# Molecular Mechanics: Principles, History, and Current Status

Valeri Poltev

*The ultimate justification for the many severe approximations and assumptions made in the present work comes from the fact that the agreement between the simple calculations and the available experimental data is as good as it is.N. L. Allinger, J. Amer. Chem. Soc.,* ***81****, 5727, 1959*

## Contents

Introduction ..... 2
Foundations and General Scheme of Molecular Mechanics. Atoms as Elementary Units of the Matter ..... 3
General Expression and Terms of Molecular System Energy ..... 4
Intramolecular Contributions to Molecular System Energy ..... 5
Intermolecular and Nonbonded Intramolecular Interactions ..... 6
General Remarks on Molecular Mechanics, Its Accuracy, and Applicability ..... 7
A Bit of History: The "Precomputer" and Early Computer-Aided MM Calculations ..... 10
First MM Applications to Three-Dimensional Structure and Thermodynamics of Organic Molecules ..... 10
The Role of Molecular Crystal Study on the First Steps of Molecular Mechanics ..... 12
Molecular Mechanics on the First Steps of Molecular Biology: Molecular Mechanics and Protein Physics ..... 13
Molecular Mechanics on the First Steps of the Biophysics of Nucleic Acids ..... 18
The Problems and Doubts of Further Development of the MM Approach ..... 21
Approaches to Improvement of MM Formulae and Parameters ..... 21
Various Schemes of Water Molecule in MM Calculations ..... 24
Modern Molecular Mechanics Force Fields: Description, Recent Development, and Applications ..... 28
Allinger's Force Fields and Programs ..... 28
Merck Molecular Force Field (MMFF94) ..... 29
The Force Fields and Programs Designed by Scheraga and Coauthors ..... 30
Force Fields and Programs Developed by Kollman and Coauthors. AMBER ..... 32
Other Popular Force Fields and MM Software. CHARMM, OPLS, GROMOS ..... 36

V. Poltev (✉)
Autonomous University of Puebla, Puebla, Mexico
e-mail: poltev@fcfm.buap.mx vpoltev@yahoo.com



J. Leszczynski et al. (eds.), *Handbook of Computational Chemistry*,
DOI 10.1007/978-94-007-6169-8_9-2

Conclusions..... 39
Bibliography..... 40

**Abstract**

A short survey of the general principles and selected applications of molecular mechanics (MM) is presented. The origin of molecular mechanics and its evolution is followed starting from "pre-computer" and the first computer-aided estimations of the structure and potential energy of simple molecular systems to the modern force fields and software for the computations of large biomolecules and their complexes. Analysis of the current state of physicochemical study of biological processes suggests that MM simulations based on empirical force fields have an ever-increasing impact on understanding the structure and functions of biomolecules. The problem of "classic mechanics" description of essentially quantum properties and processes is considered. Various approaches to a selection of force field mathematical expressions and parameters are reviewed. The relation between MM simplicity and "physical nature" of the properties and events is examined. Quantum chemistry contributions to MM description of complex molecular systems and MM contribution to quantum mechanics computations of such systems are considered.

## Introduction

Molecular mechanics (MM) deals with the classical ("mechanistic," i.e., via Newtonian mechanics) description of molecular and supramolecular systems. The simplified assumptions and approximations enable one to use MM for wide applications to various systems, starting from simple low-molecular-weight molecules (such as numerous hydrocarbons) to large biomolecular complexes (such as those of proteins, nucleic acids, and membrane fragments) or material assemblies of many thousands of atoms. During a period a bit longer than a half of a century, the number of publications using this approach roused by many orders of magnitude. It was rather easy to follow nearly all the works related to MM calculations in the 1960s and in the early 1970s of the last century (and the author of this chapter tried to do so); today it is very difficult to follow all the papers related to the MM approach even in the rather restricted area of application (e.g., to biopolymer structure and function). Thus it is impossible to consider all the applications of MM in a short survey, and the choice of the material for this chapter is "a matter of taste" of the author who refers readers to other monographs and chapters for a more general description of the modern state and/or selected aspects of extended applications of MM; for general references see, e.g., monographs (Lewars 2011; Allinger 2010; Ramachandran et al. 2008; Cramer 2004; Leach 2001). We will follow the evolution of MM from the first "precomputer" and early computer-aided (i.e., before the era of personal computers) works to modern complex simulations impossible without supercomputers.

The MM method is used now not only in theoretical and computational works but also as a part of experimental studies (e.g., many X-ray and NMR-derived structures of proteins, nucleic acids, and their complexes with other molecules deposited in Protein Data Bank (Berman et al. 2003) and Nucleic Acid Data Bank (Berman et al. 1992) are the results of MM refinements). The MM-like semiempirical terms are widely used now in some quantum mechanics-based studies. The systematic inclusion of such terms started from the addition of the dispersion term to the standard density functional theory computations, for creating the so-called DFT-D method (see, e.g., Grimme 2004; Antony and Grimme 2006; Jurecka et al. 2007, and references herein for the first DFT-D results). Recent developments and applications of different DFT-D methods (DFT-D2, DFT-D3) demonstrated their predictive power for biomolecular systems (see, e.g., Grimme 2011; Brandenburg and Grimme 2014). Since the beginning of this millennium, MM-like terms are frequently included into so-called enhanced semiempirical quantum mechanics methods (see, e.g., Yilmazer and Korth (2015) for a recent review). Both dispersion (D) and hydrogen-bond (H) correction terms are used in such methods usually referred to as SQM-DH methods.

The rapid extension of MM simulations is a result of development of new branches of natural science related to molecular biology with its numerous applications, to directed synthesis of new substances with desirable properties (including new drugs and new materials for the industry), and many other areas of so-called nanoscience. The MM simulations enable one to perform a preliminary selection of the compounds with desirable properties before experimental testing, and even before chemical synthesis itself. Sometimes it is impossible to rationalize completely experimental results without the construction of atom-level mechanistic model. In many cases, the model cannot be derived from experimental data only, e.g., the accuracy of X-ray diffraction patterns for biopolymer fragments is not sufficient for the extraction of precise atom coordinates due to a complexity of the system and inherent irregularities of atom positions (e.g., delocalization of water molecules). In other cases, the complex molecular system cannot be characterized by a single three-dimensional structure but by a set of conformations. MM simulations enable one to construct habitual atom-level molecular models of sufficiently complex systems favorable from energy viewpoint. These models can be visualized by computer graphics programs and are ready to use for further investigations and refinements (e.g., by quantum mechanics methods or by new experimental approaches).

## Foundations and General Scheme of Molecular Mechanics. Atoms as Elementary Units of the Matter

The modern MM can be considered as an extension of simple atom and bond representation of molecules and their complexes in classical chemistry. In the beginning of the 1950s, scaled paper images and hand-made primitive models of atoms and bonds from wood and wire enabled subtle scientists to construct the

first successful models of proteins and DNA. The pioneering papers of Nobel Prize authors can be mentioned in this relation (Pauling et al. 1951; Watson and Crick 1953). Stick, ball-and-stick, and space-filling plastic models as well as their computer images are widely used until now for both teaching and investigation.

## General Expression and Terms of Molecular System Energy

The MM approach is a generalization and quantitative representation of these models via mathematical expressions. A set of such expressions together with numerical coefficients is usually referred to as the force field; the MM method itself is frequently referred to as the force field method. Rigid balls of computer graphics or plastic (metal, wood, etc.) models are replaced by "soft" ones, and hard sticks are replaced by springs. "Hard Sphere" approximation can be considered as simplified (in many cases oversimplified) MM approach. Nevertheless, very important results were obtained via this approximation in the middle of the last century; we will mention some of them below.

The modern MM method and most of force fields mathematical expressions are the results of the efforts of many researchers for nearly half a century, trials and errors of both well-known and widely cited authors, as well as of those who contributed to specific parts of nonbonded energy computations for a selected class of molecules. In this section, we will describe general scheme and the main common assumptions of the MM method leaving any references to specific papers and contributions of the researchers for next sections. The main subject of this survey is rather simple MM scheme suggesting that the systems (molecules or their complexes) consist of atoms, each atom being simulated as a single point particle. Most computations for biologically important molecules use the so-called Class 1 model. More complicated schemes considering additional centers of interactions, additional terms, and more exact expressions (so-called Class 2 and Class 3 MM models) will be mentioned briefly. Molecular mechanics describes molecules in terms of "bonded atoms" (atoms in molecule), their positions are distorted from some idealized (equilibrium) geometry due to nonbonded interactions with the atoms of the same molecule and of other molecules. The MM approach implicitly uses the starting approximation for both classical and quantum computational chemistry schemes, namely, Born–Oppenheimer approximation. MM does not include electrons explicitly, but different configurations of electron shells of the same chemical element's atoms are implicitly considered in force field parameters. Some authors consider not only atoms but lone pairs of electrons, chemical bonds, and other points in a molecule as centers of interactions. We will mention such "extension" of Class 1 models below.

The first principal approximation of MM method is the additivity for several levels of calculations, namely, the additivity of energy terms responsible for the contributions of different physical nature, and the additivity of contributions of

the atoms. The simple ("minimalist") representation of the potential energy of a molecule or molecular complex as a function of atom coordinates, R, can be represented by a sum of four main terms each term being a sum of many contributions (Eq. 1).

$$\Delta E(\mathbf{R}) = \sum \mathrm{Eb} + \sum \mathrm{Ea} + \sum \mathrm{Et} + \sum \mathrm{Enb} \tag{1}$$

The summing up in the terms is over all the chemical bonds (the first term), valence angles (the second term), torsion angles (the third term), and all pairs of atoms nonbonded to each other or to common third atom (the last term). The energy terms depend on mutual positions of atoms and on the adjustable constants (parameters); the parameters are suggested to be transferable between atoms and molecules of the same type. The transferability of the force field parameters is the second important assumption of MM, and the number of atom types of the same chemical element depends on particular force field.

## Intramolecular Contributions to Molecular System Energy

The first and the second terms (sums of Eq. 1) refer to the energy changes due to variations of bond lengths, stretching (Eq. 2), and bond angles, bending (Eq. 3). These terms are usually modeled as harmonic potentials centered on equilibrium values of bond lengths and bond angles, respectively, i.e., "simple Hooke's law" dependences are used.

$$\mathrm{Eb_i} = \mathrm{kb_i}\left(\mathrm{l_i} - \mathrm{l_i}^0\right)^2 \tag{2}$$

$$\mathrm{Ea_i} = \mathrm{ka_i}\left(\alpha_\mathrm{i} - \alpha_\mathrm{i}^0\right)^2 \tag{3}$$

In these equations, $\mathrm{l_i}$ and $\alpha_\mathrm{i}$, the current values of bond lengths and angles; parameters $\mathrm{l_i}^0$ and $\alpha_\mathrm{i}^0$, equilibrium values for bond lengths and angles; $\mathrm{kb_i}$ and $\mathrm{ka_i}$, stretch and bend force constants, respectively. These adjustable parameters depend on types of atoms forming the bond or the valence angle. Some force fields may also contain cubic and higher-order contributions to these terms, or sometimes more flexible Morse potential can be used instead, as well as "cross terms" can be included to account for correlations between stretch and bend components. In the latter case, the terms depending on both bond length and angle variations are added. The third sum of the expression (Eq. 1) refers to the changes of torsion energy; it is responsible for interactions of electron shells of two atoms A and D (and of two bonds A–B and C–D) which are connected through an intermediate chemical bond B–C.

$$\mathrm{Et_i} = \mathrm{kt_i}\left(1 + \cos\left(\mathrm{n_i}\phi_\mathrm{i} - \delta_\mathrm{i}\right)\right) \tag{4}$$

This periodic function (integer $n_i$ being the periodicity) contains the current value $\phi_i$ of the angle i of the rotation around B–C chemical bond (the angle between the two planes defined by atoms A, B, and C and by B, C, and D) and three parameters ($kt_i$, $n_i$, and $\delta_i$; $n_i$ is multiplicity and $\delta_i$ is the phase angle) for each type of torsion (for each combination of four neighbor atoms of the molecule). These parameters can be estimated from experimental data on the structure and properties of the molecule considered and of related molecules, then they (as well as $l_i^0$, $\alpha_i^0$, $kb_i$, and $ka_i$ of Eqs. 2 and 3) should be adjusted by trial computations. Many force fields include terms responsible for "improper" torsions or out-of-plane bending, i.e., terms related to four atoms not forming consecutive chemical bonds, which function as correction factors for out-of-plane deviations (for example, they can be used to keep aromatic rings planar). These terms can be expressed via harmonic potentials like those for bond stretching and valence angle bending. Cross terms depending on both torsion angle and bond length or valence angle are added in some force fields. Three terms mentioned above are positive for practically all states of all the molecular systems (or equal to zero when the term corresponds to equilibrium state for contributed bond, angle or torsion).

