Generic Molecular Mechanics as Based on Local Quantum Description of Molecular Electronic Structure

A. M. TOKMACHEV, A. L. TCHOUGRÉEFF

L. Ya. Karpov Institute of Physical Chemistry, Vorontsovo pole 10, Moscow 103064, Russia, and Center for Computational Chemistry at the M. V. Keldysh Institute for Applied Mathematics of RAS, Miusskaya pl., 4, Moscow 125047, Russia

ABSTRACT: Molecular mechanics (MM) is now widely used for modeling potential energy surfaces of large organic molecules. All MM schemes use ad hoc forms for the potential functions and parameters specially adjusted to fit experimental or quantum chemical data. In this work we attempt to deduce a generic MM scheme starting from a local quantum mechanical description of molecular electronic structure. The basis for this derivation is the trial electronic wave function in the form of antisymmetrized product of strictly localized geminals. The MM scheme obtained does not require adjusting any parameters. The quality of numerical estimates obtained by this scheme is analyzed.   © 2002 Wiley Periodicals, Inc. Int J Quantum Chem 88: 403–413, 2002

Key words: molecular mechanics; potential energy surfaces; quantum mechanics; antisymmetrized product of strictly local geminals

1. Introduction

The molecular mechanics (MM) [1, 2] numerical schemes are widely used for modeling potential energy surfaces (PESs) of large molecular systems. The practical usefulness of the MM schemes is obvious: They provide the equilibrium geometries, dynamic matrices, and other features of the systems of large size when the quantum chemical procedures become inapplicable due to enormous computational costs. The MM results are sufficiently accurate for many purposes and for large objects MM schemes are the best considering the cost–efficiency ratio.

Despite the intense employment of the MM schemes they remain purely empirical and are only supported by the agreement of calculated and experimentally observed quantities. Such level of substantiation means that the success of the MM schemes in general must be considered as an experimental fact. This fact, in turn, requires theoretical explanation that as any sequential description of molecular structure and properties must be based on quantum mechanics (QM). The question of the interrelations between the MM and QM approaches is well known in the literature (see, for example, Ref. [3]). It is, however, not

Correspondence to: A. M. Tokmachev; e-mail: tokm@chat.ru

only of academic interest. Additional attention to the foundations of the MM schemes is caused by increasing interest to the hybrid QM/MM schemes [4], where a chemically active part of the whole system is described by precise quantum chemical methods while the rest (more or less inert environment) is represented by an appropriate MM method. The hybrid QM/MM schemes use a combination of sometimes high-quality QM and an empirical MM procedure. The status of the combined approach thus remains unclear, as is the form of the junction between the quantum and classic subsystems in the hybrid QM/MM schemes, which is commonly chosen ad hoc because the relations between the QM and MM approaches are not clearly defined. To derive successively the explicit form of the junction it is necessary to assume also that of the wave function underlying the MM description. In this context the concept of the MM must be clarified. The common feature of all MM schemes is that the contributions to the energy from the bonded and nonbonded atoms are treated differently, that is, the bonding is somewhat externally assumed. Different MM schemes employ different specific forms of potentials and numerous systems of their parameters fitted to reproduce a variety of characteristics [5–8] for a more or less wide class of molecules. Also during the evolution of the MM approach itself some increasingly sophisticated contributions are added to the original simple picture that allow to extend the MM approach to increasingly complex and not transparently tractable classes of the molecules where the metal-containing complexes and other molecules with significant electron redistribution must be mentioned [9–11]. For the latter cases the modern MM schemes do not represent the energy as an explicit function of the geometry parameters but include some kind of optimization scheme (for example, in Ref. [11] the scheme adjusting the charges on atoms on the ground of some simple criterion is employed). We see that the definition of MM becomes even more diffuse. Summarizing, we can define the MM scheme as one representing the energy as a combination of bonding and nonbonding contributions that are either explicit functions of molecular geometry structure or can be obtained from the latter by simple noniterative optimization procedure. In this article we suggest a derivation and numerical test of a scheme of this type—a local parametric expression for the total energy of organic molecules on the basis of a quantum description that uses a special form of the trial wave function of electrons, namely that of the antisymmetrized product of strictly local geminals (APSLG).

The reason to choose the APSLG form of the underlying wave function is some intimate similarity in the energy expressions of the APSLG and MM schemes. In our previous works [12, 13] we derived the required form of the junction between the QM and MM subsystems in an assumption that the inert part of the combined system can be represented by the wave function in the APSLG form [14–16]. At the same time, the APSLG-based QM method is a complicated, self-consistent iteration procedure with numerous diagonalizations of effective Hamiltonian matrices although of small dimensionality. Nevertheless, the APSLG approach seems appropriate for obtaining the MM-type schemes for different reasons: The APSLG approach provides the local description of the molecular electronic structure; it represents the energy as the sum of intrabond and interbond contributions and thus can be considered as a natural starting point for construction of different additive schemes. Because the MM scheme is itself a parametric empirical procedure, it is unnecessary to use a sophisticated nonempirical description for deriving the latter. It is more logical to use a semiempirical method as a starting point. A semiempirical implementation of the APSLG trial wave function has been recently developed with use of the MINDO/3 form of the Hamiltonian and resulted in an $O(N)$-scaling method suitable for describing organic molecules. This method and its parameterization are described in details in Refs. [16–18].

The article is organized as follows. In Section 2 we first briefly summarize the main features of the APSLG approach. Next, the one-step procedures for optimization of geminal amplitudes and hybridization matrices are discussed. In Section 3, we propose a series of approaches for evaluation of the electronic energy of molecules in the APSLG approximation and evaluate their accuracy. Next, we discuss the possibility of treating these approaches as generic MM-type descriptions. Section 4 contains the summary of results.

