Methods

Training Set

Level of QM Calculations. The QM calculations of molecular and dimer properties were performed with the core electrons frozen (except for Li^+, Na^+, Ca^{2+}, and Mg^{2+}), using GAMESS software. Molecular geometries were optimized at the MP2/TZ(-hp) level, which is the MP2/cc.pVTZ level with the highest orbital momentum functions removed (d-functions for hydrogens and f-functions for heavy atoms). The geometries of optimized dimer conformations also were found with this basis set, the monomers being considered as rigid molecules during the optimization. In all other calculations (dimer energies, molecular dipoles, quadrupole, and polarizability tensors) the higher-level MP2/aTZ(-hp) was used. Our choice of the QM level is a compromise between accuracy and numerical efficiency of the calculations. In contrast to QMPFF1, in the present version QMPFF2 dimer calculations were performed with the counterpoising of the basis set superposition error and the intermolecular bond functions added, which allowed essentially more accurate evaluations of the intermolecular potential energies, especially its dispersion component in the case of nonpolar molecules. The electrostatic (ES) component of dimer energy was calculated by direct averaging of the intermolecular potential by using MP2/aTZ(-hp) density matrixes of the unperturbed monomers. Exchange (EX) terms was calculated by using HF/aTZ(-hp) level and then corrected by multiplying by the ratio of MP2 and HF density overlaps. The dispersion component was approximated as the difference between the QM dimer energy and the sum of the QM electrostatics, QM EX, and QMPFF induction terms fitted to reproduce QM polarizability tensors.

Generation of Dimer Conformation Training Sets. The training set for each representative dimer is composed in a usual manner of several conformational sets. Each set starts from some locally optimized dimer geometry, and then the distance between the

monomers is varied while retaining the fixed orientation of the monomers. Similar procedure of generation of the training set was used for parameterization of e.g., TTM water model (1).

Typical energy dependence on the distance between the monomers is illustrated in Fig. 6 by example of the water homodimer conformation sets used (along with other dimers) in parameterization of the atom/bond types found in water molecule.

Atom/Bond Type Classification

Table 4 presents the QMPFF2 atom type classification scheme for atom types found in proteins. Table 5 shows some additional atom types. There are a total of 63 atom types (43 heavy atoms and 20 hydrogen atoms). The bond types are determined by the types of bonded atoms, aromaticity attribute, and bond multiplicity; the total number of bond types is 94.

The Functional Form

The total potential energy of a molecule or molecular complex in QMPFF2 is given by:

$$
U^{TOTAL}(\mathbf{r},\mathbf{R}) = \sum_{ab} \left\{ U_{ab}^{ES}(\mathbf{r}_a,\mathbf{r}_b;\mathbf{R}_a,\mathbf{R}_b) + U_{ab}^{EX}(r_{ab}) + U_{ab}^{DS}(R_{ab}) \right\} + \sum_{a} U_{a}^{IN}(\mathbf{r}_a;\{\mathbf{R}\}_a)
$$
$$
+ \sum_{\{ab\}} U_{ab}^{ST}(R_{ab}) + \sum_{\{abc\}} U_{abc}^{BN}(\theta_{abc}) + \sum_{\{abcd\}} U_{abcd}^{TR}(\varphi_{abcd}) \qquad , \qquad \textbf{[1]}
$$

The first line represents nonbonded terms. The vectors $\mathbf{R}_a$ and $\mathbf{r}_a$ define, respectively, the positions of the atomic core and the center of electron cloud for atom a in the reference coordinate system; $r_{ab} \equiv |\mathbf{r}_b - \mathbf{r}_a|$, $R_{ab} \equiv |\mathbf{R}_b - \mathbf{R}_a|$. The four interaction terms, ES, EX, induction (IN), and dispersion (DS), are summed over interacting atoms assuming the 1-3 rule with the terms being dropped for atom pairs separated by one or two chemical bonds (along with the interaction of the clouds with their nuclear cores). For 1-4 interactions, a special renormalization is used (see below).

The second line of the relation (**1**) represents the potential energy caused by valence interactions as a sum of commonly used terms for stretching, U^{ST}, bend deformation U^{BN} of valence angles θ_{abc}, and torsion strain U^{TR} of chemically bonded atom quartets characterized by torsion angle φ_{abcd}. In Eq. **1** the sign {.} indicates summation over bonded atom groups.

The functional form of each component is presented below.

Nonbonded Interactions

The functional form of the components of intermolecular energy is closely linked to the model of atomic charge density, which is approximated as a superposition of the positive atomic core point charge (representing the atomic nucleus and inner electrons) and the negatively charged exponential electron density (the electron cloud), their positions, **R** and **r**, being the QMPFF dynamic variables.