## Intermolecular and Nonbonded Intramolecular Interactions

The last sum of Eq. 1 refers to so-called nonbonded interactions, Enb, of all the atom pairs not bonding to each other or to the same third atom, Eq. 5. Each atom-atom term is usually represented by a sum of electrostatic, Coulomb (the first term of Eq. 5), and van der Waals (the second and the third terms of Eq. 5) interactions.

$$E_{ij}\left(r_{ij}\right) = Kq_iq_j/r_{ij} - A_{ij}/r_{ij}^{\,6} + B_{ij}/r_{ij}^{\,12} \tag{5}$$

This equation contains $r_{ij}$, the current distance between i and j atoms; $q_i$ and $q_j$, effective atom charges; $A_{ij}$ and $B_{ij}$, adjustable parameters responsible for dispersion (London) attraction and short-range repulsion interactions respectively. The atomic charges are usually derived using calculations via various quantum chemistry methods, effective dielectric constant implicitly accounting for surrounding can be used (this value may be distant dependent). The $A_{ij}$ and $B_{ij}$ coefficient can be preliminarily estimated via equilibrium inter-atomic distance and energy values at equilibrium for neutral pairs of atoms ($\rho_{ij}$ and $\varepsilon_{ij}$ respectively) and followed by the adjustment to reference experimental data. Most of the early force fields used for description of van der Waals interactions are Buckingham (6-exp) potential instead of Lennard-Jones (6-12) potential in Eq. 5. The total expression for nonbonded interaction term are usually referred to as (1-6-12) or (1-6-exp) potential relating to the dependency of the terms on the inter-atom distance. The Buckingham potential is more flexible (it has three adjustable parameters instead of two for 6-12 potentials for each atom pair type) and has more physics justifications for the range of experimentally observed distances (due to exponential dependence of electron wave functions on the distance from nuclei), but it is less convenient for computations. It

has a maximum at short distance and then trend to negative infinite value. Majority of the modern force fields utilize 6-12 expressions for description of van der Waals interactions, the total atom-atom potential being referred to as 1-6-12 one. Some force fields substitute 6-12 potential by 10-12 one for the interactions of hydrogen atoms of hydrogen bonds in order to describe more sharp distance dependence in the most important area of energy minimum corresponding to H-bond formation (referred to as 1-10-12 potential). More complex expressions (including those dependent on the angles between two straight lines connecting three atoms of H-bond) were used for H-bond description in some early potential sets. The nonbonded terms of the intramolecular energy related to 1-4 interactions (i.e., the interactions between atoms in a molecule separated by three chemical bonds) are frequently accounted for with a coefficient less than 1 (1-4 scaling) as these interactions are already included into torsion term (Eq. 4). To reduce the number of adjustable $A_{ij}$ and $B_{ij}$ parameters of Lennard-Jones potential (and corresponding parameters of other potentials), the combination rules for equilibrium inter-atom distance, $\rho$, and minimal energy values, $\varepsilon$, for pairs of different atoms are usually applied.

$$\rho_{ij} = \rho_i + \rho_j;\ \varepsilon_{ij} = \left(\varepsilon_i \varepsilon_j\right)^{1/2} \tag{6}$$

Some force fields apply the combination rules directly to the coefficients of van der Waals terms.

The calculations of potential energy via Eq. 1 are used to search for local energy minima (mutual atom positions corresponding to possible stable configurations), to construct and analyze multidimensional potential energy surfaces (PES), to follow trajectory of movement (in MD, molecular dynamics simulations), or to study averaged thermodynamic and geometry characteristics (via MC, Monte Carlo sampling) of the systems. Description of MD and MC methods is beyond the survey, while some results of these applications of MM method will be mentioned briefly.

## General Remarks on Molecular Mechanics, Its Accuracy, and Applicability

The first computer (and all "precomputer") applications of mechanistic approach to molecule conformations and interactions ignored certain energy terms (e.g., stretching, bending, torsion, or electrostatic ones). Some modern works can ignore certain terms as well in order to reduce the number of variables of energy function, e.g., considering bond lengths as constants (their changes in many cases are very small and have no influence on energy and geometry of minimal energy structures). The simplest one of such approaches considers bond lengths and valence angles as constants, ignores torsion energy (the contributions of the first three terms of Eq. 1 being zeros), and utilizes "hard sphere approach" to nonbonded interactions. This approach is a mathematical representation of plastic (or wood, iron, etc.) space filling mechanic models or their computer images. The configurations are forbidden

when any two nonbonded atoms are closer to each other than a sum of van der Waals radii (these configurations have infinite positive energy), all other being allowed (with zero energy). Already this oversimplified approach enables one to obtain some important results, e.g., to reject certain configurations and even a possibility to synthesize the molecule with inevitable too close positions of nonbonded atoms. The first "Ramachandran maps" for proteins (which will be discussed in the next section) have demonstrated allowed and forbidden regions on two-dimensional plots of the fragment of polypeptide chain. These maps were subsequently improved using more realistic MM functions or quantum mechanics calculations.

Most modern MM computations include additional terms besides those already mentioned. These terms refer to direct imposition of experimental data (e.g., NMR-derived restrains on inter-atom distances or global characteristics of the macromolecule) and to description of specific quantum effects not accounting for by standard MM force field formulae.

The complexity of mathematical expressions and the number of parameters depend on the systems considered. The problem of "which atoms pertain to the same type and which ones are of different types" is considered by the authors of specific force field and software depending on the tasks and computer resources. The atom type may depend not only on the chemical element and electron shell configuration but on neighbor atoms and on the structure of the whole molecular fragment (e.g., the carbons of six-member and five-member aromatic rings having the same three bonded atoms may be considered as pertaining to different atom types). The more broad the applications that are planned for the force field, the greater the number of atom types that should be involved and the more complex force field formulae that should be constructed. The first works which dealt with the tasks related to specific systems (e.g., the conformations of saturated hydrocarbons or peptide fragments) usually contained a few parameters; the modern force fields may contain thousands of parameters (in spite of use of combination rules mentioned above). The mathematical forms of the equations and numbers of atom types and adjustable parameters represent a compromise between the simplicity (for elaboration and for use) and the accuracy of the results obtained.

Various physical considerations can be used for preliminary estimation of mathematical expressions and parameter values (rather simplified considerations were used in Eqs. 2, 3, 4, and 5). It is important to emphasize that neither dependences nor values of parameters can be "derived" (directly calculated) from universal principles or measured by any experimental method. The stretch and bend constants (of Eqs. 2 and 3) can be evaluated using infrared spectra; equilibrium bond lengths and valence angles can be estimated from X-ray data for crystals of simple molecules. The $A_{ij}$ coefficients of the attraction part of van der Waals interactions can be evaluated (and really were calculated and used without refinement in the first MM works) via approximate formulae for dispersion interactions; however, their exact values for a certain class of systems should be adjusted by comparison with experimental data or with the most exact quantum chemistry results after trial computations for reference set of related systems. The same is valid for other terms and their parameters. Some parameters have rather simple physical meaning

and restricted areas of possible values (e.g., equilibrium distances between bonded and nonbonded atoms or barriers to rotation about the bonds); other parameters have only approximate relation to physical values (atom charges, $B_{ij}$ coefficients of Lennard-Jones potential). As all the parameters are adjustable ones, only the values of total energy and the equilibrium geometry of the molecular system can be compared with experimental data, and consequently have the strict physical meaning, not the individual contributions or the values of the individual parameters. As various force fields utilize different reference sets of data, the individual parameters are not transferable between different force fields even in cases where they use the same mathematical expressions. Different force fields may result in the nearly equal energy and geometry of local minima configurations but rather different values of the individual term contributions. Thus individual terms of the energy may have very approximate physical interpretation, although in some cases it is interesting to evaluate the certain energy contributions and to follow their changes for different molecular complexes and different configurations (and many researchers include these evaluations to their publications).

It is worth mentioning that preliminary consideration of MM scheme has resulted already in some doubts and objections. Generally speaking, the classical description of the essentially quantum molecular systems cannot be exact and full. Most of the terms in Eqs. 2, 3, 4, and 5 refer to the first approximation or to the first term of expansion of the corresponding interaction energy. The atoms are not points, they have dipole and quadrupole moments (not only charges), and charge distribution in a molecule is continuous. The polarization or electron delocalization interactions are not considered in the classical "minimalist" MM approach, and the contributions of three-body and four-body interactions can be essential ones. Many attempts have been undertaken to overcome these inherent difficulties of the MM method as well as to justify the assumptions and simplifications; we will consider some of these attempts below. Few remarks for justification of the main principles of MM method are described here.

The possibility of consideration of atoms as elementary subunits of the molecular systems is a consequence of Born-Oppenheimer or adiabatic approximation ("separation" of electron and nuclear movements); all the quantum chemistry approaches start from this assumption. The additivity (or linear combination) is a common approach to a construction of complex functions for a physics-based description of the systems of various levels of complexity (cf. orbital approximation, MO LCAO approximation, basis sets of wave functions, and some other approximations in quantum mechanics). The final justification of the method is good correlation of the results of its applications with the available experimental data and a possibility to predict the characteristics of molecular systems before experimental data become available. It can be achieved after careful parameter adjustment and proper use of the force field in the area of its validity. The contributions not considered explicitly in the force field formulae are included implicitly into parameter values of the energy terms considered.

## A Bit of History: The "Precomputer" and Early Computer-Aided MM Calculations

The quantitative estimations of molecular properties via simple atom level mechanics representations originate from the communications of Hill (1946, 1948), Westheimer and Mayer (1946), and Barton (1948, 1950). All these papers refer to conformations of organic molecules. It is interesting to mention that mathematical expression of potential energy suggested in the pioneering work of Hill (1946) contains common for all the modern force fields stretch and bend components (Eqs. 2 and 3 of previous section) as well as the Lennard-Jones terms of nonbonded interaction energy. Westheimer and Mayer (1946) suggested the use of exponential term for description of steric repulsion. The first calculations for selected conformations of rather simple ("medium size") molecules, such as diphenil derivatives in paper (Westheimer and Mayer 1946), cis-decalin and steroids in the papers of Barton (1948, 1950) were performed manually or using desk calculators. Some researchers constructed "hand-made" models of steel or wood (e.g., of cyclic saturated hydrocarbons in the papers of Allinger (1959)) for careful measurements of geometry parameters. The importance of quantitative estimations of nonbonded interactions for considerations of three-dimensional structure of organic molecules was emphasized starting from the first mechanical considerations, which was clearly shown by Bartell in 1960. He illustrated the preference of "soft sphere" over "hard sphere" approach to the analysis of hydrocarbon structure and suggested one of the first (6-12) parameters for hydrocarbons (Bartell 1960). As mentioned above, rather approximate calculations clearly demonstrated the utility of MM approach to the problems of organic chemistry as well as the need for further extensive computations and searching for more reliable parameters. We will refer to all these quantitative theoretical considerations of molecular properties as MM, not depending on the use of this term by the authors or on the method of estimation of different types of interactions.

Rapid expansion of MM method starting from the 1960s was provoked by an introduction of computers into all the branches of natural science. In this section, we will briefly consider some examples of the first computer-aided applications of MM method to three research areas, namely, the physical organic chemistry (these works can be considered as a continuation of the "precomputer" papers mentioned above), the structure and properties of molecular crystals, and the interactions and conformations of biopolymers.

## First MM Applications to Three-Dimensional Structure and Thermodynamics of Organic Molecules

The first paper on computer study of organic molecule conformations was related to saturated hydrocarbons by Hendrickson (Hendrickson 1961). The angle bending, torsion, and (6-exp) van der Waals contributions to the conformation energy were

taken into account, while constant values were assigned to bond lengths. The computer calculations of cyclo-alkanes containing 5, 6, and 7 carbon atoms enabled him to consider various conformers and to reproduce and rationalize the experimental data. During the 1960s and the beginning of the 1970s such computations were performed by several groups of investigators. Hendrickson (1962, 1973, and references herein), Allinger et al. (1971), and Allinger and Sprague (1973) extended MM approach to more complex hydrocarbons including those with delocalized electronic systems. Allinger's computations took into account bond stretching in addition to terms used by Hendrickson. Wiberg (1965) was the first researcher who suggested use of the Cartesian coordinates of atoms (instead of "internal coordinates," i.e., bond lengths, bond angles, torsions) for nonbonded ("strain") energy minimization. Such approach enables researchers to use the same scheme and the same software to study completely different systems. Most modern MM programs use such an approach. The electrostatic term has not been included in the papers mentioned in this paragraph as hydrocarbons are nonpolar molecules; it was introduced later when more broad sets of molecules became to be considered.

The most important results of early MM computations of organic compounds can be illustrated by Engler et al.'s paper (1973) titled "Critical Evaluation of Molecular Mechanics." The calculations for various hydrocarbons have been performed using two rather different force fields, their own and that of Allinger et al. (1971). The two force fields have substantially different parameters as different sets of experimental characteristics were used for parameter adjustment. It results in significant difference of separate terms of the energy, which may vary by several times or even be positive for one force field and negative for another one. Nevertheless, total energies and relative values for various conformations calculated by two force fields are rather close to each other for most of the compounds considered. Two phrases of the paper's abstract can be considered as characteristics of the whole situation with the MM approach for that time: "Most of the available data are reproduced with an accuracy rivaling that achieved by the experimental methods" and "The molecular mechanics method, in principle, must be considered to be competitive with experimental determination of the structures and enthalpies of molecules."

Since the abovementioned publications, the theoretical conformational analysis using MM force fields has become an inherent part of physical studies of organic molecules. It became clear that in spite of difficulties with the parameter choice, the MM calculations are not only a useful tool to rationalize experimental observations, to reproduce and to predict the structure and energy characteristics of "medium size" organic molecules with experimental accuracy, but they can also help in elaboration of pathways of chemical synthesis. Different mathematical expressions and parameters transform into *force fields*, i.e., to complete and verified sets of the formulae and constants and ready-to-use computer programs. One of the first such force fields and programs was MM1 program based on the paper by Allinger and Sprague (1973), which then was followed by MM2, MM3, and MM4 force fields and software for broad class of organic molecules. The inclusion of new atom types and new terms of energy as well as parameter refinement continued till last years,

and we will discuss these program sets in the next sections together with other modern force fields and software.