2. Construction of Generic MM Schemes

2.1. WAVE FUNCTION AND ENERGY OF UNDERLYING APSLG-BASED QM METHOD

The APSLG wave function of electrons in the molecule has the form

$$|\Phi\rangle = \prod_m g_m^+|0\rangle \qquad (1)$$

and is an example of the general "group function method" [19] wave function. Each geminal corresponds to a chemical bond or a lone electron pair. In the case of a chemical bond, the geminal (singlet two-electron function) is constructed of two orbitals $|r\rangle$ and $|l\rangle$ assigned to the right and left ends of the bond, respectively. In the second quantization formalism it can be written as

$$g_m^+ = u_m r_{m\alpha}^+ r_{m\beta}^+ + v_m l_{m\alpha}^+ l_{m\beta}^+ + w_m(r_{m\alpha}^+ l_{m\beta}^+ + l_{m\alpha}^+ r_{m\beta}^+). \qquad (2)$$

The first two terms are ionic configurations (two electrons on the right or left end of the bond, respectively), while the last term is the covalent (Heitler–London) configuration. The normalization condition for the geminal Eq. (2) reads

$$u_m^2 + v_m^2 + 2w_m^2 = 1. \qquad (3)$$

In the case of a lone pair only the first (ionic) term makes sense and the geminal can be written as

$$g_m^+ = r_{m\alpha}^+ r_{m\beta}^+. \qquad (4)$$

The orbitals $|r\rangle$ and $|l\rangle$ are chosen as strictly localized hybrid orbitals (HOs) [20], that is, they are obtained by applying 4×4 orthogonal hybridization matrices $h^A \in SO$ (4) to the set of atomic orbitals (AOs) attached to each heavy (nonhydrogen) atom A. The amplitudes of the configurations (u_m, v_m, w_m) for each bond and the matrices h^A for each heavy atom are to be determined [16–18] with use of the variational principle.

The above function is applied to find an estimate for the system energy with the Hamiltonian of the MINDO/3 approximation. In the HO basis the latter can be rewritten as a sum of one- and two-center contributions:

$$H = \sum_A H_A + \frac{1}{2} \sum_{A \neq B} H_{AB}, \qquad (5)$$

where

$$H_A = \sum_{\substack{t_m \in A \\ \sigma}} (U_m^t - \sum_{B \neq A} \gamma_{AB} Z_B) t_{m\sigma}^+ t_{m\sigma}$$
$$- \sum_{\substack{t_{m_1}, t_{m_2} \in A \\ m_1 < m_2}} \sum_\sigma \beta_{t_{m_1} t_{m_2}'}^{AA} (t_{m_1\sigma}^+ t_{m_2\sigma}' + h.c.)$$
$$+ \frac{1}{2} \sum_{\substack{t_{m_1}, t_{m_2} \in A \\ t_{m_3}'', t_{m_4}''' \in A}} \sum_{\sigma\tau} \begin{pmatrix} t & t' & | & t'' & t''' \\ m_1 & m_2 & | & m_3 & m_4 \end{pmatrix}^A t_{m_1\sigma}^+ t_{m_3\tau}''^+ t_{m_4\tau}''' t_{m_2\sigma}' \qquad (6)$$

and

$$H_{AB} = - \sum_{\substack{t_{m_1} \in A \\ t_{m_2}' \in B}} \beta_{t_{m_1} t_{m_2}'}^{AB} \sum_\sigma (t_{m_1\sigma}^+ t_{m_2\sigma}' + h.c.)$$
$$+ \gamma_{AB} \sum_{\substack{t_{m_1} \in A \\ t_{m_2}' \in B}} \sum_{\sigma\tau} t_{m_1\sigma}^+ t_{m_2\tau}'^+ t_{m_2\tau}' t_{m_1\sigma}, \qquad (7)$$

where $h.c.$ stands for the hermitean conjugation and t refers to the r or l HO. Molecular integrals entering this Hamiltonian depend on molecular geometry and on the hybridization matrices h^A. The parameter describing attraction of electrons to a core reads

$$U_m^t = \sum_{i \in A} (h_{mi}^A)^2 U_i(A). \qquad (8)$$

The electron attraction to other cores and repulsion of electrons on the orbitals of different atoms γ_{AB} are invariant under the basis transformations allowed. The one-center, two-electron matrix elements of the Coulomb repulsion are

$$(t_{m_1} t_{m_2}' | t_{m_3}'' t_{m_4}''')^A = \sum_{i \in A} h_{m_1 i}^A h_{m_2 i}^A h_{m_3 i}^A h_{m_4 i}^A (ii|ii)^A$$
$$+ \sum_{i < j \in A} (ii|jj)^A (h_{m_1 i}^A h_{m_2 i}^A h_{m_3 j}^A h_{m_4 j}^A$$
$$+ h_{m_1 j}^A h_{m_2 j}^A h_{m_3 i}^A h_{m_4 i}^A) + \sum_{i < j \in A} (ij|ij)^A$$
$$\times (h_{m_1 i}^A h_{m_2 j}^A h_{m_3 i}^A h_{m_4 j}^A$$
$$+ h_{m_1 i}^A h_{m_2 j}^A h_{m_3 j}^A h_{m_4 i}^A + h_{m_1 j}^A h_{m_2 i}^A h_{m_3 i}^A h_{m_4 j}^A$$
$$+ h_{m_1 j}^A h_{m_2 i}^A h_{m_3 j}^A h_{m_4 i}^A), \qquad (9)$$

where i and j are the AOs. The resonance integral describing one-electron transfers between the HO t_{m_1} (centered on the atom A) and the HO t_{m_2}' (centered on the atom B) is expressed through the resonance integrals in the AO basis:

$$\beta^{AB}_{t_{m_1}t'_{m_2}} = \sum_{i \in A} \sum_{j \in B} h^A_{m_1 i} h^B_{m_2 j} \beta^{AB}_{ij}. \tag{10}$$