Electrostatic interaction is derived by using an explicit expression for the charge density of atom a:

$$\rho_a(\mathbf{r}) = \tilde{Z}_a \delta(\mathbf{r} - \mathbf{R}_a) + \frac{Q_a}{8\pi \, \tilde{w}_a^3} \exp\left(- |\mathbf{r} - \mathbf{R}_a| / \tilde{w}_a\right), \quad [2]$$

where $\tilde{Z}_a$ is the atom core charge, Q_a is the charge of the electron cloud, and $\tilde{w}_a$ is the scale parameter of the cloud of atom type a. Here and below the QMPFF2 parameters are marked by tilde signs, with the subscript(s) referring to atom types rather than atoms. Note that the electron cloud charge Q_a is not a parameter but depends on the atoms bonded to atom a, through the following formula based on the bond charge transfer parameters, $\tilde{Q}_{ab}$:

$$Q_a = -\tilde{Z}_a + \sum_b \tilde{Q}_{ba}, \quad [3]$$

where the summation is over all atoms bonded to atom a; calculating the ES term for 1-4 interactions the sum over bonded atoms is dropped in this equation for pairs 1-2 and 3-4.

In accordance with Eq. **2**, ES interaction between two atoms, a and b, includes core-core, core-cloud, and cloud-cloud terms:

$$U_{ab}^{ES} = \tilde{Z}_a \tilde{Z}_b \varphi(R_{ab};0,0) + Q_a Q_b \varphi(R_{ab};\tilde{w}_a,\tilde{w}_b)$$
$$+ \tilde{Z}_a Q_b \varphi(R_{ab};0,\tilde{w}_b) + Q_a \tilde{Z}_b \varphi(R_{ab};\tilde{w}_a,0), \quad [4]$$

all expressed through the potential $\varphi(R;w_1,w_2)$ of electrostatic interaction of two exponentially distributed charge densities with scale parameters w_1 and w_2 separated by a distance R; the analytical expression for φ when w_1 and w_2 are nonzero can be found for instance in ref. 2:

$$\varphi(r;w_1,w_2) = \frac{1}{r}\left[1 - f(w_1,w_2)e^{-r/w_1} - f(w_2,w_1)e^{-r/w_2}\right], \quad [5]$$

where

$$f(w_1,w_2) = \frac{w_1^4(3w_2^2 - w_1^2)}{(w_2^2 - w_1^2)^3} + \frac{w_1^3 r}{2(w_2^2 - w_1^2)^2}.$$

In the case of interactions of point and diffuse charges or of two diffuse charges with the same value of the scale parameter, we have respectively:

$$\varphi(R;w,0) = \frac{1}{R}\left\{1 - \left[1 + \frac{1}{2}\left(\frac{R}{w}\right)\right]e^{-R/w}\right\}$$

and

$$\varphi(r;w,w) = \frac{1}{r}\left\{1 - \left[1 + \frac{11}{16}\left(\frac{r}{w}\right) + \frac{3}{16}\left(\frac{r}{w}\right)^2 + \frac{1}{48}\left(\frac{r}{w}\right)^3\right]e^{-r/w}\right\}.$$

EX repulsion is known to decay exponentially in the asymptotic region. Hence, the EX interaction of two atoms, *a* and *b*, is represented as:

$$U_{ab}^{EX} = \tilde{C}_a^{EX}\tilde{C}_b^{EX}\,\chi(R_{ab};\tilde{v}_a,\tilde{v}_b) \quad [6]$$

where

$$\chi(R;v_1,v_2) = \frac{v_1 v_2}{v_2 - v_1}\frac{e^{-R/v_2} - e^{-R/v_1}}{R}. \quad [7]$$

In the case of $v_1 = v_2 = v$ this formula reduces to $\chi(R;v,v) = e^{-R/v}$.

The QMPFF parameters $\tilde{v}_a$ are associated with the sizes of the exponential distributions that effectively manifest themselves in Pauli repulsion; these parameters generally differ from the scale parameters appearing in Eqs. **2** and **4**. The parameters $\tilde{C}_{a1}$, $\tilde{C}_{a2}$, are the EX charges characterizing the intensity of the EX interaction.