## The Role of Molecular Crystal Study on the First Steps of Molecular Mechanics

The first publications on MM approach to molecule conformations mentioned above deal with *intramolecular* interactions. The first works on such approach to *intermolecular* interactions belong to one of the pioneers and founders of MM, A. I. Kitaigorodsky (other spellings, Kitaigorodskii, Kitajgorodskij, Kitaygorodsky), though he (and all the researchers at that time) did not use the term "Molecular Mechanics." Nearly at the same period (the 1950s) as Hill and Westheimer, usually mentioned as initiators of MM approach, he suggested to consider mechanical models of molecular systems quantitatively via mathematical expressions. In the 1950s, he foresaw the applications of the method to various problems of physics and physical chemistry of organic and biological molecules, including the problems of crystal structure, adsorption, and conformational transitions. We refer to the most frequently cited works of Kitaigorodsky (1960, 1961, 1973). Kitaigorodsky used structure and thermodynamic data on the crystals of organic molecules for adjustment of parameters of atom-atom potential functions for calculations of nonbonded interaction energy. This term of the total energy is dominant for molecular crystals. The method of atom-atom potentials was suggested as a generalization of the "principle of close packing" of molecules in molecular crystals he discovered earlier; his book (Kitaigorodski 1959) describing this principle is still widely cited. Later he demonstrated that this principle is a consequence of more general atom-atom approach to potential energy of molecular crystals. Kitaigorodsky was the first researcher who suggested to consider the interactions of nonbonded atoms in a molecule and of such atoms of different molecules via the same mathematical expressions and parameters (applications of MM approach to biopolymers would be impossible without this suggestion). The studies of molecular crystals via MM approach enabled one to derive potential functions for nonbonded interactions and to test the fidelity and accuracy of the approach itself using extensive sets of available quantitative data on structure and energy of molecular crystals.

Some energy terms were ignored in the first MM works on conformations of organic molecules, but it was impossible to predict a priori what types of contributions to intramolecular interactions energy (sums in Eq. 1) can be reasonably neglected for specific types of molecules. Considering crystal structure of nearly rigid molecules (e.g., aromatic rings) it is possible to ignore (at least in the first approximation) all the terms but the last one in Eq. 1. This assumption is just a consequence of examination of the geometry of series of related molecules in crystals (e.g., aromatic hydrocarbons consisting six-atom rings); all of them have nearly equal values of bond lengths and angles not depending on molecular complexity and the type of crystal packing. For hydrocarbons, it is possible (again at least for the initial studies) to ignore electrostatic contributions as well (the

molecules have no dipole moment, and all the estimations of atom charges result in small values, less than 0.1 of electron charge). Kitaigorodsky suggested step-by-step selection of potential functions starting from molecules of nearly neutral atoms of two types (hydrocarbons) and introducing next atom types one-by-one for selecting next parameters. His pioneering works on molecular crystals were followed by more extensive computations of other authors. Some of these works were inspired by his ideas and followed his methodology (e.g., first papers of Williams (1966, 1967)), other researchers performed computations using different expressions for nonbonded potentials and additional terms of the energy (e.g., Lifson and Warshel 1968; Warshel and Lifson 1970). The common set of (6-exp) parameters for C . . . C, C . . . H, and H . . . H intermolecular interactions was obtained by Williams (1967) from energy and structure calculations for crystals of both aromatic and aliphatic hydrocarbons. Warshel and Lifson (1970) derived "consistent force field" parameters for description of both intermolecular and intramolecular interactions in crystals. This set contains both parameters for van der Waals interactions and other terms of energy mentioned in the previous section. Mason and coauthors (e.g., Craig et al. 1965; Rae and Mason 1968; Mason 1969) used combined approach to calculations of intermolecular interactions in crystals, namely, repulsive terms were calculated at atom-atom level, while attractive ones were considered on bond-bond level. Such approach and other changes of simple atom-atom scheme are not convenient for computations, and we would like to cite a phrase from Mason's paper comparing his and Kitaigorodsky's methods: "Although the representation of the attractive potential by spherically symmetric atom-atom interactions cannot be justified theoretically, it has been outstandingly successful in predicting the properties of crystals" (Mason 1969). The calculations for molecular crystals became a part of the selection and test of the parameters during elaboration of nearly all modern force fields. Later MM calculations for molecular crystals enable researchers to examine the approximations accepted in force field elaboration and to derive force field versions with additional terms of energy (including polarization and atom centers, e.g., Williams and Weller (1983)).

## Molecular Mechanics on the First Steps of Molecular Biology: Molecular Mechanics and Protein Physics

Development of molecular biology and biophysics in the 1960s required a quantitative consideration of the conformations and interactions of proteins, nucleic acids, and biomolecules in general. The problems of biological importance were to rationalize the structure and conformational properties of the proteins and nucleic acids, namely, to understand the contributions of the subunits and to construct the models of the most favorable and biologically important conformations of the fragments. We already mentioned in previous section that the first successful models of regular structures of proteins and DNA duplex have been constructed using "hand-made" fragments of paper, wood, wire in the beginning of the 1950s. For understanding, explanation, and prediction of molecular level mechanisms of

biopolymer functions, it was necessary to work out the method of quantitative simulation of the three-dimensional structure and properties of biomolecules. The first molecular mechanics considerations of biopolymer fragments were performed in 1960s. We will follow briefly such studies for proteins in a few paragraphs below.

The general problems on proteins which can be in principle solved via MM simulations are (1) the construction of three-dimensional structure of the macromolecule and prediction of the pathways of "protein folding" using restricted experimental data (ideally, the primary structure only); (2) the refinement of experimental structure (X-ray diffraction patterns usually do not supply us with information sufficient for precise atom coordinate assignment). The whole problem of the proteins functioning cannot be solved via MM only as chemical reactions are beyond the MM approach. Changes in electronic structure in processes that involve bond-breaking and bond-forming, charge transfer, and/or electronic excitation require quantum mechanics considerations. Nevertheless, MM approach is useful for the problems of enzyme-substrate complex formation and of molecular recognition which are crucial for the protein functions.

First of all, two important results on protein structure obtained before classical MM computations should be mentioned. The atom-level structures of α-helical and β-sheet fragments of polypeptide chains have been designed by Linus Pauling and Robert Corey in 1951 using "hand-made" models of wood (Pauling and Corey 1951; Pauling et al. 1951). Such regular structures are the intrinsic parts of the majority of the proteins. The success in the construction of these first models of the regular protein fragments (as well as of the DNA duplex) primarily depended on correct subunits geometry and a potential to predict the correct scheme of hydrogen-bond formation, all other contributions to intramolecular interactions being of secondary importance. The regular structures imply H-bond formation between N-H donors and C=O acceptors of the peptide groups of the same chain for α-helix, and of other chain or of distant parts of the same chain for β-sheet. As it was pointed in more than half of century (Eisenberg 2003), "In major respects, the Pauling-Corey-Branson models were astoundingly correct, including bond lengths that were not surpassed in accuracy for >40 years."

The second important "precomputer" result refers to construction via hard-sphere models of two-dimensional map for possible (and impossible) conformations of the dipeptide unit, the so-called Ramachandran plot (Ramachandran et al. 1963). The polypeptide backbone consists of repeating peptide group connected via $C_\alpha$ atoms, the latter being connected to hydrogen atom and amino-acid residue (the first atom of the residue is designated as $C_\beta$ atom). The peptide group is practically planar, and the only single bonds in the peptide chain are those between $C_\alpha$ atom and two neighbor peptide groups ($C_\alpha$–N–H of one group, the torsion angle for rotation about this bond is designated as Φ, and $C_\alpha$–C=O of the next group, the torsion angle for rotation about this bond is designated as Ψ). Using mathematical expressions for atom coordinates suggesting fixed bond lengths and bond angles, Ramachandran et al. (1963) have constructed the two-dimensional Φ-Ψ maps with two sets of van der Waals atom radii, "normally allowed" and "outer limit," i.e., normal and shortened radii. The main part of the maps for all the amino-acid

residues except glycyl corresponds to the conformations forbidden due to shorten atom-atom contacts, while α-helix and β-sheet fall into the allowed regions. For glycyl residue, the allowed regions were considerably more extended. It took about half a year to construct these maps using a desk calculator (Ramachandran 1990); later such maps were computationally constructed using various force fields and quantum mechanics methods; superposition of Φ-Ψ combinations corresponding to experimentally or computationally constructed three-dimensional structure of proteins and peptides on the Ramachandran plot helped to check and rationalize the protein models nearly for half a century. The consideration of dependence of Ramachandran plot on amino acid residue via hard sphere approach enables Leach et al. (1966a) to evaluate semi-quantitatively the parts of two-dimensional map corresponding to allowed conformations for different amino-acid residues (from about 50 % of all conceivable conformations for glycyl to only 5 % for valyl). The evaluation of steric restrictions emphasizes their important role as a determinant in protein conformation; the consideration of α-helices demonstrated that the preference of the right-handed ones in comparison to left-handed helices is due essentially to interactions of the $C_\beta$ atom of the side chains with atoms in adjacent peptide units of the backbone (Leach et al. 1966b). The "hard sphere" approach was applied to search for allowed conformations of cyclic oligopeptides, for example (Némethy and Scheraga 1965).

The main conclusion of the "hard sphere" works was that steric effects are one of the most important factors in determining polypeptide conformations. "Many conformations of a polypeptide can be classed as energetically unfavorable without consideration of other kinds of interactions; however, the method breaks down to the extent that it cannot discriminate between those conformations which are sterically allowed . . . The contribution that this method has made is that more than half of all the conceivable polypeptide conformations are now known to be ruled out by steric criteria alone" (Scott and Scheraga 1966a). Nearly at the same time (the middle of 1960s) as abovementioned hard sphere considerations of peptides, the first applications of MM formulae to protein structure were started using "soft atoms," and the first parameters of potential functions suitable for peptide subunits calculations were suggested. The first such works were published by three groups of researchers, namely, those of De Santis and Liquori (De Santis et al. 1965), Flory (Brant and Flory 1965), and Scheraga (Scott and Scheraga 1966a). All these works were preceded by the publications of the authors related to synthetic polymers. The bond lengths, valence angles, and planar peptide group in all the early studies of polypeptides were fixed.

De Santis et al. (1963) have carried out the calculations of van der Waals term of the energy using different potentials for regular conformations of linear polymers as functions of torsion angles of monomer units. The deepest minima of the conformational diagrams were found very near to the experimental structures, as obtained by X-ray fiber diffraction methods, for a series of polymers investigated including polyethylene, poly(tetrafluoroethylene), poly(oxy-methylene), and polyisobutylene. When the calculations were extended to the polypeptides (polyglycine, poly-L-alanine, and poly-L-proline), good agreement with the experiments was obtained

as well (De Santis et al. 1965). After nearly 30 years De Santis wrote "While other contributions to the conformational energy are included in the calculations, the dominant role of the van der Waals interactions remains well established as the main determinant of the conformational stability of macromolecules" (De Santis 1992).

Brant and Flory (1965) have carried out the calculations on peptide unit energy in which they included torsion potentials, van der Waals (6-exp) interactions, and dipole-dipole interactions between the permanent dipole moments of near-neighbor peptide groups. The conformational energies were used by Brant and Flory in statistical mechanical calculations of the mean-square unperturbed end-to-end distance of various polypeptide chains. They concluded that it was necessary to include the electrostatic interactions in the energy calculations to obtain agreement between the calculated and experimentally determined chain dimensions and thus that these interactions are important in determining polypeptide conformations.

The first publication of Scheraga's group related to "soft atoms" MM simulations related to proteins (on helical structures of polyglycine and poly-L-alanine (Scott and Scheraga 1966a)) included besides contributions considered in paper of Brant and Flory (1965) explicit accounting of hydrogen bonds and interactions of atom charges of peptide group. Rather complex expression for dipole-dipole energy of interactions between peptide groups used by Brant and Flory (1965) was substituted by Coulomb atom-atom terms, and (6-exp) expressions for van der Waals interactions were substituted in this paper and subsequent publications by (6-12) potentials, but additional complicated expression for description of hydrogen bond energy was added. The method and parameters for calculation of torsion and van der Waals energies were earlier developed by Scott and Scheraga (1966b) for normal hydrocarbons and polyethylene. It is interesting to note that both groups of authors (those of Flory and of Scheraga) applied without further refinement approximate Slater-Kirkwood equation for calculation of coefficient for attractive term of van der Waals energy (the coefficients were expressed via polarizabilities and effective numbers of electrons of the interacting atoms) and van der Waals radii for calculation of repulsion term parameters. Approximate character of such parameters is evident, but rather good agreement with experimental results was obtained in the both publications (Brant and Flory 1965; Scott and Scheraga 1966a). The existence of polyglycine and poly-L-alanine regular α-helices was rationalized as a consequence of favorable both H-bond and van der Waals interactions, while helices of other possible types (ω and $3_{10}$) lie in relatively high-energy regions, and for isolated helices these structures have been excluded (Scott and Scheraga 1966a).

The abovementioned calculations of Scheraga et al. on protein simulations were followed by a series of publications related to both improvements of the method (both parameter adjustments and new term additions) and of simplifications of computations. Such changes in the methods enable them to obtain preliminary results on peptides of up to 20 amino acid residues already in 1967 (Gibson and Scheraga 1967a, b). The improvements of the method refer to enlarging the equilibrium radii of atoms, assigning partial charges to each atom (but neglecting the electrostatic contribution when the atoms are not close together), and introduction of

orientation-dependent term for H-bond description instead of complicated formulae of the previous paper (Scott and Scheraga 1966a). The simplified account for solvent contribution to the energy has been included considering the effect of removing the nearest neighbor to peptide water molecules (Gibson and Scheraga 1967a). The simplification of the model refers to introducing "extended atom," i.e., CH, $CH_2$, and $CH_3$ groups were considered as the single atom to reduce the number of atom-atom interactions that had to be computed (this approach was used later in many force fields and such potential functions are commonly referred to as "united atom" potentials, we will discuss this simplification below). Scheraga's group prolongs the improvement of the parameters and the method continuously till recent years. We will discuss these studies below, now we will mention some improvements performed in the first half of the 1970s. The most important improvements refer to description of H-bond potential by (10-12) expression (McGuire et al. 1972) (some other force fields adopted such dependence for description of interaction between hydrogen atom capable to participate in H-bond and atom-acceptor of H-bond), to calculation of partial charges via semiempirical CNDO/2 method (Yan et al. 1970), and to adjustment of repulsion part of van der Waals interactions to crystal data for extensive set of molecules including amino acid crystals (Momany et al. 1974, 1975). The results of the crystal calculations led to an internally self-consistent set of interatomic potential energies for interactions between all types of atoms found in polypeptides (ECEPP force field (Momany et al. 1975)). The bond lengths and bond angles were obtained from a survey of the crystal structures and were considered as fixed for each of amino acid residue (Scheraga's force fields differ in this feature from most other widely used force fields). The preliminary set of potential functions they used to refine X-ray structures of some proteins, e.g., lysozyme (Warme and Scheraga 1974), α-lactalbumin (Rasse et al. 1974), and rubredoxin (Warme et al. 1974).