The electronic energy is then the sum of four terms:

$$E^{el} = E_1 + E^{(1)}_{Coul} + E_{res} + E^{(2)}_{Coul}, \tag{11}$$

originating from (1) the electron–cores attraction (E_1), (2) one-center electron–electron repulsion ($E^{(1)}_{Coul}$), (3) resonance intrabond interaction (E_{res}), and (4) two-center electron–electron repulsion ($E^{(2)}_{Coul}$). These contributions to the energy have the form

$$E_1 = 2 \sum_A \sum_{t_m \in A} \left(U^t_m - \sum_{B \neq A} \gamma_{AB} Z_B \right) P^{tt}_m,$$

$$E^{(1)}_{Coul} = \sum_A \sum_{t_m \in A} (t_m t_m | t_m t_m)^A \Gamma^{tt}_m$$

$$+ 2 \sum_A \sum_{\substack{t_{m_1}, t_{m_2} \in A \\ m_1 < m_2}} [2(t_{m_1} t_{m_1} | t'_{m_2} t'_{m_2})^A$$

$$- (t_{m_1} t'_{m_2} | t'_{m_2} t_{m_1})^A] P^{tt}_{m_1} P^{t't'}_{m_2},$$

$$E_{res} = -4 \sum_m \beta^{R_m L_m}_{r_m l_m} P^{rl}_m,$$

$$E^{(2)}_{Coul} = 4 \sum_{A < B} \gamma_{AB} \sum_{\substack{t_{m_1} \in A \\ t'_{m_2} \in B \\ m_1 \neq m_2}} P^{tt}_{m_1} P^{t't'}_{m_2} + \sum_m \gamma_{R_m L_m} \Gamma^{rl}_m, \tag{12}$$

where we use the notation R_m for the right- and L_m for the left-end atoms, respectively, of the mth bond and introduce the intrabond matrix elements of one- and two-electron density matrices:

$$P^{tt'}_m = \langle 0 | g_m t^+_{m\sigma} t'_{m\sigma} g^+_m | 0 \rangle,$$

$$P^{rr}_m = u^2_m + w^2_m, \quad P^{ll}_m = v^2_m + w^2_m,$$

$$P^{rl}_m = P^{lr}_m = (u_m + v_m) w_m,$$

$$\Gamma^{tt'}_m = \langle 0 | g_m t^+_{m\beta} t'^+_{m\alpha} t'_{m\alpha} t_{m\beta} g^+_m | 0 \rangle,$$

$$\Gamma^{rr}_m = u^2_m, \quad \Gamma^{ll}_m = v^2_m, \quad \Gamma^{rl}_m = \Gamma^{lr}_m = w^2_m. \tag{13}$$

The nuclear–nuclear (core–core) repulsion in the MINDO/3 approximation differs from the pure Coulomb repulsion. It has the form

$$E^{nn} = \frac{1}{2} \sum_{\alpha \neq \beta} C_{\alpha\beta}, \tag{14}$$

where α and β are the indices of the respective nuclei and the quantities $C_{\alpha\beta}$ are defined by the expression

$$C_{\alpha\beta} = Z_\alpha Z_\beta (\gamma_{\alpha\beta} + D_{\alpha\beta}) \tag{15}$$

with

$$\frac{D_{\alpha\beta}}{\gamma_{\alpha\beta}} \rightarrow 0, \text{ when } R_{\alpha\beta} \rightarrow \infty,$$

$$D_{\alpha\beta} \rightarrow \infty, \text{ when } R_{\alpha\beta} \rightarrow 0. \tag{16}$$

The standard MM potentials [7, 11] usually implicitly include the Coulomb interaction of some charges residing on the atoms. The Mulliken atomic charges in the framework of the APSLG approximation are simply given by

$$Q_A = 2 \sum_{t_m \in A} P^{tt}_m - Z_A. \tag{17}$$

The total energy can be written as a sum of contributions with obvious classification:

$$E = \sum_m E_m + \sum_A \sum_{k < m \in A} E^A_{km} + E_{nonbond} + E_{Coul}. \tag{18}$$

This expression contains the intrabond contributions E_m, the contributions from the interaction between the vicinal chemical bonds at one atom A (E^A_{km}), the core interaction between nonbonded atoms, and the Coulomb interaction of atomic charges. Eq. (18) thus resembles the MM-type form of the potential energy. The explicit form of the contributions is the following:

$$E_m = \sum_{t \in \{r, l\}} [2U^t_m P^{tt}_m + (t_m t_m | t_m t_m)^{T_m} \Gamma^{tt}_m]$$

$$+ 2\gamma_{R_m L_m} [\Gamma^{rl}_m - 2P^{rr}_m P^{ll}_m] - 4\beta^{R_m L_m}_{r_m l_m} P^{rl}_m;$$

$$E^A_{km} = \delta_{AT_k} \delta_{T_k T_m} [4(t_k t_k | t'_m t'_m)^{T_k}$$

$$- 2(t_k t'_m | t_k t'_m)^{T_k}] P^{tt}_k P^{t't'}_m;$$

$$E_{nonbond} = \frac{1}{2} \sum_{\alpha \neq \beta} Z_\alpha Z_\beta D_{\alpha\beta};$$

$$E_{\text{Coul}} = \frac{1}{2} \sum_{\alpha \neq \beta} Q_\alpha Q_\beta \gamma_{\alpha\beta}. \tag{19}$$