Dispersion interaction is simulated in QMPFF2 by superposition of Tang-Toennies (3) R^{-6} and R^{-8} terms:

$$U_{ab}^{DS} = -\tilde{C}_a^{DS6}\tilde{C}_b^{DS6}\psi_6(R_{ab};\tilde{u}_a,\tilde{u}_b) - \tilde{C}_a^{DS8}\tilde{C}_b^{DS8}\psi_8(R_{ab};\tilde{u}_a,\tilde{u}_b) \quad [8]$$

where

$$\psi_n(R; u_1, u_2) = R^{-n}\left[1 - e^{-\frac{2R}{u_1 + u_2}} \sum_{k=1}^{n} \frac{1}{k!}\left(\frac{2R}{u_1 + u_2}\right)^k\right]. \quad [9]$$

Induction in QMPFF2 is based on the functional form deduced from consideration of the two-level atomic system with the electron wave function approximated as a superposition of s- and p-states. The restraint potential is written as:

$$U_a^{IN}(\mathbf{r}_a; \mathbf{R}) = \frac{Q_a^2 \tilde{t}_{max}^2}{\tilde{\alpha}_a} \frac{1 + \sum_{b \in B_a} \mathfrak{s}_{ab}(\mathbf{n}_a \mathbf{n}_{ab})^2}{1 + \sum_{b \in B_a} \mathfrak{s}_{ab}/3}\left(1 - \sqrt{1 - \tau_a^2}\right), \quad [10]$$

where $\boldsymbol{\tau}_a = \left(\mathbf{r}_a - \mathbf{R}_a - \tilde{\mathbf{t}}_a^0 - \sum_{b \in B_a} \tilde{t}_{ab}\mathbf{n}_{ab}\right)/\tilde{t}_{max}$, $\mathbf{n}_{ab} \equiv \mathbf{R}_{ab}/R_{ab}$ is the unit vector directed from the core of atom a to that of atom b, and $\mathbf{n}_a \equiv \boldsymbol{\tau}_a/\tau_a$; with summation being performed over the set B_a of atoms bonded to atom a. The QMPFF parameter $\tilde{\alpha}_a$ is associated with the atomic polarizability; $\tilde{t}_{max}$ is the maximum value of the cloud shift and is independent of atom type. The parameters $\mathfrak{s}_{ab}$ and $\tilde{t}_{ab}$ depending on bond type ab characterize, respectively, the anisotropy of the restraint potential and the tendency of shifting of the electron cloud from atom a toward atom b (note that generally $\mathfrak{s}_{ab} \neq \mathfrak{s}_{ba}$ and $\tilde{t}_{ab} \neq \tilde{t}_{ba}$).

At small values of τ_a, U^{IN} is close to the oscillator potential, whereas for large values it approaches a constant, but its derivative (i.e., force) becomes infinite. Therefore for any external field τ_a is always <1, that eliminates the polarization catastrophe and provides a universal existence of a physically meaningful solution of the optimization problem for the cloud positions.

Bonded Interactions

QMPFF2 uses the simplest quadratic form for the energy potential related to deviation of valence bond length and valence angles from their equilibriums, whereas the threefold cosine potential is used to represent the molecular energy from torsion deformations:

$$U_{\{ab\}}^{ST}(R_{ab}) = \frac{\tilde{k}_{ab}}{2}\left(R_{ab} - \tilde{R}_{ab}^{0}\right)^2, \quad [11]$$

$$U_{\{abc\}}^{BN}(\theta) = \frac{\tilde{k}_{abc}}{2}\left(\theta - \tilde{\theta}_{abc}^{0}\right)^2, \quad [12]$$

$$U_{\{abcd\}}^{TR}(\varphi) = \frac{1}{2}\left(\tilde{V}_{\{abcd\}}^{(1)}(1+\cos\varphi) + \tilde{V}_{\{abcd\}}^{(2)}(1-\cos 2\varphi) + \tilde{V}_{\{abcd\}}^{(3)}(1+\cos 3\varphi)\right), \quad [13]$$

where $\tilde{k}_{ab}$, $\tilde{k}_{abc}$ and $\tilde{V}_{\{abcd\}}^{(1,2,3)}$ are the QMPFF2 parameters defined solely by fitting to QM data at the level of MP2(frozen core)/TZ(-hp). The parameters are fitted to minimize the deviation of total potential energy U^{TOTAL}, Eq. **1**, of a molecule in a vacuum from the *ab initio* calculated energy at specially deformed molecular geometries, with the QMPFF2 parameters for nonvalence interactions frozen during the fitting process.

For QMPFF2 nonbonded interactions, each atom type requires eight parameters and each general bond type requires five parameters (two parameters for symmetric bonds). For the bonded interactions, each bond length, bond angle, and torsion angle function uses two, two, and three parameters, respectively, which is comparable with most empirical force fields. The atom and bond types in a water molecule use a total of 18 nonbonded and 4 bond parameters.