We will not consider details of these papers; instead we will mention the first publication on MM refinement of X-ray protein structure of Levitt and Lifson (1969). The refinements of protein structures (myoglobin and lysozyme) have been performed via two steps by minimization of the energy function containing all the MM terms and an additional term describing deviation of the atom coordinates from their values obtained from X-ray diffraction studies (penalty function). Three energy terms (corresponding bond stretching, bond angle bending, and torsion potentials) supplemented by penalty functions were used at the first step. At the second step nonbonded and hydrogen-bonded interactions were included, and penalty functions were omitted. The authors mention that energy function parameters are preliminary ones, intended for protein simulation in aqueous solution, differed significantly from those of Scheraga and subject to further improvement. The refine structures were close to X-ray diffraction ones and do not contain short atom-atom contacts or unusual bond lengths and valence angles. Starting from rough model coordinates, the method minimizes the assumed total potential energy of the protein molecule to give a refined conformation (Levitt and Lifson 1969).

## Molecular Mechanics on the First Steps of the Biophysics of Nucleic Acids

The problems related to MM approach to nucleic acid structure and functions differ in some aspects from those for proteins. The main computational tasks for the 1960s in the area of nucleic acids were to rationalize the structure of native and modified DNA duplexes, t-RNAs, synthetic polynucleotides, i.e., to understand the contribution of the subunits (the bases, sugars, ionized phosphate groups, counterions, and surrounding water) to three-dimensional structure, and to evaluate the contributions of interactions of different physical nature to structure and functions of nucleic acids. The extensive applications of MM approach to nucleic acids were started a few years after those for proteins. The same is true for the adjustment of MM parameters for calculations of nucleic acids. Such a situation can be explained by "special role" of DNA in cells as a heredity material as well as by expectations of "specific" forces between the bases as conjugated molecules with delocalized electron system. Many researchers considered the electron exchange via H-bonds in base pairs and exchange interactions between stacked bases as the crucial contributions to nucleic acid functioning. It took some years before pioneers of quantum mechanics approach to biochemical problems would write: "The possible contributions of resonance energy stabilization through electronic delocalization across the hydrogen bond for horizontal interactions and of overlapping of their π-electronic cloud or of charge-transfer complexation for the vertical interactions appear to be of much smaller order the magnitude than the stabilization increments due to van der Waals-London forces" (Pullman and Pullman 1968).

The author of this survey was the first researcher since the application of MM formulae in 1966 to consider the interactions of nucleic acid subunits. We will now briefly follow the route of investigations of nucleic acid interactions to "classic" MM approach. The first quantitative estimation of interactions of bases in DNA duplex via formulae of intermolecular interactions was performed by De Voe and Tinoco (1962). They used so-called molecule dipole approximation, i.e., permanent and induced electric point dipoles were placed into the center of each base. The energy of interaction between two paired or stacked bases was suggested to be a sum of the interactions of permanent dipoles (Coulomb electrostatic interactions), permanent dipole-induced dipole (polarization interactions), and of induced dipoles (London dispersion interaction). Such an approximation does not enable them to obtain even qualitatively correct results (e.g., the energy of interaction between the bases in Watson-Crick A:T pair was positive), but the computational problem of evaluation of interaction energy changes on the formation of the specific nucleic acid conformation has been defined in this chapter. That is why it is cited in hundreds of publications during nearly a half of century, many of them (including those of the author of this survey) were inspired by the paper of De Voe and Tinoco (1962). Shortly it became clear that such an approximation is invalid for this system not depending on dipole location as well as after replacement of point dipoles by "real" ones, i.e., by dipoles of definite size. Bradley et al. (1964)

suggested that the main contribution into base-base interaction energy is of Coulomb electrostatic nature, which can be calculated in "monopole approximation." The point charges were calculated via semiempirical methods of quantum chemistry and placed on each atom of the bases. This paper was followed by the papers of Nash and Bradley related to calculations of base-base interactions. One of their papers should be mentioned in relation to the general progress and development of MM approach. Starting from various mutual in-plane positions for all the combinations of RNA bases, the calculations and search for minima of electrostatic energy was performed (Nash and Bradley 1966). Van der Waals energy was taken into account in "hard sphere" approximation, i.e., any non-hydrogen atom pair should not be closer than sum of van der Waals radii. As a result of the calculations, 27 energy minima were obtained which have two or more short N-H$\cdots$O or N-H$\cdots$N contacts (with N$\ldots$N or N$\ldots$O distances about 3 Å) corresponding to nearly linear or bifurcated hydrogen bonds. The results enable Nash and Bradley to rationalize the available up-to-date experimental data on base crystals data as well as base-base complex formation in solutions and polynucleotides. It appears that the observed base pairings with two H-bonds in crystals always correspond to one of the computed geometries of lowest energy (Nash and Bradley 1966). It is important to emphasize that Bradley et al. (1964) "monopole" approach to calculations of the base interaction was the first attempt to consider the electrostatic energy in the framework of the modern MM scheme. The abovementioned calculations of Scheraga's group with inclusion of point charges of peptide group were started 2 years after the Bradley's publication.

The subsequent studies on the quantitative evaluation of the interactions and three-dimensional structure of nucleic acid fragments were continued via two routes. One of the approaches suggested construction of MM force field parametrized to nucleic acids interactions; another approach implied calculations using more rigorous physics concepts and molecular characteristics in a hope to understand physical nature of interactions. The first works on this way were performed by Pullman's group and referred to interactions of the bases in fixed positions (e.g., Pullman et al. 1966). Considering the energy of interactions as a sum of electrostatic, polarization, and dispersion terms, two approximations were applied, "dipole" and "monopole". The monopole approximation was applied to electrostatic energy only (following Bradley et al. 1964), while two other terms were really calculated in molecular dipole approximation. This approach was later augmented by inclusion of the short range repulsion term and representation of the molecule as a set of "many-centered multipole distributions obtained from ab initio SCF calculations (charges, dipoles, and quadrupoles located on the atoms and the middles of segments joining pairs of atoms)" (Langlet et al. 1981). We will not consider subsequent progress on this way of nucleic acid computations as new schemes (e.g., Gresh et al. 1986 and references herein) do not correspond to the MM approach. Nevertheless, it is worth to mention that the sophisticated schemes of base-base computations had no advantage for rationalization and prediction of experimental results in comparison to "standard" MM scheme reviewed in the next paragraph.

In the middle of the 1960s, the author of this survey proposed the first atom-atom scheme and numerical parameters for the calculation of interactions of nucleic acid bases (the terms "force field" and "molecular mechanics" were not widely used in those years). This scheme had three source points, namely, idea of De Voe and Tinoco (1962) on quantitative estimations of interactions of nucleic acid subunits, "monopole" approximation of Bradley et al. (1964) for electrostatic interactions of the bases, and Kitaygorodsky (1961) atom-atom approach to calculations of van der Waals interactions in molecular crystals. At that time, there was no "MM-like" calculation scheme suitable for quantitative considerations of nucleic acid interactions and structure. We refer here only to later papers summarizing some preliminary adjustments and applications of the approach (Poltev and Sukhorukov 1970; Poltev and Shulyupina 1986). When calculating the interactions of nucleic acid bases or other conjugated heterocyclic molecules the only term $\sum$Enb of Eq. 1 was considered, corresponding to all the pair-wise interactions of atoms not pertaining to the same molecule. The bond lengths and valence angles were fixed and corresponded to averaged experimental geometry for each of the bases (like those for amino-acid residues and peptide backbone in Scheraga's force field (Momany et al. 1975)). The atom-atom terms contained electrostatic, London attraction, and short-range repulsion contributions. This scheme was the first one assigning different parameters of van der Waals terms to the atoms of the same chemical element but different electron shell configuration (e.g., aromatic and aliphatic carbons, pyridine and pyrrol nitrogens, keto and enol oxygens). These parameters were preliminarily evaluated from approximate London formulae (for $A_{ij}$ coefficient of Eq. 5) and enlarged van der Waals radii of atoms (for $B_{ij}$ coefficient of Eq. 5 or corresponded values for the firstly used 6-exp potentials). The final values of the parameters were selected after step-by-step calculations and comparison with experimental data for crystals containing the atoms of some types only (starting from graphite, then aromatic hydrocarbons, pyrazine, benzoquinone, purine, and finally the bases). The effective charges of atoms were calculated via simple semiempirical quantum chemistry method (of Huckel and Del Re for $\pi$-electrons and $\sigma$-electrons, respectively) with the parameters adjusted to reproduce experimental values of dipole moments for related molecules. The primary scheme (Poltev and Sukhorukov 1970) contained polarization term as well, but it was shortly eliminated due to its small value for many cases and to avoid difficulties related to nonadditivity of atom-atom and molecule-molecule interactions. The parameter set was than amplified by consistent parameters for sugar-phosphate backbone (Zhurkin et al. 1980) and for water-DNA interactions (Poltev et al. 1984). This simple MM scheme specially adjusted to nucleic acid interactions enables us to rationalize extended set of experimental data and to forecast some nucleic acid properties before experimental evidence. Variability of DNA helix parameters (Khutorsky and Poltev 1976), dependence of mutual base-pair positions in energy minima on nucleotide sequence (Polozov et al. 1973), pathways for all the base-substitution errors (Poltev and Bruskov 1978), and DNA duplexes with mispairs (Chuprina and Poltev 1983) were predicted by such calculations before experimental data became available. The parameters of this simple scheme were refined later several times (like

any other force field scheme discussed above and below) via adjustment to new experimental data and quantum mechanics considerations. We will mention here only one more series of early MM calculations related to nucleic acids performed by Indian scientists Renugopalakrishnan, Lakshminarayanan, and Sasisekharan (coauthor of Ramachandran et al. (1963) paper on Φ-Ψ plot for proteins). They suggested the first complete set of parameters for calculations of the conformations of nucleic acid fragments (Renugopalakrishnan et al. 1971, and references herein). The atom charges crucial for the base interactions have been calculated via the same procedure as in our earlier papers (e.g., Poltev and Sukhorukov 1970).

## The Problems and Doubts of Further Development of the MM Approach

As follows from the previous section, the usefulness of MM approach was already demonstrated in the first decade of its extensive applications and modifications. At the same time, the problems of justification and of pathways of future development of the approach had arisen. From the very beginning, it was clear that the method cannot provide us with exact description of the structures and processes due to its semiempirical nature. The interesting problem of the method is the relative extent of empiricism and physical meaning of the scheme and its parameters.

### Approaches to Improvement of MM Formulae and Parameters

Several ways of improvement of MM scheme were proposed during the 1960s and the 1970s. Two hypothetical approaches to refinement of MM representation and force fields can be considered. In the framework of the first approach, it is suggested to describe the physical nature of the system in the most exact way, e.g., to employ more exact equations for description of the interaction energy (e.g., not restricting expressions to first term in the expansion for bond stretching and angle bending, use of better grounded expressions for van der Waals interactions, etc.) and/or more detailed representation of the system considering not only point atoms but other centers of interactions, not restricting the scheme with scalar values but introducing vectors (such as dipole moments) or tensors (such as anisotropic polarizability). In the framework of the second approach, it would be possible to consider energy expressions as entirely empirical ones and to adjust the parameters to experimental data (or, later, to reliable quantum mechanics results) until the closest coincidence (e.g., via the least square method). Unsuccessful attempts of consistent and strict use of only one of the approaches had been undertaken during the first two decades of the MM computer simulations.

The second approach mentioned above is practically impossible to realize; usually, there is no sufficient quantitative experimental data not only to derive the form of dependences of the energy on inter-atomic distances but to obtain the coefficients of already accepted potential form (e.g., of Lennard-Jones). Even in a

case we have sufficient number of experimental values, the equations for parameter calculations have no single definite solution. It was demonstrated in the paper of Momany et al. (1968). The authors of the abovementioned paper tried to adjust the coefficient of (6-12) potentials for C . . . C, C . . . H, and H . . . H interactions for benzene molecule to the energy and structure data of benzene crystal. Using three sets of experimental data for different temperatures, they obtained entirely repulsive contribution from C-C interactions. It means that qualitative estimations (relative values of attraction terms for various atoms, of equilibrium radii, of atom charges in the same or similar molecules, etc.) should be taken into account in adjusting the parameters of the force field. Most of the modern force fields originate from previous sets of expressions and parameters derived from simple qualitative considerations (e.g., simplified formulae of dispersion term, van der Waals radii of atoms) and take into account these considerations automatically (e.g., using previous approximate parameters' values as starting points).

As for the first approach, the more complex expressions of the energy were suggested since the end of 1960s, and such expressions are used in some modern force fields, i.e., next terms in the energy expansions for stretching, bending, and van der Waals contributions. The inclusion of additional terms requires new parameters. Usually the number of reliable quantitative values is not sufficient, and such an approach becomes useful in case the area of applications of the force field is restricted to certain types of similar molecules (e.g., hydrocarbons). In any case, the expressions for each term remain the approximate ones, and more coefficients should be adjusted to target data. The more complex representation of the molecular system as compared to a set of point atoms was proposed by several researchers during the 1960s and the 1970s (including the author of this survey). Such a representation looks the most natural for atoms with lone pairs of electrons (inclusion of additional points for electrostatic interactions of pyridine nitrogen or keto oxygen located at the centers of lone-pair orbitals) or for atoms with π-orbitals (additional points for aromatic heterocycles located above and below the ring planes to account for quadrupole moments of the molecules). Additional lone pair centers for electrostatic interactions are included in some modern force fields for better reproduction of electric field around the molecules and the directionality of hydrogen bonds.