At the same time this resemblance is relative because the contributions of Eq. (19) to Eq. (18) are not explicit functions on geometry parameters but depend on the bond geminal amplitudes u_m, v_m, w_m through the matrix elements of the density matrices and on matrices h^A through the molecular integrals. In the framework of the APSLG approach [16], these quantities are determined by iterative optimization procedure with alternating diagonalizations and multidimensional direct minimizations. At the same time, our previous calculations [16] have shown the bond amplitudes to be close for the identical bonds in the similar environment. The hybridization is also close for the atoms of definite type in the definite environment (e.g., carbon atoms in ethane, butane, and isobutane molecules have similar parameters of hybridization). The quantities u_m, v_m, and w_m and h^A determined variationally completely characterize the molecular electronic structure in the APSLG approximation. These quantities can be termed as electronic structure parameters (ESPs). An approximate transferability of these quantities can be confirmed by numerical estimates performed in the present and in our previous studies on the semiempirical version of the APSLG approximation. This allows to construct parametric expressions for the energy. Two classes of approaches can be constructed. In the first, the zero approximation transferable ESPs can be taken and inserted into Eq. (19). In the second class of approaches, the zero approximation ESPs are tuned by some one-step (noniterative) procedure to accommodate the variations of the ESPs under geometry variations. In the next section we consider some procedures to be used for noniterative ESP tuning.

2.2. NONITERATIVE TUNING OF ESPs

2.2.1. Bond (Geminal) Amplitudes

The optimal values of u_m, v_m, and z_m are the solutions of the eigenvalue problem:

$$\begin{pmatrix} A_m & 0 & D_m \\ 0 & B_m & D_m \\ D_m & D_m & C_m \end{pmatrix} \begin{pmatrix} u_m \\ v_m \\ z_m \end{pmatrix} = E^m \begin{pmatrix} u_m \\ v_m \\ z_m \end{pmatrix}, \tag{20}$$

where w_m is replaced by $z_m = \sqrt{2}w_m$, which simplifies the normalization condition for the geminals:

$$u_m^2 + v_m^2 + z_m^2 = 1, \tag{21}$$

corresponding to the lowest eigenvalue. The 3×3 matrix here is that of the effective two-electron Hamiltonian for the mth bond. The quantities A_m, B_m, C_m, and D_m are

$$A_m = 2\left(U_m^r - \sum_{B \neq R_m} \gamma_{R_m B} Z_B\right) + (r_m r_m | r_m r_m)^{R_m}$$
$$+ \sum_{\substack{t_{m_1} \in R_m \\ m_1 \neq m}} [4(r_m r_m | t_{m_1} t_{m_1})^{R_m}$$
$$- 2(r_m t_{m_1} | r_m t_{m_1})^{R_m}] P_{m_1}^{tt}$$
$$+ 4 \sum_{B \neq R_m} \gamma_{R_m B} \sum_{\substack{t_{m_1} \in B \\ m_1 \neq m}} P_{m_1}^{tt};$$

$$B_m = 2\left(U_m^l - \sum_{B \neq L_m} \gamma_{L_m B} Z_B\right) + (l_m l_m | l_m l_m)^{L_m}$$
$$+ \sum_{\substack{t_{m_1} \in L_m \\ m_1 \neq m}} [4(l_m l_m | t_{m_1} t_{m_1})^{L_m}$$
$$- 2(l_m t_{m_1} | l_m t_{m_1})^{L_m}] P_{m_1}^{tt}$$
$$+ 4 \sum_{B \neq L_m} \gamma_{L_m B} \sum_{\substack{t_{m_1} \in B \\ m_1 \neq m}} P_{m_1}^{tt};$$

$$C_m = \frac{1}{2}(A_m + B_m) + \gamma_{R_m L_m}$$
$$- \frac{1}{2} \sum_{t_m = r_m l r_m} (t_m t_m | t_m t_m)^{T_m};$$
$$D_m = -\sqrt{2}\,\beta_{r_m l_m}^{R_m L_m}. \tag{22}$$

As mentioned above, the geminal amplitudes are remarkably transferable in a series of similar compounds. The quality of this approximation can be further improved by adjusting the zero approximation values of the geminal amplitudes. To do so one can notice that all the matrix elements Eq. (22) of the effective Hamiltonian Eq. (20) depend only on the diagonal elements of the one-electron density matrix $P_{m_1}^{rr}$ and $P_{m_1}^{ll}$. Keeping in mind the normalization condition for the diagonal one-electron densities:

$$P_m^{rr} + P_m^{ll} = 1, \tag{23}$$

we (following Ref. [21]) introduce for each single chemical bond the polarity parameter

$$q_m = P_m^{rr} - P_m^{ll} = u_m^2 - v_m^2. \tag{24}$$

Then, the diagonal densities can be expressed through these parameters:

$$P_m^{rr} = \frac{1 + q_m}{2}, \quad P_m^{ll} = \frac{1 - q_m}{2}. \qquad (25)$$

The calculation of A_m, B_m, and C_m for the mth geminal requires the information on polarities of all other bonds (geminals). These polarities are defined by the transferable values of the zero approximation amplitudes u_m, v_m, and w_m for all the chemical bonds. Also, intuitively transparent values of the effective atomic charges can be conveniently used. They also can be expressed through the bond polarities (see above):

$$Q_A = \sum_{t_m \in A} [1 + (\delta_{tr} - \delta_{tl})q_m] - Z_A. \qquad (26)$$