MD Simulations

MD simulations were performed with an isothermal–isobaric NPT ensemble consisting of 256 flexible water molecules in a cubic box under periodic boundary conditions. The path integral discretization index P for quantum MD was generally chosen to be equal to

4. Certainly simulations with $P = 4$ do not provide the convergence with respect to internal molecular motions (e.g., bond stretching). However, this level is adequate to describe quantum effects in intermolecular motions for which we are the most interested in this article. A few calculations were performed with simulation time up to 5 ns and a P value of up to 20.

Details of QMPFF2 Implementation in MD. QMPFF implementation in MD is based on an adiabatic approximation typical for polarizable systems, according to which the electronic degrees of freedom are considered to be fast in comparison with motion of the nuclei. Thus, in every MD step the cloud positions are first optimized at fixed atomic cores locations, and then the forces acting on the cores are calculated and atoms move to the new positions.

Pressure is evaluated as

$$P = \frac{2}{3V}(E_{kin} + W),$$

where W is the virial. For QMPFF it is written as:

$$W = \tfrac{1}{2}\sum_{AB}\{\mathbf{F}_{AB}\mathbf{R}_{AB} + \mathbf{F}_{Ab}\mathbf{R}_{Ab} + \mathbf{F}_{aB}\mathbf{R}_{aB} + \mathbf{F}_{ab}\mathbf{R}_{ab}\} + \sum_{a}\mathbf{F}_{a}^{IN}\mathbf{R}_{a} + \sum_{A}\mathbf{F}_{A}^{IN}\mathbf{R}_{A}.$$

Here $\mathbf{F}_{xy}$ is the force on object x due to object y, indexes A and B are related to atomic cores, and a and b numerate the electron clouds. Pair terms in this equation include nuclei-nuclei interaction and nuclei-cloud and cloud-cloud contribution; the last two terms relate to polarization:

$$\mathbf{F}_{A}^{IN} = -\partial U^{IN}/\partial \mathbf{R}_{A}, \mathbf{F}_{a}^{IN} = -\partial U^{IN}/\partial \mathbf{R}_{a}.$$

Details of Long-Range Interactions Treatment in MD Simulations. Reaction field formalism is used in MD simulations with QMPFF, in which the formula for the modified Coulomb term is written as (4):

$$U_{RF} = \varphi(R_{ij})\left[1 + \frac{\varepsilon_{RF} - 1}{2\varepsilon_{RF} + 1}\frac{R_{ij}^3}{R_{cutoff}^3}\right].$$

φ is given by Eq. **5**, R_{cutoff} is the cut-off radius, and the dielectric constant ε_{RF} satisfies equation:

$$\frac{\varepsilon_{RF} - 1}{\varepsilon_{RF} + 2} = \frac{4\pi}{3}\alpha_{eff}n_0,$$

where $\alpha_{eff} = \alpha_O + 2\alpha_H$, α_O, and α_H are the atomic polarizabilities of oxygen and hydrogen, respectively, and $n_0 = N_{H_2O}/V$ is the current average density of water, N_{H_2O} and V being the total number of molecules in the system and the current cell volume, respectively.

In addition to long-range electrostatics effects, it is also necessary to take into account the long-range corrections to the total energy and pressure caused by truncation of Van der Waals forces (or dispersion term in QMPFF2). The asymptotic behavior of dispersion term at long distances can be written as $U^{DS} \sim -C_6/r^6 - C_8/r^8$, so the long-range energy correction can be represented in the form:

$$\Delta U_{LR} = -\frac{2\pi}{3}\frac{1}{VR_{cutoff}^3}\left(\langle C_6 \rangle + \frac{3}{5}\frac{\langle C_8 \rangle}{R_{cutoff}^2}\right), \quad [14]$$

where $\langle C_6 \rangle = \sum_{i \neq j} N_i N_j C_{6ij}$ and $\langle C_8 \rangle = \sum_{i \neq j} N_i N_j C_{8ij}$ are the interaction constants summed over atom types i, j. Correction to the pressure is derived as:

$$\Delta P_{LR} = -\frac{4\pi}{3} \frac{1}{V^2 R_{cutoff}^3} \left(\langle C_6 \rangle + \frac{4}{5} \frac{\langle C_8 \rangle}{R_{cutoff}^2} \right). \quad [\textbf{15}]$$

1. Burnham, C. J. & Xantheas, S. S. (2002) *J. Chem. Phys.* **116**, 1479-1492.

2. Slater, J. C. (1963) *Quantum Theory of Molecules and Solids* (McGraw-Hill, New York), Vol. 1.

3. Tang, K. T. & Toennies, J. P. (1984) *J. Chem. Phys.* **80,** 3726-3741.

4. Neumann, M. & Steinhauser, O. (1980) *Mol. Phys.* **39,** 437-454.