Although the introduction of such centers looks physically based, an inconsistency of such an approach can be detected rather easily. When assigning the negative charged interaction centers to some electron orbitals (and hence positively charged centers to nuclei locations), it looks natural to assign additional centers to chemical bonds (sites of the highest electron density between the nuclei). As a result, we have several times more centers (more computer resources are needed) and several times more adjustable parameters. Such an approach was consistently implemented into force field suggested by Scheraga and coauthors. The first version was titled "empirical potential using electrons and nuclei" (EPEN ; Shipman et al. 1975) and an improved version EPEN/2 was released in 3 years (Snir et al. 1978). In this model, the molecular interactions are modeled through distributed interaction sites, which account for the nuclei and electronic clouds. The positive charges are

located at the atomic nuclei, and the negative charge centers are located off the nuclei. The bonding electrons are located along the bonds, while the lone-pair and π- electrons are located off the bonds. The energy of interaction is approximated by the sum of the coulomb interactions between all point charge centers, an exponential repulsion to represent electron–electron overlap repulsion, and an $R^{-6}$ (R = distance) attraction to simulate attractive energies between the nuclei. The distances between the charge locations are the method parameters; they are fixed in the molecule fragments. The parametrization was optimized by least square fits to spectroscopic, crystallographic, and thermodynamic data. This approach (as well as other approaches considering many in addition to atom centers of interactions) was not widely used in MM calculations. The situation can be accounted for by restrictions imposed by fixed geometry of the molecule fragments as well as by need for the parameter adjustment for molecules other than considered by the approach authors.

A separate problem is the choice of the force field parameters for most exact reproduction of experimental and/or quantum mechanics data. All the experimental values are determined with a definite accuracy; all the quantum mechanics data refer to gas phase and do not include implicitly third body interactions. The use of quantum mechanics data only (not depending on level of theory) results in errors in reproduction of condensed phase properties. The adjustment of parameters to characteristics of condensed phase (most experimental data refer to crystals or solutions) leads to wrong values of characteristics for separate molecules or their simple complexes in gas phase. Hence, various experimental and quantum mechanics data should be considered with different weights in the process of the adjustment of force fields, the weights being dependent on areas of intended applications of the force field.

Recently, a method and software package named ForceBalance for systematic force field optimization has been published (Wang et al. 2013a). The method enables researchers to parametrize a wide variety of functional forms using flexible combinations of target data. These data may come from experimental measurements or from quantum mechanics calculations. Various properties are included in a single objective function, which is then integrated over the entire configuration space. The method was used by the authors (Wang et al. 2013a) for developing a new polarizable water model with 27 parameters adjustable to quantum mechanics data for single water molecule and H-bonded water clusters. The water model contains five fluctuating charge sites including three virtual sites. The positions of virtual sites have been fully optimized; they do not coincide with lone pair centers. An interesting suggestion was made for description of van der Waals interactions. The authors propose specially modified (6-exp) potential. As it was mentioned in the first section, “classical” (6-exp) Buckingham function widely used on the first steps of MM-like calculations has unphysical dependence on interatomic distance at small distances. The authors of the model (Wang et al. 2013a) have chosen empirical van der Waals potential $E_{vdW}(r)$ with three parameters.

$$E_{vdW}(r) = 2\varepsilon\left(\rho^6/\left(\rho^6 + r^6\right)\right)/(1 - 3/(\gamma + 3))\left[3\exp(\gamma(1 - r/\rho))/(\gamma + 3) - 1\right] \quad (7)$$

This function depends on equilibrium interatomic distance, ρ; minimal energy values, ε; and a dimensionless constant, γ, describing the steepness of the repulsion. Like Buckingham function, it tends to an attractive $r^{-6}$ potential at large interatomic distances, tends to a repulsive exponential term at small distances, and has well-defined unique minimum. The authors note that their water model is quite accurate for some, but not all, properties of liquid water. We should note that this disadvantage is common for all the force fields derived from quantum mechanics results only, without adjustment to experimental data. Improvements of water model using ForceBalance software and both quantum mechanics and experimental data will be considered below.

## Various Schemes of Water Molecule in MM Calculations

In view of importance of water for life and for description of the biomolecular interactions as well as of "anomalous" physical properties of water in liquid and solid states, hundreds of models and their modifications for water molecule have been proposed. We will mention here only few of most frequently used and cited rigid models and some recent polarizable models. Simple atomic structure of the water molecule enables researchers to perform quantum mechanics calculations of rather high level unreachable for most organic molecules. The results of these computations for a single molecule and for H-bonded water clusters together with availability of quantitative experimental data for bulk water at various temperatures and pressures stimulate elaboration of new models and permanent improvement of existent ones. These models can be used for demonstration of possibilities and restrictions of the MM method in general, for comparison of various approaches to MM model construction.

Most water models describe the molecule by the number of interaction centers not coinciding exactly with atom centers. This diversion from simple MM scheme can be understood as a consequence of small size of the molecule and the inability of any simple charge distribution on atom centers to adequately describe the molecular dipole and quadrupole of the monomer at the same time. Inclusion of additional centers or/and displacement of interaction centers from atoms is necessary to reproduce charge distribution of this simple molecule. (The author of this survey noticed this fact earlier for heterocyclic compounds, but fortunately in this case the contributions of permanent quadrupole interactions can be implicitly accounted for by other terms of interaction energy.)

The first water molecule model (frequently referred to as BF model) has been proposed by Bernal and Fowler (1933). This model may be considered as the earliest molecular model suitable for MM calculations, and it was used for Monte Carlo calculations of liquid water performed in nearly a half of century for comparison with some later models (Jorgensen et al. 1983). The model contains four interaction sites: three charged sites, namely, two positive centers located at the hydrogen atoms and one negative center (so-called M-site) displaced by 0.15 Å from oxygen atom

in the direction of H-O-H bisector, and one center for van der Waals interactions located on oxygen atom. It is interesting to note that Bernal and Fowler (1933) performed approximate evaluations of intermolecular interaction energy by the (1-6-12) potential functions. Jorgensen et al. (1983) used nearly the same position of centers and slightly modified charges in TIP4P four-site water model (one of their popular Transferable Intermolecular Potentials). Matsuoka et al. (1976) derived more complex four-site model from potential surface of water dimer constructed via ab initio quantum mechanics calculations with configuration interactions. The analytical expression for interaction potential function between two water molecules consists of pair-wise Coulomb interactions of three charges located similar to BF or TIP4P models and exponential interactions of all the three atoms of one molecule with the atoms of the other one. This model (the so-called MCY model) was widely used during the first decade of liquid water simulations.

We will not consider many models including earliest ones which consist of three, five, or more centers but only briefly mention some of the rigid models cited till the present time and used in modern force fields. Stillinger and Rahman (1974) suggested the five-site model with two positive centers (hydrogen atoms), two negative centers (located at oxygen lone pairs), and one center (oxygen atom) for Lennard-Jones interactions. The model has been used successfully to study a wide variety of properties of liquid water. Mahoney and Jorgensen (2001), after extensive Monte Carlo simulations and parameters adjustment, suggested five-site model TIP5P which enables them to describe the density anomaly of liquid water better than previously existing models. Nada and Van der Eerden (2003) have designed six-site model of water molecule. Three of these sites are the three atoms of the molecule (they interact through a Lennard-Jones potential), and three other sites are point charges (O atom is electrically neutral; the two H atoms carry positive charges). The model can be considered as a combination of TIP4P and TIP5P models of Jorgensen. The Monte Carlo simulations of ice and water show that the six-site model reproduces well the real structural and thermodynamic properties of ice and water near the melting point.

Two three-site rigid water models, namely, TIP3P of Jorgensen et al. (1983) and SPC of Berendsen et al. (1981) should be specially mentioned, as these models (or their modifications) are used in modern force fields for biomolecular simulations. These models have (in accord with general MM principles for class I models) three interaction sites centered on the nuclei. Each site has a partial charge for computing the Coulomb energy and only one site (oxygen atom) for Lennard-Jones interactions. The TIP3P model demonstrates better reproduction of experimental thermodynamic and structural data for bulk water (Jorgensen et al. 1983).

Further progress in the study of water properties and its interactions with other molecules is related to new models introducing flexibility to the molecule (incorporating bond stretching and angle bending terms into interaction energy) and with the inclusion of electronic polarization. These effects are important for the quantitative description of water properties in liquid and solid states as well as of phase transitions. We mention here a few of recent water models beyond class I MM methods.

The model elaborated by Ren and Ponder (2003, 2004), AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications), includes full intramolecular flexibility, permanent atomic monopole, dipole, and quadrupole moments placed on each atomic center and buffered 7-14 potential for pairwise atom-atom van der Waals interactions. Polarization effects are explicitly treated in the AMOEBA force field via mutual induction of dipoles at atomic centers. The functional forms for bond stretching and angle bending were taken from the Allinger's MM3 force field and include anharmonicity through the use of higher-order deviations from ideal bond lengths and angles. An additional intramolecular term, the Urey-Bradley term, consists of a simple harmonic function. The force field does not contain off-atom sites. It was parametrized by hand to fit results from ab initio calculations on gas phase clusters up to hexamers (Ren and Ponder 2003) and temperature and pressure dependent bulk phase properties (Ren and Ponder 2004). The model enables authors to reproduce well both the high-level ab initio results for various gas phase clusters and bulk water thermodynamics data, including the second virial coefficients over the whole temperature range. The authors mention that AMOEBA water model is approximately eight times slower than the TIP3P one (for MD simulations of a periodic system of 216 water molecules); nevertheless, they consider it as "the first step in the derivation of a next-generation polarizable force field for biomolecular simulation." Later, AMOEBA force field was extended to organic molecules (Ren et al. 2011) and to proteins (Shi et al. 2013).

The iAMOEBA, i.e., "inexpensive AMOEBA" (Wang et al. 2013b) was developed using the ForceBalance automated procedure (Wang et al. 2013a). A direct approximation to describe electronic polarizability reduces the computational cost relative to a fully polarizable AMOEBA model, and the application of systematic optimization ForceBalance method results in raising the accuracy of iAMOEBA.

Recently (Laury et al. 2015), a set of improved parameters for the AMOEBA water model was developed using the automated procedure, ForceBalance (Wang et al. 2013a). An adjustment of 21 force field parameters has been performed to both high-level ab initio results for gas phase clusters, containing 2–20 water molecules, and six condensed phase properties at a range of temperatures (249–373 K) and pressures (1 atm to 8 kilobar). The enthalpy of vaporization and density dependence on temperature were given a greater weight (as compared to the other properties) in the optimization. The model represents a significant refinement of the original AMOEBA parametrization, namely, it provides a substantial improvement in the ability to predict experimental data, particularly for a number of liquid properties across a range of temperatures. The authors (creators of original AMOEBA models are among them) refer to new model as AMOEBA14 and to the original model (Ren and Ponder 2003, 2004) as AMOEBA03. The AMOEBA14 model has a similar accuracy level as the AMOEBA03 model for the interaction energy of the gas phase cluster ranging from dimers to 20-mers. They promise compatibility of AMOEBA14 water model with the previously determined AMOEBA force fields for organic molecules (Ren et al. 2011) and for proteins (Shi et al. 2013).

The authors of CHARMM force field and software package, widely used for biomolecular simulation (see description below), elaborate their own polarizable

models of water based on classical Drude oscillator. Last model of their SWM (simple water model) series, six-site SWM6 one (Yu et al. 2013) was developed and optimized for liquid phase simulations under ambient conditions using CHARMM family of force fields. The atoms of the molecule kept its gas phase experimental geometry, and the positions of three other massless interaction sites correspond to the so-called M point located at a fixed distance from the oxygen atom along the bisector of the H–O–H angle, and two oxygen's lone pairs (LP sites). The model is less computer time consuming as compared to more sophisticated AMOEBA; it has an improved balance between descriptions of microscopic and bulk phase properties compared to most additive and polarizable water models developed previously. Besides, the local hydrogen bonding structures in the liquid phase are better reproduced in the model as indicated by the O-O distance distribution function. However, the model does not reproduce correctly the experimental data on liquid water density maximum. Authors note that the model is not developed for routine use but for studies when the local hydrogen bonding is important and high accuracy is needed.

We will not discuss many other recent sophisticated water models, including those developed using both experimental data and quantum mechanics computations (via both ab initio and DFT methods), such as Tröster's polarizable molecular mechanics (PMM) model (Tröster et al. 2014) and Pinilla's polarizable and dissociable force field (Pinilla et al. 2012). Selected old and recent MM water models, both simple and rather sophisticated, enable us to arrive at some conclusions on possibilities of molecular mechanics description of simple and complex molecular systems. Some of such conclusions are rather trivial, but additional evidence from the experience of hundreds of researchers on water molecule modeling can help to illustrate and reinforce them.

Till recent time, complete and reliable MM description of even this very simple biologically important molecule is a challenge for computationally oriented researchers. More complex expressions for separate energy terms grounded on physical theories, increase of the number of adjustable parameters, improved algorithms for parameter adjustment to target data, and the use of more reliable quantum mechanics results, all together help to better reproduce and rationalize experimental data, but some results are not accurately reproduced. Each model of the molecule, each force field, reproduces (and predicts) some characteristics better as compared to other force fields. The more sophisticated model is not necessarily the better one for specific MM applications and specific tasks of the study. This is true not only because of restrictions of computer resources but due to approximate nature of the MM description. This conclusion is a result of many trials and errors in water simulations. Class I molecular mechanics representations, i.e., simple three-site rigid water models (TIP3P and its modifications) are successfully used in many biomolecular computations in spite of the development of new force fields. This success can be accounted for not by the importance of additional terms but by their implicit compensation by parameter choice for simple model (Class I force fields). More sophisticated force fields (Class II and Class III), together with quantum

mechanics, both ab initio and semiempirical, computations are useful for detailed description of mutual positions of molecules in water solutions of biomolecules.