Using this notation A_m and B_m can be rewritten:

$$A_m = 2U_m^r + (r_m r_m | r_m r_m)^{R_m} + 2 \sum_{B \neq R_m} \gamma_{R_m B} Q_B$$

$$- 2\gamma_{R_m L_m} + \sum_{\substack{t_{m_1} \in R_m \\ m_1 \neq m}} [2(r_m r_m | t_{m_1} t_{m_1})^{R_m}$$

$$- (r_m t_{m_1} | r_m t_{m_1})^{R_m}][1 + (\delta_{tr} - \delta_{tl})q_{m_1}];$$

$$B_m = 2U_m^l + (l_m l_m | l_m l_m)^{L_m} + 2 \sum_{B \neq L_m} \gamma_{L_m B} Q_B$$

$$- 2\gamma_{R_m L_m} + \sum_{\substack{t_{m_1} \in L_m \\ m_1 \neq m}} [2(l_m l_m | t_{m_1} t_{m_1})^{L_m}$$

$$- (l_m t_{m_1} | l_m t_{m_1})^{L_m}][1 + (\delta_{tr} - \delta_{tl})q_{m_1}]. \qquad (27)$$

Analytic formulae for the roots of the corresponding cubic equation are used in calculation. The lowest of the three roots E_m must be taken. Then, the amplitudes of the geminal are given by the formulae

$$u_m = z_m \frac{D_m}{A_m - E_m},$$

$$v_m = z_m \frac{D_m}{B_m - E_m},$$

$$z_m = \frac{1}{\sqrt{1 + \left(\dfrac{D_m}{A_m - E_m}\right)^2 + \left(\dfrac{D_m}{B_m - E_m}\right)^2}}. \qquad (28)$$

These expressions represent explicit formulae for the amplitudes u_m, v_m, and w_m.

2.2.2. Hybridization Matrices

According to Ref. [22] each hybridization matrix can be represented by the product of six Jacobi rotation matrices acting in the 2-D subspaces of the whole 4-D space. Therefore, it depends on six angles ω_{sx}^A, ω_{sy}^A, ω_{sz}^A, ω_{xy}^A, ω_{xz}^A, and ω_{yz}^A. These six angles, however, play different roles: The first triple determines the hybridization itself and must be at least approximately transferable while three others rotate the set of the HOs attached to the atom A as a whole and cannot possess any transferability. It is congenial to the modern MM schemes to consider the six angles determining hybridization matrices as quantities that are determined by minimizing a simple functional:

$$E^A = \sum_{t_m \in A} [2U_m^t P_m^{tt} + (t_m t_m | t_m t_m)^A \Gamma_m^{tt} - 4\beta_{r_m l_m}^{R_m L_m} P_m^{rl}]$$

$$+ 2 \sum_{\substack{t_{m_1}, t_{m_2}' \in A \\ m_1 < m_2}} [2(t_{m_1} t_{m_1} | t_{m_2}' t_{m_2}')^A$$

$$- (t_{m_1} t_{m_2}' | t_{m_2}' t_{m_1})^A] P_{m_1}^{tt} P_{m_2}^{t't'}, \qquad (29)$$

which is a sum of the intraatomic energy and resonance energy of bonds incident to the atom A. This approach is implemented in this article. In the next section we consider a series of approximate schemes based on the energy expression Eq. (19) but employing different approximate procedures for the ESPs (geminal amplitudes and hybridization matrices).

2.3. PARAMETRIC FORMS FOR THE ENERGY

The important feature of the APSLG-MINDO/3 method [16] is a remarkable transferability of its relevant ESPs—the geminal amplitudes and hybridization angles. It distinguishes the APSLG method from others (based, e.g., on the SCF approximation where the corresponding ESPs—MO LCAO coefficients—are not transferable). The geminal amplitudes vary slightly when going from one molecule to another: For example, the quantities u_m, v_m, and w_m for the C—H bond in methane are 0.4897, 0.4178, and 0.5412, respectively, while in ethane the analogous quantities are 0.4805, 0.4272, and 0.5416. These amplitudes also remain approx-

imately the same in the case of the C—H bonds in significantly different environments (e.g., in methylamin molecule these amplitudes are 0.4814, 0.4223, and 0.5431). That degree of transferability is even more accurate than one could guess.

At this point we are in a position to formulate a generic MM scheme with certain QM foundation. Indeed, if we assume the expression Eq. (19) for the energy to be an exact QM one, then schemes fitting into a diffuse concept of MM can be suggested. The key point with them is to accept specific procedures for defining the ESPs entering Eq. (19).

Within such a setting the parameters of the method can be subdivided in two groups: The first comprises the parameters of the Hamiltonian of the underlying semiempirical QM approach (in our case those of the APSLG–MINDO/3 [16]). They do not require special adjusting; other parameters reflect the transferability of the electronic structure characteristics obtained by the APSLG approach and can be taken from calculations of simple molecules by the APSLG approach or specially adjusted.

Using such definition of the HO matrices we can define zero-order approximation to the energy Eq. (19) by the following procedure: The amplitudes u_m, v_m, w_m are chosen to be constant for each type of the bond and the angles $\omega_{sx}^A, \omega_{sy}^A, \omega_{sz}^A$, determining the hybridization, are also constant for each type of the atom (the type of atom can be determined in the way analogous to that of MM: It depends on the closest environment). In other words, the values of the hybridization angles define the type of atom in terms of its hybridization. The quantities $\omega_{xy}^A, \omega_{xz}^A, \omega_{yz}^A$ are chosen on the basis of purely geometric correspondence between directions of HOs and those of the chemical bonds. The above assumption of perfect transferability of ESPs u_m, v_m, w_m and $\omega_{sx}^A, \omega_{sy}^A, \omega_{sz}^A$ allows to represent the total energy as an explicit function of geometric structure. At the same time, this approximation may turn out to be crude because the transferability of the zero approximation ESPs is good only for similar geometric structure and the large distortions of the molecule can impair the quality of this simple model. In fact, the transferability is characteristic only for the interatomic separations close to the equilibrium bond lengths. If the bond becomes elongated, the ionic contributions diminish while the amplitude (w_m) of the homeopolar configuration tends to $1/\sqrt{2}$. In this case, the transferability of the zero approximation amplitudes is broken and they must be readjusted. Analogously, if the valence angles are far

from the equilibrium ones the degree of transferability of the hybridization angles is questionable as well. Incidentally, the quality of the approximation to the amplitudes u_m, v_m, w_m can be significantly improved without loss of the explicit character of the energy expression.

