 1993, 97, 12760.
39. Laidig, K. E. Can J Chem 1996, 74, 1131.
40. Bader, R. F. W.; Larouche, A.; Gatti, C.; Carroll, M. T.; MacDougall, P. J.; Wiberg, K. B. J Chem Phys 1987, 87, 1142.
41. Bader, R. F. W.; Carroll, M. T.; Cheeseman, J. R.; Chang, C. J Am Chem Soc 1987, 109, 7968.
42. Bader, R. F. W.; Becker, P. Chem Phys Lett 1988, 148, 452.
43. Bader, R. F. W.; Zou, P. F. Chem Phys Lett 1992, 191, 54.
44. McLean, A. D.; Yoshimine, M. D. J Chem Phys 1967, 47, 1927.
45. Applequist, J. Chem Phys 1984, 85, 279.
46. Applequist, J. J Chem Phys 1985, 83, 809.
47. Stogryn, D. E.; Stogryn, A. Mol Phys 1966, 11, 371.
48. Sokalski, W. A.; Hariharan, P. C.; Kaufman, J. J. Int J Quantum Chem Quantum Biol Symp 1987, 14, 111.
49. Bader, R. F. W. J Chem Phys 1980, 73, 2871.
50. Bader, R. F. W.; Anderson, S. G.; Duke, A. J. J Am Chem Soc 1979, 101, 1389.
51. Bader, R. F. W. Atoms in Molecules: A Quantum Theory; Oxford: Oxford Press, 1990.
52. Bader, R. F. W.; Tang, T.-H.; Tal, Y.; Biegler-Konig, F. W. J Am Chem Soc 1982, 104, 946.
53. Thompson, T. R. In Chemistry; Rensselaer Polytechnic Institute: Troy, 1994; p 1000.
54. Nye, J. F. Physical Properties of Crystals, Their Representation by Tensors and Matrices; Oxford: Clarendon Press, 1957; p 322.
55. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head–Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian 94; Gaussian, Inc.: Pittsburgh, PA, 1995.
56. Murray, J. S.; Politzer, P.; Famini, G. R. Theochemistry 1998, 454, 299.