## Modern Molecular Mechanics Force Fields: Description, Recent Development, and Applications

The MM computations during the first two decades were mainly performed by the researchers using their own force fields and homemade computer programs. Most of the modern publications refer to standard, ready-to-use software with implemented (or sometimes slightly modified) force field parameters. The number of program and parameter packages is rather great and new or newly modernized software and force fields appear every year. We will present below a short overview of the selected program sets and force fields most frequently used for study of biomolecular systems. Some of them have been already mentioned in previous sections. During last decade, some software packages such as NAMD (Phillips et al. 2005) and TINKER (Ponder 2015) have appeared which give an opportunity to perform computations using various molecular mechanics force fields. Some popular quantum mechanics software packages such as GAUSSIAN and SCHRODINGER supply users with the possibility to use MM force fields as well.

### Allinger's Force Fields and Programs

We started description of the modern force field with short characterization of Allinger's parameters and programs as they were the first ones ready for use by any researcher. The first parameter set (MM1) was described in 1973 (Allinger and Sprague 1973), and the MM1 program was released through Quantum Chemistry Program Exchange in 1976. The force field and the program MM1 were followed by a series of improved and extended versions for various classes of organic molecules, MM2, MM3, MM4 (see Allinger et al. 1989, 1996; Lii et al. 2005 for references). These force fields have been parameterized based on experimental data, including the electron diffraction, vibrational spectra, heats of formation, and crystal structures. The calculation results were verified via comparison with high-level ab initio quantum chemistry computations, and the parameters were additionally adjusted. The last program versions include rather complex force field equations (including additional terms to harmonic potentials for stretching and bending and cross-terms; thus MM3 and MM4 force field are those of class II) and various calculation options (e.g., prediction of frequencies and intensities of vibrational spectra, calculating the entropy and heat of formation for a molecule at various temperatures). MM4's molecular dynamics option creates a description of the vibrational and conformational motion of molecules as a function of time. The program examines the atoms of a molecule and their environment and decides which atom type is appropriate. The program includes more energy terms and more complicated expressions for some terms than most other popular MM force

fields. The force field contains greater number of the constants which are adjusted to extended set of experimental data. The stretching and bending energy terms include higher order contributions, several cross-terms are inserted; the electrostatic contribution includes interactions of bond dipoles.

The MM family of force fields designed by Allinger and coauthors is used till recent time for calculations on small and medium size organic molecules, e.g., steroidal-porphyrin aggregates (Lorecchio et al. 2014). New improvements to MM4 force field were suggested by Allinger (2011) recently. The classic of molecular mechanics, after some examples make the conclusion: "A good (well developed, Class 3) molecular mechanics force field will usually give molecular structures that are approximately of chemical accuracy. When it fails to do so, there is a good chance that nature knows something that we don't" (Allinger 2011).

## Merck Molecular Force Field (MMFF94)

This force field of class II was developed for a broad range of "medium size" molecules primary of importance for drug design. It differs from Allinger's and many other popular force fields in several aspects. The core portion of MMFF94 has primarily been derived from high-quality computational quantum chemistry data (up to MP4SDQ/TZP level of theory) for a wide variety of chemical systems of interest to organic and medical chemists (Halgren 1996, 1999b). Nearly all MMFF94 parameters have been determined in a mutually consistent fashion from the full set of available computational data. The force field reproduces well both computational and experimental data, including experimental bond lengths (0.014 Å rms), bond angles (1.2° rms), vibrational frequencies, conformational energies, and rotational barriers. The mathematical expressions of the force field differ from many other force fields. The expressions for stretching and bending energies (like Allinger's force fields) contain in addition to harmonic terms stretch-bend cross terms and out-of-plane terms. The van der Waals energy is expressed by buffered (7-14) potentials (Eq. 8, the designations are the same as in Eqs. 2, 3, 4, 5, 6, and 7)

$$E_{vdW} = \varepsilon_{ij}\left(1.07\ \rho_{ij}/\left(r_{ij} + 0.07\ \rho_{ij}\right)^7\right)\left(\left(1.12\rho_{ij}{}^7/\left(r_{ij}{}^7 + 0.12\ \rho_{ij}{}^7\right)\right) - 2\right) \quad (8)$$

The electrostatic term contains buffering constant as well. The specially formulated combination rules for $\varepsilon_{ij}$ and $\rho_{ij}$ are used. The MMFF94 version was developed for molecular dynamics simulations, and MMFF94*s* version (Halgren 1999a) was modified for energy minimization studies (*s* in the title means static). Both versions provide users with nearly the same results for majority of molecules and complexes. Additional parameters of force field and careful adjustment to high-level reference data result in good reproduction of extended set of experimental data on conformational energies of the molecules appeared after force field publication (Halgren 1999b). Nevertheless, Bordner et al. (2003) reveal that MMFF94 values of sublimation energy for a set of compounds are systematically 30–40 % lower

than experimental data. It can be explained by "computational nature" of MMFF parameters and by the adjustment to quantum mechanics results without corrections for the experimental data, i.e., nonadditive contributions do not accounted for implicitly.

This force field is used in many recent studies of biologically important molecules, which employ combined experimental and computational approaches. Anti-inflammatory drugs (Pawar et al. 2014; Sawant et al. 2014) and anticholinergic and antiasthmatic drugs (Guo et al. 2014) are among the compounds considered recently with MMFF.

## The Force Fields and Programs Designed by Scheraga and Coauthors

As we had already mentioned that the first MM parameter set for the simulation of proteins was described in 1975 by Scheraga and coauthors (Momany et al. 1975); their empirical conformational energy program for peptides (ECEPP) was released through *Quantum Chemistry Program Exchange* in the same year. This force field family was refined continuously till last years; we will briefly describe two versions of Scheraga's force field used for studies of protein three-dimensional structure and protein folding up to recent time, ECEPP/3 (Némethy et al. 1992) and ECEPP-05 (Arnautova et al. 2006). The ECEPP/3 force field in turn is the modified version of ECEPP/2; it contains updated parameters for proline and oxyproline residues. The ECEPP force fields utilize fixed bond lengths and bond angles obtained from a survey of the crystal structures for each of amino acid residue. The peptide energy is calculated and minimized as a function of torsion angles only, and it is a significant advantage for minimization approaches as it drastically reduces the variable space that must be considered. The energy function of the ECEPP/3 force field consists of torsion and nonbonded (Coulomb and van der Waals) terms (Eqs. 4 and 5). The partial charges of atoms are calculated by the molecular orbital CNDO/2 method, the parameters of (6-12) term are adjusted by comparison with crystal data. The (6-12) terms are substituted by (10-12) ones for the interactions of the hydrogen atoms capable to participate in H bonds with H-bond acceptor atoms. In spite of appearance of new improved and augmented versions of ECEPP force field, the ECEPP/3 is used until recently for protein simulations (e.g., McAllister and Floudas 2010).

The ECEPP-05 force field (Arnautova et al. 2006 and references herein) is adjusted to both new experimental data and quantum mechanics results. It has some distinctive features as compared to both ECEPP/3 version and many other popular force fields. The van der Waals term of the energy is modeled by using the "6-exp" potential function (potential of Buckingam)

$$E_{ij} = -A_{ij} r_{ij}^{-6} + B_{ij} \exp\left(-C_{ij} r_{ij}\right) \tag{9}$$

where $r_{ij}$ is the distance between atoms i and j; $A_{ij}$, $B_{ij}$, and $C_{ij}$ are parameters of the potential. The combination rules are applied directly to these parameters.

$$A_{ij} = \left(A_{ii}A_{jj}\right)^{1/2};\ B_{ij} = \left(B_{ii}B_{jj}\right)^{1/2};\ C_{ij} = \left(C_{ii} + C_{jj}\right)/2 \tag{10}$$

The parameters of (6-exp) potentials were derived by using so-called global-optimization-based method consisting of two steps. An initial set of parameters is derived from quantum mechanics interaction energies (at MP2/6-31G* level of ab initio theory) of dimers of selected molecules; at the second step, the initial set is refined to satisfy the following criteria: the parameters should reproduce the observed crystal structures and sublimation enthalpies of related compounds, and the experimental crystal structure should correspond to the global minimum of the potential energy.

The atomic charges were fitted to reproduce the molecular ab initio electrostatic potential, calculated at HF/6-31G* level; the fitting was carried out using the restrained electrostatic potential (RESP) method taken from the AMBER program. The method was applied to obtain a single set of charges using several conformations of a given molecule (the multiple-conformation-derived charges). An additional point charge (with zero nonbonded parameters) is used to model the lone pair electrons of $sp^2$ nitrogen. The torsional parameters of the side chains of the 20 naturally occurring amino acids were computed by fitting to rotational energy profiles obtained from ab initio MP2/6-31G** calculations. The peptide backbone torsional parameters were obtained by fitting the energies to the Φ-Ψ energy maps of terminally blocked glycine and alanine amino acids constructed at MP2/6-31G**//HF/6-31G** level of theory. The performance of the force field was evaluated by simulating crystal structures of small peptides.

As the important problems of the protein structure and functions as well as of the natural science as a whole refer to prediction of three-dimensional structure and pathways of folding of proteins, Scheraga and coauthors have modified and augmented the force fields mentioned above by some theoretical approaches and concepts. We will notice a few of them related directly to MM calculations. The all-atom ECEPP05 force field was coupled with an implicit solvent-accessible surface area model (ECEPP05/SA, Arnautova and Scheraga 2008). Recognizing the impossibility of searching enormous conformational space of real protein with an all-atom potential functions, a hierarchical procedure was developed. Its main features are the initial use of a united residue (UNRES) potential functions and an efficient procedure, conformational space annealing (CSA) (Scheraga et al. 2002, and references herein). The protein is first optimized with the low-resolution "coarse-grained" UNRES model (Liwo et al. 1999). In UNRES force field, the backbone is represented as a virtual-bond chain of $C_\alpha$ atoms, and the side chains are depicted as ellipsoids. The interaction sites are the united-atom side chains and the centers of the peptide groups between $C_\alpha$ atoms. The lowest-energy UNRES conformation, as well as a set of distinct low-energy conformations, is then converted to all-atom models. Finally, the all-atom models are refined with the EDMC method, with inclusion of continuum hydration model.

The search for low-energy "native-like" structures includes several minimization and Monte Carlo procedures. The simplest of the MC methods is Monte Carlo with

minimization, i.e., a Metropolis Monte Carlo (MCM) algorithm in which every trial state is first energy-minimized before the Boltzmann acceptance criterion is applied. The electrostatically driven Monte Carlo (EDMC) method employs a move set in which individual peptide groups are selected at random and rotated "in place" (i.e., the conformational change is localized to the peptide group as much as possible) so as to optimize the alignment of its dipole moment with the local electric field. The methodology assumes that a protein molecule is driven toward the native structure by the combined action of electrostatic interactions and stochastic conformational changes associated with thermal movements (Ripoll and Scheraga 1988). The use of the force fields and methods enables Scheraga and coauthors not only to rationalize but to predict some interesting features of protein three-dimensional structure and pathways of folding. "With increasing refinement of the computational procedures, the computed results are coming closer to experimental observations, providing an understanding as to how physics directs the folding process" (Scheraga 2008).

Several recent successful applications of UNRES force to simulations of biological functions of proteins can be mentioned. The mechanism of growth of amyloid β-peptide fibrils (the main component of plaques formed in the human brain during Alzheimer disease) was suggested (Rojas et al. 2011) and supported by simulated 2D ultraviolet spectroscopy (Lam et al. 2013). The mechanism of HSP70 heat-shock protein action was proposed which was based on ATP-induced scissor-like motion of the two sub-domains of the nucleotide-binding domain (NBD) (Golas et al. 2012). The computationally obtained structure of this chaperon was confirmed then by experimental study. Some additional examples of UNRES applications are mentioned in very recently published review of Scheraga (2015). Similar approach was developed by Scheraga and coauthors for nucleic acids interactions. The first paper on coarse-grained models of nucleic acids represents NARES approach (Maciejczyk et al. 2010) to nucleic acid bases. Each base is simulated by a combination of three to five interaction centers of 6-12 van der Waals and permanent dipole interactions. Then, UNRES-2P model was developed (He et al. 2013) representing a polynucleotide chain by a sequence of virtual sugar atoms, located at the geometric center of the sugar ring, linked by virtual bonds with attached united base and united phosphate groups. This oversimplified model reproduces the double helix formation, some its thermodynamics properties (He et al. 2013), and was very recently applied to DNA-protein interactions (Yin et al. 2015).

## Force Fields and Programs Developed by Kollman and Coauthors. AMBER

The computer software and force fields referred to as "AMBER" (Assisted Model Building with Energy Refinement) are the most popular of those designed for wide class of biomolecules including proteins and nucleic acids. The first description of the computer program AMBER appeared in 1981 (Weiner and Kollman 1981), and

the first detailed specification of force field was published 3 years later (Weiner et al. 1984). The force field contained usual "minimalist" set of energy terms (Eqs. 1, 2, 3, 4, 5, and 6 of this survey), the parameters being adjusted to both experimental and quantum mechanics data. The parameters of their previous publications and those of other authors were used as starting points for the adjustment. Atomic charges are obtained by fitting a partial charge model to electrostatic potentials calculated by ab initio quantum mechanics calculations (with minimal STO 3G basis set). The (10-12) potential was included for description of H-bonding (this term was deleted in new versions of AMBER force fields and software). The force field accounted for implicit solvent by using distance dependent dielectric constant ($\varepsilon = R_{ij}$), united atom approach (i.e., several atoms, e.g., four atoms of a methyl group are considered as one particle) enabled one to perform calculations of sufficiently extended systems (such as fragments of DNA double helix of a few nucleotide pairs or peptides of dozens of amino acid residues). In spite of several deficiencies of the first AMBER force field (e.g., too short distances in stacking complexes due to short equilibrium radii for van der Waals interactions, defects in atomic charges due to very approximate quantum mechanics data), hundreds of papers were published with this force field and the program. In 2005, even after release of several new and improved AMBER force field versions, the paper (Weiner et al. 1984) was the 10th most-cited one in the history of the *Journal of the American Chemical Society* (Case et al. 2005).