The general idea of constructing local parametric expressions based on the APSLG description of the molecular electronic structure can be verified by numerical estimates of the validity of different schemes for obtaining the ESPs. The results of previous sections allow to propose a range of schemes for estimating the total energy by Eq. (19) differing by approximations used for obtaining the ESPs:

1. The geminal amplitudes and hybridization angles ω are taken as perfectly transferable parameters and are not additionally tuned. This scheme can be termed as FAFO (fixed amplitudes and fixed orbitals).

2. The geminal amplitudes are additionally adjusted by using formulae Eq. (28) while the HOs remain fixed at their transferable values—the TAFO scheme (tuned amplitudes and fixed orbitals).

3. Only the HOs are adjusted by minimizing the functionals Eq. (29) for each heavy atom but the geminal amplitudes are taken equal to their transferable values—the FATO scheme.

4. The geminal amplitudes are adjusted first and the HOs are corrected for the adjusted values of the bond amplitudes—the TATO scheme.

5. This is analogous to the previous scheme but the order of two adjustment procedures is changed—the TOTA scheme.

3. Results and Discussion

In this article, we performed test calculations on some organic molecules with use of the above five schemes. As the first test, we applied the above schemes to constructing the energy profile for the process of variation of one of the C—H bond's length in the methane molecule. The transferable hybridization matrix is taken to correspond to the sp^3 hybridization and the zero approximation C—H bond amplitudes are taken from a calculation on the methane molecule at its experimental equilibrium geometry. When such a choice is made, all five schemes give in the equilibrium point the electronic structure (and other properties, e.g., heat of forma-
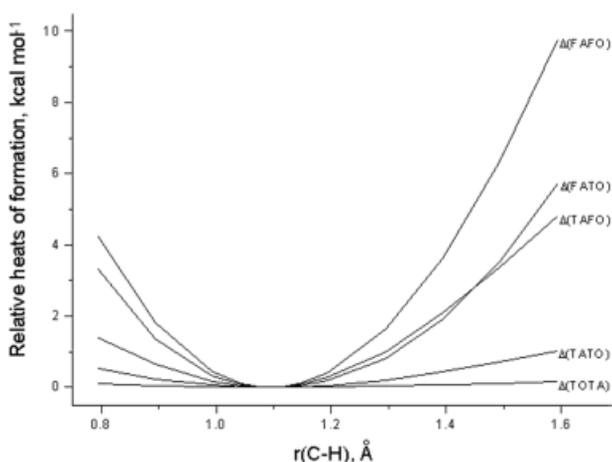
FIGURE 1. Relative heats of formation calculated by all parametric schemes for methane with one C—H bond elongated.

tion) coinciding with those obtained by the original APSLG approach. Figure 1 represents the differences between heats of formation calculated by each of the five above MM-type schemes and by the APSLG method as a function of the C—H bond length. This figure shows that all the five schemes work well in the vicinity of the equilibrium geometry. It can be concluded that both the minimum position and the heat of formation in the minimum remain the same as in the original QM APSLG approach [16]. Because the APSLG approach reproduces the experimental equilibrium geometry of methane with perfect accuracy, the MM-type schemes do the same as well. At the same time, the results of the zero-order approximation (FAFO) deviate from the precise APSLG one strongly for interatomic distances far from the equilibrium. This level of approximation can, however, satisfactorily cover the PES within a ± 0.1 Å range near the equilibrium position. It is important to find the quantitative measure of deviation between PES calculated by different computational methods. In our case, the difference between an MM-type PES and the APSLG PES must be characterized. As a convenient and representative measure in the 1-D case, we choose the area (integral) between two energy profiles in the definite interval of variation of the geometric parameter. Table I contains such data for three different ranges of variation of the C—H bond length in methane for all five schemes. The methods TAFO and FATO somewhat improve the results of the FAFO scheme but the range of their validity is not strongly extended as compared to that of the FAFO scheme. The methods combining the optimization of HOs with adjustment of the geminal amplitudes work satisfactorily for all interatomic distances studied. At the same time, the TOTA method yields the numerical values closer to those obtained by the underlying APSLG approach than the TATO scheme. The maximal deviation of the TOTA scheme from the exact APSLG one does not exceed 0.2 kcal/mol in the whole considered range of geometry parameters.

The above example is not representative enough because in the methane molecule the sp^3 hybridization can be a good approximation even for large geometry distortions. The hybridization can be significantly perturbed when a less symmetrical molecule is considered. To show how these schemes

TABLE I _____

Areas between energy profiles (kcal/mol Å or kcal/mol deg).