Since those publications new versions of both force fields and program sets were released. Cornell et al. (1995) published "second generation" AMBER force field for simulating the structures, conformational energies, and interaction energies of proteins, nucleic acids, and many related organic molecules in condensed phases. The charges were determined using the calculations with 6-31G* basis set and restrained electrostatic potential (RESP) fitting and had been shown to reproduce interaction energies, free energies of solvation, and conformational energies of small molecules to a good degree of accuracy. The new van der Waals parameters were derived to account for liquid properties. Retaining most of torsion parameters of "the previous generation" of force field, the improvement of potentials for peptide backbone was performed for $\Phi$ and $\Psi$ dihedrals as well as for the nucleoside $\chi$ dihedrals by adding extra Fourier terms. This "second generation" force field is usually referred to as *ff94* (parameter set *parm94*).

There are a number of AMBER parameter sets, with names beginning with "ff" and followed by a two-digit year number, such as *ff99*. We will briefly mention some new features of these force fields. Some versions differ from *ff94* by minor (but important for some problems) modifications. The *ff96* force field differs from *ff94* only in the peptide $\Phi$-$\Psi$ torsional potential, and *parm96* was developed because it was clear that parm94 overstabilized $\alpha$ helices relative to $\beta$ sheets. Thus, *parm94* seems adequate for comparing structures with similar secondary structures, but *parm96* seems to more accurately represent the relative stability of $\alpha$ and $\beta$ secondary structures. The improved version *ff99* (Wang et al. 2000) was developed using the restrained electrostatic potential (RESP) approach to derive the partial charges. The adjustment of atom point charges and torsion constants for the set of 34

molecules to the highest level ab initio model (GVB/LMP2) and experimental data enables authors to reproduce data for 55-molecule set with the better accuracy than MM3, MMFF, and early versions of CHARMM force fields. It was demonstrated that "a well-parameterized harmonic force field with a reliable charge method can describe the structure and intramolecular energies for organic systems very well" (Wang et al. 2000). The authors tried to use the adjusted parameters (*parm99*) in nonadditive models (which include polarization), and obtained "reasonable results using the same parameters derived for the additive model, although additional torsional parameterization is required to achieve the same high level accuracy as that found using the additive model" (Wang et al. 2000). This force field is widely used during more than 10 years. The other AMBER force field, "a third-generation point-charge all-atom force field for proteins" (referred to as *ff03*) was developed by Duan et al. (2003). The main difference from previous force fields is that all quantum mechanics calculations were done in the condensed phase with continuum solvent models and an effective dielectric constant of $\varepsilon = 4$. Initial tests on peptides demonstrated a high degree of similarity between the calculated and the statistically measured Ramanchandran maps for Glycine and Alanine peptide fragments. One more AMBER force field referred to as general Amber force field (GAFF) should be mentioned (Wang et al. 2004). GAFF have been designed as an extension of the AMBER force fields to be compatible with existing versions for proteins and nucleic acids and has parameters for most organic molecules of pharmaceutical importance that are composed of H, C, N, O, S, P, and halogens. The parameterization is based on more than 3000 MP2/6-31G* optimizations and 1260 MP4/6-311G(d,p) single-point calculations. Unlike most conventional force fields, parameters for all combinations of atom types are not contained in an exhaustive table but are determined algorithmically for each input molecule based on the bonding topology (which determines the atom types) and the geometry. A completely automated, table-driven procedure (called *antechamber,* a part of the late versions of AMBER program set) was developed to assign atom types, charges, and force field parameters to almost any organic molecule (Wang et al. 2006). New versions of AMBER software work with several force fields. The newest AMBER 11 version (Case et al. 2010) supports CHARMM fixed-charge force fields as well.

The initial AMBER force fields and program packages (due to the state of the theory and computer facilities) were generally aimed at the search for energy minima of separate molecular systems in vacuum. The current AMBER versions are strongly aimed at simulations (molecular dynamics, evaluations of free energy changes) of biomolecules in water solutions. Both explicit solvent models (using TIP3P, TIP4P, TIP5P, and SPC water models as well as models of some organic solvents) and implicit ones are supported. By default, explicit solvation model involves electrostatic interactions handled by a particle-mesh Ewald (PME) procedure, and long-range Lennard-Jones attractions are treated by a continuum model. The PME is a modified form of Ewald summation, a method to efficiently calculate the infinite range Coulomb interaction under periodic boundary conditions, and PME is a modification to accelerate the Ewald reciprocal sum to near linear scaling using the three-dimensional fast Fourier transform (3DFFT).

An accurate description of the aqueous environment is essential for atom-level biomolecular simulations but may become very expensive computationally. An implicit model replaces the discrete water molecules by an infinite continuum medium with some of the dielectric and "hydrophobic" properties of water. The continuum implicit solvent models have several advantages over the explicit water representation, especially in molecular dynamics simulations (e.g., they are often less expensive and generally scale better on parallel machines; they correspond to instantaneous solvent dielectric response; the continuum model corresponds to solvation in an infinite volume of solvent, there are no artifacts of periodic boundary conditions; estimating free energies of solvated structures is much more straightforward than with explicit water models). Despite the fact that the methodology represents an approximation at a fundamental level, it has in many cases been successful in calculating various macromolecular properties (Case et al. 2005).

The total energy of a solvated molecule can be written as a sum of molecule's energy in vacuum (gas phase) and the free energy of transferring the molecule from vacuum into solvent, that is, solvation free energy typically assumed as can be decomposed into the electrostatic and nonelectrostatic parts. The above decomposition is the basis of the widely used PB/SA (Poisson-Boltzman solvent accessible surface area) scheme. AMBER follows this way to compute the nonelectrostatic part of the energy. The most time-consuming part of solvation energy estimations is the computation of the electrostatic contribution. AMBER employs one of the most popular biomolecular applications the finite-difference method (FDPB) and analytic generalized Born (GB) method. In the GB model, a molecule in solution is represented as a set of spherical cavities of dielectric constant of 1 with charges at their centers embedded in a polarizable dielectric continuum solvent of a higher dielectric constant.

The AMBER software and force fields are widely used for the simulation of various systems of biological and material science importance. Some researchers develop proper additions and corrections to "standard" parameters for better description of specific systems. We refer to only few examples. Perez et al. (2007) reparameterized the *parm99* force field for nucleic acid simulations improving the representation of the α/γ conformational space. Zgarbova et al. (2011) refined description of glycosidic torsion angles of nucleic acids. Other work of Zgarbova et al. (2013) refers to ε and ζ DNA torsions. Very recently, Ogata and Nakamura (2015) suggested the improvement of parameters of the force field for phospholipids. Song et al. (2008) adjusted the AMBER force field to reproduce conformational properties of an oxidative DNA lesion, 2,6-diamino-4-hydroxy-5-formamidopyrimidine. New improved parameters for modified peptide chain were suggested by Grubišić et al. (2013) recently. Some of such improvements have been included into newer versions of AMBER package. AMBER force field can be used via other MM program packages, such as CHARMM, GROMACS, NAMD, and with widely used QM package GAUSSIAN.

The latest version of AMBER program package is AMBER14 distributed with AmberTools 15 (Case et al. 2015). The new versions of AMBER software support the option of allowing part of the system to be described quantum mechanically

in an approach known as a hybrid (or coupled potential) QM/MM simulation. Both semiempirical and the density functional theory-based tight-binding (DFTB) Hamiltonian can be used. The system may contain both two nonbonded parts and covalently bonded QM and MM subsystems (Case et al. 2010).

## Other Popular Force Fields and MM Software. CHARMM, OPLS, GROMOS

The force fields and program packages listed above are used for nearly the same type of simulations as those considered above. The main differences refer to the organization of the programs and to relative role of experimental data, and quantum mechanics results in force field parameter adjustment. The first versions of force fields and programs as well as the whole research programs were initiated by such pioneers of MM applications as Karplus (CHARMM, Chemistry at HARvard Molecular Mechanics), Jorgensen (OPLS, Optimized Potentials for Liquid Simulations), Gunsteren and Berendsen (GROMOS, GROningen MOlecular Simulation package), and GROMACS (GROningen MAchine for Chemical Simulation). The progress and improvement of the force fields and programs are connected to comparison with the approaches and results of other programs. The parametrization of second generation AMBER force field (Cornell et al. 1995) had been influenced by Jorgensen OPLS potentials for proteins (Jorgensen and Tirado-Rives 1988), whereas these OPLS united atom potentials adopted bond stretch, angle bend, and torsional terms from the AMBER united-atom force field (Weiner et al. 1984). The OPLS parameters for the 20 neutral peptide residues were obtained primarily via Monte Carlo simulations for the 36 organic liquids, and the parametrization for the charged protein residues was performed via comparisons with ab initio results for ion-molecule complexes (6-31G(d) basis set) and on Monte Carlo simulations for hydrated ions (Jorgensen and Tirado-Rives 1988). The OPLS-AA (all atom) force field (Jorgensen et al. 1996) retain most of bond stretch and angle bending parameters from AMBER all-atom force fields, but torsion and nonbonding constants are reparametrized utilizing both experimental and ab initio (RHF/6-31G* level) data. This force field was improved for peptides by means of refitting the key Fourier torsional coefficients using accurate ab initio data (LMP2/cc-pVTZ(-f)//HF/6-31G** level) as the target (Kaminski et al. 2001).

Most of popular program sets utilize various force fields, not only those developed in the research groups of the authors of the program with the same title. The software suite GROMACS was developed by Berendsen group at the University of Groningen, The Netherlands, starting in early 1990s. This fast program for molecular dynamics simulation does not have a force field of its own but is compatible with GROMOS, OPLS, and AMBER force fields. The GROMACS (van der Spoel et al. 2005) was developed and optimized especially for use on PC-clusters. It originated from the Fortran package GROMOS developed by van Gunsteren et al. (1987). The GROMOS force fields are not pure all-atom force fields; aliphatic $CH_n$ groups are treated as united atoms. The recent version, GRO-

MOS05 (Christen et al. 2005), utilize new force field sets (53A5 and 53A6, the first number in the names of a parameter sets refer to the number of atom types) based on extensive reparameterization of the previous GROMOS force field (Oostenbrink et al. 2004). In contrast to the parameterization of most other biomolecular force fields, this parameterization is based primarily on reproducing the free enthalpies of hydration and apolar solvation for a range of organic compounds. Recently, 53A6 force field has been reparametrized by Pol-Fachin et al. (2014) by adjustment to QM data (HF/6-31G* level of theory) for applications to study Glycoproteins.

Very popular for the computations of wide range of macromolecular systems are CHARMM (Chemistry at HARvard Macromolecular Mechanics) program and force field sets. The computations include energy minimization, molecular dynamics, and Monte Carlo simulations. CHARMM originated at Martin Karplus's group at Harvard University, the first publication on CHARMM program and force field (refer to Brooks et al. 1983). Several versions of both software and force field were released since the first ones. The improvement of the programs and force fields are in progress till now. The mathematical expressions for calculation of CHARMM energy are nearly the same as "minimalist" ones for other force fields (MacKerell 2004). The minor differences with e.g., AMBER or OPLS force fields refer to inclusion of harmonic improper and Urey-Bradley terms. The parameters of various versions of force field are consistently adjusted with emphasis to specific molecules to be applied. The first versions of CHARMM force fields were aimed primarily to protein molecular dynamics simulations. Then, the special adjustments were performed for the simulations of the nucleic acids (Foloppe and MacKerell 2000) and lipids (Klauda et al. 2008). The improvement of parameter for specific molecules continues till recent times, while other parameters can be used from the previous versions. Raman et al. (2010) recently extended the CHARMM additive carbohydrate all-atom force field to enable modeling glycosidic linkages in carbohydrates involving furanoses. Like in most popular modern force fields, both the ab initio quantum mechanics computations and experimental data for solid and liquid states are used in adjustment of the parameters of CHARMM force fields. The distinct feature of this adjustment refers to charge adscription. Like OPLS force fields (Jorgensen et al. 1996), CHARMM utilize so-called supramolecular approach. The partial atomic charges are adjusted to reproduce ab initio (HF/6-31G* level) minimum interaction energies and geometries of model compounds with water or for model compound dimers (with energy and distance corrections for reproducing the correct experimental densities). On the other hand, the use of ab initio data on small clusters to optimize the van der Waals parameters leads to poor condensed phase properties. This requires that the optimization of these parameters be performed by empirical fitting to reproduce thermodynamics properties from condensed phase simulations, taking into account the relative values of the parameters obtained from high-level ab initio data on interactions of the model compounds with rare gases (Klauda et al. 2008).

Very recently, Vanommeslaeghe and MacKerell (2015) published detailed description of the development of CHARMM method and software. We will mention here only some most important for biological and medical applications

recent advancements of both all-atom additive force field (on the way to CHARMM36 version, modern class I model) and polarizable force fields (class II models). Best et al. (2012) presented a combination of refinements important with respect to the equilibrium between helical and extended conformations in protein folding simulations. Peptide backbone (φ, ψ) potential has been refined against experimental solution NMR data. Side-chain $\chi_1$ and $\chi_2$ torsion parameters have been optimized by fitting to quantum-mechanical energy surfaces, followed by additional empirical optimization targeting NMR data. A comprehensive validation of the revised force field indicates that the revised CHARMM 36 parameters represent an improved model for modeling and simulation studies of proteins, including protein folding, assembly, and functionally relevant conformational changes (Best et al. 2012). The refinement of the CHARMM27 all-atom additive force field for nucleic acids via adjustment of backbone parameters to QM data (MP2/6-31 + G(d) level of theory) followed by trial MD simulations of DNA duplexes enabled Hart et al. (2012) to improve treatment of the BI/BII equilibrium of DNA conformations. Many other specific refinements of the CHARMM additive force fields for these and other classes of biologically important molecules have been mentioned in the recent paper of Vanommeslaeghe and MacKerell (2015).