Variation range for a geometry parameter (Å or deg)	FAFO	TAFO	FATO	TATO	TOTA
Change of one C—H bond length in methane					
0.994–1.194	0.031	0.021	0.013	0.004	0.0006
0.794–1.594	2.119	1.241	1.075	0.246	0.038
0.794–3.094	73.973	17.618	56.527	2.121	0.207
Change of one O—H bond length in water					
0.880–1.060	3.131	2.676	0.047	0.405	0.037
0.800–1.200	7.904	6.732	0.229	1.112	0.113
Change of two O—H bond lengths in water					
0.880–1.060	3.039	2.835	0.028	0.409	0.006
0.800–1.200	9.325	7.978	0.273	1.230	0.014
Change of angle H—O—H in water					
101.0–108.0	115.23	98.48	0.32	13.67	0.30
97.0–112.0	246.64	211.00	0.81	29.29	0.66
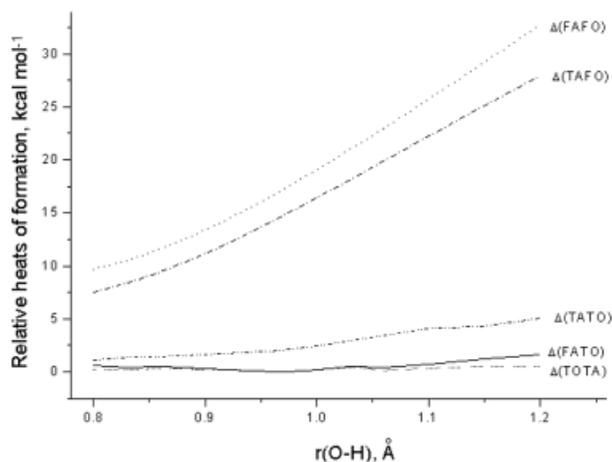
FIGURE 2. Relative heats of formation calculated by all parametric schemes for water with one O—H bond elongated.

FIGURE 3. Relative heats of formation calculated by all parametric schemes for water with both O—H bonds elongated.
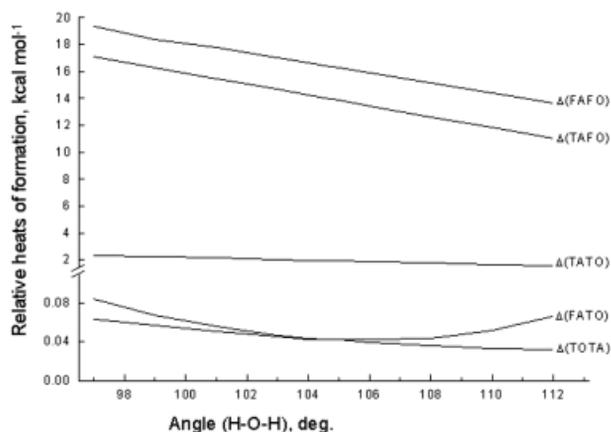
FIGURE 4. Relative heats of formation calculated by all parametric schemes for water with angle H—O—H distorted.

work under larger deviation of the hybridization from the one calculated by the APSLG at the equilibrium geometry, we consider as the next example distortions of the water molecule taken with the sp^3 hybridization for the oxygen atom as the zero approximation. The amplitudes u_m, v_m, and w_m of the O—H bond were taken from the APSLG–MINDO/3 calculation of the water molecule: 0.5946, 0.3317, and 0.5179, respectively. The distortions considered were (1) the variation of the length of one O—H bond, (2) simultaneous variation of the lengths of both O—H bonds, (3) variation of the valence angle H—O—H (these three cover all the possible distortions of the water molecule). Figures 2–4 represent for these processes the deviations of the heats of formation calculated by MM-type

schemes and the underlying QM APSLG method for these processes. Analogously, Table I contains the integral characteristics of these deviations.

These data show the importance of the quality of the zero approximation hybridization employed. The FAFO and TAFO methods result in large deviations of the heats of formation from the APSLG method even in the case of geometries close to the equilibrium. Also in the cases of variation of the lengths of O—H bonds the minimum position is significantly displaced, and in the case of the variation of the valence angle the shift of the minimum position is large. Better results are obtained with use of the FATO scheme. This scheme in this case works even better than the TATO scheme. At the same time the best of these schemes is the TOTA scheme. It gives results that are in perfect agreement with those of the underlying QM APSLG approach itself. The data of Figures 1–4 and Table I unambiguously demonstrate that only the TOTA scheme can be successfully and surely applied for description of the large portions of molecular PESs. Only this scheme is applied in our further analysis.

The molecules considered previously have only one type of bonds. Now we consider how the TOTA scheme works with different types of bonds present. Table II shows the results of calculations on the heats of formation for the ethane molecule as a function of the C—C interatomic separation. The zero approximation amplitudes u_m, v_m, and w_m for the C—H bond were taken from the APSLG calculation on the methane molecule in its equilibrium geometry and the analogous quantities for the C—C bond were taken from the calculation on the

TABLE II _____

Heats of formation of ethane for different values of the C—C bond length calculated by the APSLG and TOTA methods (kcal/mol).

r(C—C), Å	APSLG	TOTA	Δ
1.370	−8.456	−8.158	0.298
1.420	−15.585	−15.383	0.202
1.450	−17.902	−17.037	0.865
1.470	−18.739	−18.044	0.695
1.490	−19.065	−18.519	0.546
1.510	−18.923	−18.508	0.415
1.520	−18.689	−18.333	0.356
1.530	−18.354	−18.051	0.303
1.536	−18.106	−17.833	0.273
1.540	−17.922	−17.668	0.254
1.550	−17.397	−17.188	0.209
1.560	−16.785	−16.615	0.170
1.580	−15.312	−15.209	0.103
1.600	−13.538	−13.484	0.054
1.620	−11.493	−11.471	0.022
1.650	−7.977	−7.973	0.004
1.700	−1.138	−1.087	0.051

ethane molecule (0.4502, 0.4502, and 0.5453). The data in Table II show that the ethane molecule is well described by the TOTA scheme for all interatomic distances considered. Moreover, also for the large C—C distances the difference between the APSLG and TOTA heats of formation becomes small. The position of the minimum on the PES remains unchanged.

Table III contains the results of calculations on a series of organic molecules (for their experimental geometry parameters) by the QM APSLG–MINDO/3 method and the TOTA scheme for the two sets of zero approximation ESPs (geminal am-

plitudes). The first type of ESP contains no adjusted parameters. The geminal amplitudes for the C—H, C—C, O—H, and C—O bonds are taken from the APSLG calculations on the methane, ethane, water, and methanol molecules at their experimental geometry parameters. The sp^3 hybridization was taken as a zero approximation one for all heavy atoms. These results show that the TOTA scheme provides the heats of formation well corresponding to those of the APSLG approach. The deviation from the underlying APSLG scheme is significantly smaller than the precision of this scheme itself. It allows one to make a conclusion of a validity of the MM-type TOTA scheme for calculations of the heats of formation and equilibrium geometry structures of organic compounds with well-defined chemical bonds and lone pairs. It can be noted that this parameterization takes zero approximation ESPs from calculations of the simplest representatives of organic molecules. At the same time, these ESPs are somewhat different from those characteristic for larger molecules because the simplest molecules have somewhat specific structure in the class of homologues. Therefore, it can be reasonable to improve the results on the heats of formation by slight change of the parameters and, thus, the second procedure is an attempt to improve the results of calculations by changing the initial amplitudes for the C—H bond. They were taken as follows: 0.4739, 0.4306, and 0.5431. The change of the parameterization allows to improve the agreement between the heats of formation calculated by the APSLG and TOTA schemes. At the same time, this improvement is not significant and the initial geminal amplitudes can be left unadjusted.

TABLE III _____

Heats of formation of molecules calculated by the APSLG and TOTA methods with two sets of initial amplitudes for the C—H bond (kcal/mol).

Molecule	APSLG	TOTA(1)	Δ(1)	TOTA(2)	Δ(2)
CH_4	−8.414	−8.414	0.000	−8.397	0.017
H_2O	−55.648	−55.605	0.043	−55.605	0.043
C_2H_6	−18.106	−17.833	0.273	−17.844	0.262
C_3H_8	−23.438	−22.805	0.633	−22.869	0.569
C_4H_{10}	−26.951	−26.200	0.751	−26.295	0.656
C_5H_{12}	−32.093	−31.157	0.936	−31.283	0.855
Cyclopropane	18.520	18.877	0.357	18.806	0.286
CH_3OH	−47.677	−47.014	0.653	−47.001	0.666
C_2H_5OH	−59.112	−57.905	1.207	−57.997	1.115
C_3H_7OH	−63.811	−62.386	1.425	−62.520	1.291

4. Conclusion

The present work is aimed at constructing parametric expressions for the energy of molecules by starting from an underlying QM description of molecular electronic structure. Such a construct could serve as a generic form of MM. The APSLG trial wave function was proven to be the proper choice for underlying the desirable MM description. In such a way, a range of different MM-type schemes has been constructed for description of the total molecular energy as a function of molecular geometry. These generic MM schemes can be characterized as the QM APSLG–MINDO/3 method [16] accompanied by special procedures of selection of the relevant electronic structure parameters—geminal amplitudes and hybridization angles. Within such a setting, the parameters of generic MM schemes are (1) those of the underlying QM Hamiltonian and (2) the ESPs. The latter are calculated by the APSLG–MINDO/3 method for some simple molecules and are used further according to a procedure adopted. The deformations of methane and water molecules have been used to discriminate the quality of different schemes for selecting the ESPs. The TOTA scheme results in a reliable agreement with the underlying QM APSLG method. The performed calculations on the organic molecules of different classes confirm this conclusion. An attempt to change the initial electronic structure parameters leads to slight improvement of the results.

References

1. Dashevskii, V. G. Conformations of Organic Molecules; Khimiya: Moscow, 1974 (in Russian).
2. Burkert, U.; Allinger, N. L. Molecular Mechanics; ACS: Washington, DC, 1982.
3. Bernardi, F.; Olivucci, M.; Robb, M. A. J Am Chem Soc 1992, 114, 1606.
4. CECAM-NSF Meeting on QC/MM methods. Int J Quantum Chem 1996, 60(6) (special issue).
5. Allinger, N. L. J Am Chem Soc 1977, 99, 8127.
6. Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. J Am Chem Soc 1989, 111, 8551.
7. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
8. Jorgensen, W. L.; Tirado-Rives, J. J Am Chem Soc 1988, 110, 1657.
9. Pletnev, I. V.; Melnikov, V. L. Russ Chem Bull 1997, 7, 1278.
10. Comba, P.; Hambley, T. W. Molecular Modeling; VCH: New York, 1995.
11. Kozelka, J. In Metal Ions in Biological Systems, vol. 33; Sigel, A.; Sigel, H., Eds.; Marcel Dekker: New York, 1996, p. 1.
12. Tokmachev, A. M.; Tchougréeff, A. L.; Misurkin, I. A. J Mol Struct Theochem 2000, 506, 17.
13. Tokmachev, A. M.; Tchougréeff, A. L. Int J Quantum Chem 2001, 84, 99.
14. Surján, P. R. In The Concept of the Chemical Bond, Theoretical Models of Chemical Bonding, part 2; Maksić, Z. B., Ed.; Springer: Heidelberg, 1989, p. 205.
15. Wu, W.; McWeeny, R. J Chem Phys 1994, 101, 4826.
16. Tokmachev, A. M.; Tchougréeff, A. L. J Comput Chem 2001, 22, 752.
17. Tokmachev, A. M.; Tchougréeff, A. L. Russ J Phys Chem 1999, 73, 259.
18. Tokmachev, A. M.; Tchougréeff, A. L.; Misurkin, I. A. Russ J Phys Chem 2000, 74(2), S205.
19. McWeeny, R. Methods of Molecular Quantum Mechanics, 2nd ed.; AP: London, 1992.
20. Náray-Szabó, G. Acta Chim Acad Sci Hung 1976, 40, 261.
21. Parks, J. M.; Parr, R. G. J Chem Phys 1958, 28, 335.
22. Smirnov, V. I. Course of Higher Mathematics, vol. 3, part I; Moscow: Glz FizMatLit, 1958 (in Russian).