Along with comprehensive refinements and biomolecular applications of class I CHARMM force fields, polarizable CHARMM models are extensively developed during recent years for more accurate and physically based description of biologically important molecular systems. These force fields use the classical Drude oscillator (i.e., "charge-on-spring") model. In the model, a virtual negatively charged particle ("Drude particle") is connected to the core of each polarizable atom by means of a harmonic potential ("spring"). Description of this class II force field is out of scope of this survey, but we will mention some results of proteins and nucleic acid computations with such models. Very recently, Huang et al. (2014) summarized recent efforts to include the explicit treatment of induced electronic polarization in biomolecular force fields. They arrive with conclusion that the inclusion of explicit electronic polarizability leads to significant differences in the physical forces affecting the structure and dynamics of proteins (Huang et al. 2014). The polarizable force field more accurately models peptide-folding cooperativity (Huang and MacKerell 2014). Development of a CHARMM Drude polarizable model for nucleic acids (Baker et al. 2011) requires significant adjustments of the phosphodiester backbone and glycosidic bond dihedral parameters of base-sodium interactions and of selected electrostatic parameters in the model (Savelyev and MacKerell 2014a). The modern polarizable model satisfactorily treats the equilibrium between the A, BI, and BII conformations of DNA (Savelyev and MacKerell 2014b).

It is interesting to mention that multiple adjustments of different additive force fields mentioned above and other force fields result in an increase of similarity in both corresponding parameters values and the results of applications to complex molecular systems (such as protein and nucleic acid fragments), but many differences both in corresponding parameter values and computation results still remain.

## Conclusions

The molecular mechanics computation approach to the study of the molecular systems arose a half of century ago and today has become a useful and powerful research tool for many branches of natural science. During this period the number of publications utilizing this approach increased by a few orders of magnitude. This increase continues up to present days, followed by new journals appearance and a flow of MM publications in the journals of rather wide area of coverage. The MM considers the molecular systems via classic Newtonian mechanics approach, i.e., using classical approximation to essentially quantum systems. The MM suggests representation of the system energy as a sum of terms responsible for interactions of various physical natures.

The minimalist MM approach involves the simple expressions for chemical bond stretching, bond angle bending, torsions about the bonds, and nonbonded van der Waals and Coulomb contributions calculated via additive atom-atom scheme. The force fields (sets of mathematical formulae and numerical parameters) are derived from simple physical considerations followed by adjustment to the experimental data and precise quantum mechanics results. An additivity of the terms responsible for separate atom contributions and for interactions of different physical nature and transferability of the parameters between atoms and molecules of similar structure are the main assumptions of MM approach. Development of the approach results in elaboration of a variety of computer software and force fields for simulating different molecular systems of biological, medical, and industrial importance. Various force fields differ in the complexity of mathematical expressions and in the relative role of experimental and quantum mechanics results used for the parameter adjustment. The development of algorithms for the MM applications (including those for molecular dynamics and Monte Carlo techniques) and the increase of computer power enable researchers to approach such complex problems as predicting the pathways of the biopolymers three-dimensional structure, molecular recognition in various biological processes, assistance in creation of compounds with desirable properties (including drugs and industrial materials). Analysis of the current state of physicochemical study of biological processes suggests that MM simulations based on empirical force fields have an ever-increasing impact on understanding structure and functions of biomolecules. Rather close results obtained with different force fields can be accounted for by existing various solutions of the task of parameter adjustment even to the same target data (really, target data used when elaborating various force fields differ, slightly or considerably). Force fields are internally consistent. Their parameters are highly correlated within a force field. Several force fields have been elaborated recently going beyond the atom-atom approach and the additivity principle. Some force fields utilize a hierarchical procedure for consideration of complex systems.

## Bibliography


Allinger, N. L. (1959). Conformational analysis. III. Applications to some medium ring compounds. *Journal of the American Chemical Society, 81*, 5727.

Allinger, N. L. (2010). *Molecular structure: Understanding steric and electronic effects from molecular mechanics*. Hoboken: Willey.

Allinger, N. L. (2011). Understanding molecular structure from molecular mechanics. *Journal of Computer-Aided Molecular Design, 25*, 295.

Allinger, N. L., & Sprague, J. T. (1973). Calculation of the structures of hydrocarbons containing delocalized electronic systems by the molecular mechanics method. *Journal of the American Chemical Society, 95*, 3893.

Allinger, N. L., Tribble, M. T., Miller, M. A., & Wertz, D. W. (1971). Conformational analysis. LXIX. Improved force field for the calculation of the structures and energies of hydrocarbons. *Journal of the American Chemical Society, 93*, 1637.

Allinger, N. L., Yuh, Y. H., & Lii, J. H. (1989). Molecular mechanics. The MM3 force field for hydrocarbons. *Journal of the American Chemical Society, 111*, 8551.

Allinger, N. L., Chen, K., & Lii, J. H. (1996). An improved force field (MM4) for saturated hydrocarbons. *Journal of Computational Chemistry, 17*, 642.

Antony, J., & Grimme, S. (2006). Density functional theory including dispersion corrections for intermolecular interactions in a large benchmark set of biologically relevant molecules. *Physical Chemistry Chemical Physics, 8*, 5287.

Arnautova, Y. A., & Scheraga, H. A. (2008). Use of decoys to optimize an all-atom force field including hydration. *Biophysical Journal, 95*, 2434.

Arnautova, Y. A., Jagielska, A., & Scheraga, H. A. (2006). A new force field (ECEPP-05) for peptides, proteins and organic molecules. *The Journal of Chemical Physics, 110*, 5025.

Baker, C. M., Anisimov, V. M., & MacKerell, A. D., Jr. (2011). Development of CHARMM polarizable force field for nucleic acid bases based on the classical Drude oscillator model. *The Journal of Physical Chemistry B, 115*, 580.

Bartell, L. S. (1960). On the effects of intramolecular van der Waals forces. *The Journal of Chemical Physics, 32*, 827.

Barton, D. H. R. (1948). Interaction between non-bonded atoms, and the structure of cis-decalin. *Journal of the Chemical Society*, 340.

Barton, D. H. R. (1950). The conformation of the steroid nucleus. *Experientia, 6*, 316.

Berendsen, H. J. C., Postma, J. P. M., von Gunstaren, W. F., & Hermans, J. (1981). Interaction models for water in relation to protein hydration. In B. Pullman (Ed.), *Intermolecular forces* (pp. 331–342). Dordrecht: Reidel.

Berman, H. M., Olson, W. K., Beveridge, D. l., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R., & Schneider, B. (1992). The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophysical Journal, 63*, 751.

Berman, H. M., Henrick, K., & Nakamura, H. (2003). Announcing the worldwide protein data bank. *Nature Structural & Molecular Biology, 10*, 980.

Bernal, J. D., & Fowler, R. H. (1933). A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. *The Journal of Chemical Physics, 1*, 515.

Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., & MacKerell, A. D., Jr. (2012). Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone $\varphi$, $\psi$ and side-chain $\chi_1$ and $\chi_2$ dihedral angles. *Journal of Chemical Theory and Computation, 8*, 3257.

Bordner, A. J., Cavasotto, C. N., & Abagyan, R. A. (2003). Direct derivation of van der Waals force field parameters from quantum mechanical interaction energies. *The Journal of Physical Chemistry B, 107*, 9601.

Bradley, D. F., Lifson, S., & Honig, B. (1964). Theory of optical and other properties of biopolymers: Applicability and elimination of the first-neighbor and dipole-dipole approximations. In

B. Pullman (Ed.), *Electronic aspects of biochemistry*. New York: Academic.

Brandenburg, J. G., & Grimme, S. (2014). Accurate modeling of organic molecular crystals by dispersion-corrected density functional tight binding (DFTB). *The Journal of Physical Chemistry Letters, 5*, 1785.

Brant, D. A., & Flory, P. J. (1965). The configuration of random polypeptide chains. II. Theory. *Journal of the American Chemical Society, 87*, 2791.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry, 4*, 187.

Case, D. A., Cheatham, T. E., III, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computation Chemistry, 26*, 1668.

Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., & Kollman, P. A. (2010). *AMBER 11*. San Francisco: University of California.

Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., Cheatham, T. E., III, Darden, T., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, K. M., Monard, G., Needham, P., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., York, D. M., & Kollman, P. A. (2015). *AMBER 2015*. San Francisco: University of California.

Christen, M., Hunenberger, P. H., Bakowies, D., Baron, R., Burgi, R., Geerke, D. P., Heinz, T. N., Kastenholz, M. A., Krautler, V., Oostenbrink, C., Peter, C., Trzesniak, D., & Van Gunsteren, W. F. (2005). The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry, 26*, 1719.

Chuprina, V. P., & Poltev, V. I. (1983). Possible conformations of double-helical polynucleotides containing incorrect base-pairs. *Nucleic Acids Research, 11*, 5205.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I., Merz, K., Jr., Ferguson, D., Spellmeyer, D., Fox, T., Caldwell, J., & Kollman, P. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society, 117*, 5179.

Craig, D. P., Mason, R., Pauling, P., & Santry, D. P. (1965). Molecular packing in crystals of the aromatic hydrocarbons. *Proceedings of the Royal Society A, 286*, 98.

Cramer, C. J. (2004). *Essentials of computational chemistry: Theories and models*. Chichester: Wiley.

De Santis, P. (1992). Conformational energy calculations of macromolecules. *Current Contents, 34*, 8.

De Santis, P., Giglio, E., Liquori, A. M., & Ripamonti, A. (1963). Stability of helical conformations of simple linear polymers. *Journal of Polymer Science: Part A, 1*, 1383.

De Santis, P., Giglio, E., Liquori, A. M., & Ripamonti, A. (1965). Interaction and stability of helical polypeptide chains. *Nature, 206*, 456.

De Voe, H., & Tinoco, I., Jr. (1962). The stability of helical polynucleotides: Base contributions. *Journal of Molecular Biology, 4*, 500.

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., & Kollman, P. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry, 24*, 1999.

Eisenberg, D. (2003). The discovery of the α-helix and β-sheet, the principal structural features of proteins. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 11207.

Engler, E. M., Andose, J. D., & Schleyer, P. R. (1973). Critical evaluation of molecular mechanics. *Journal of the American Chemical Society, 95*, 8005.

Foloppe, N., & MacKerell, A. D., Jr. (2000). All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of Computational Chemistry, 21*, 86.

Gibson, K. D., & Scheraga, H. A. (1967a). Minimization of polypeptide energy. I. Preliminary structures of bovine pancreatic ribonuclease S-peptide. *Proceedings of the National Academy of Sciences of the United States of America, 58*, 420.

Gibson, K. D., & Scheraga, H. A. (1967b). Minimization of polypeptide energy. II. Preliminary structures of oxytocin, vasopressin, and an octapeptide from ribonuclease. *Proceedings of the National Academy of Sciences of the United States of America, 58*, 1317.

Golas, E., Maisuradze, G. G., Senet, P., Oldziej, S., Czaplewski, C., Scheraga, H. A., & Liwo, A. (2012). Simulation of the opening and closing of Hsp70 chaperones by coarse-grained molecular dynamics. *Journal of Chemical Theory and Computation, 8*, 1750.

Gresh, N., Claverie, P., & Pullman, A. (1986). Intermolecular interactions: Elaboration on an additive procedure including an explicit charge-transfer contribution. *International Journal of Quantum Chemistry, 29*, 101.

Grimme, S. (2004). Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry, 25*, 1463.

Grimme, S. (2011). Density functional theory with London dispersion corrections. *WIREs Computational Molecular Science, 1*, 211.

Grubišić, S., Brancato, G., & Barone, V. (2013). An improved AMBER force field for α, α-dialkylated peptides: Intrinsic and solvent-induced conformational preferences of model systems. *Physical Chemistry Chemical Physics, 15*, 17395.

Guo, X., Wang, Z., Zuo, L., Zhou, Z., Guo, X., & Sun, T. (2014). Quantitative prediction of enantioseparation using b-cyclodextrin derivatives as chiral selectors in capillary electrophoresis. *Analyst, 139*, 6511.

Halgren, T. A. (1996). Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry, 17*, 490.

Halgren, T. A. (1999a). MMFF VI. MMFF94s option for energy minimization studies. *Journal of Computational Chemistry, 20*, 720.

Halgren, T. A. (1999b). MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *Journal of Computational Chemistry, 20*, 730.

Hart, K., Foloppe, N., Baker, C. M., Denning, E. J., Nilsson, L., & MacKerell, A. D., Jr. (2012). Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *Journal of Chemical Theory and Computation, 8*, 348.

He, Y., Maciejczyk, M., Oldziej, S., Scheraga, H. A., & Liwo, A. (2013). Mean-field interactions between nucleic-acid-base dipoles can drive the formation of the double helix. *Physical Review Letters, 110*, 098101.

Hendrickson, J. B. (1961). Molecular geometry. I. Machine computation of the common rings. *Journal of the American Chemical Society, 83*, 4537.

Hendrickson, J. B. (1962). Molecular geometry. II. Methyl-cyclohexanes and cycloheptanes. *Journal of the American Chemical Society, 84*, 3355.

Hendrickson, J. B. (1973). Molecular geometry. VIII. Proton magnetic resonance studies of cycloheptane conformations. *Journal of the American Chemical Society, 95*, 494.

Hill, T. L. (1946). On steric effects. *The Journal of Chemical Physics, 14*, 465.

Hill, T. L. (1948). Steric effects. I. Van der Waals potential energy curves. *The Journal of Chemical Physics, 16*, 399.

Huang, J., & MacKerell, A. D., Jr. (2014). Induction of peptide bond dipoles drives cooperative helix formation in the $(AAQAA)_3$ peptide. *Biophysical Journal, 107*, 991.

Huang, J., Lopes, P. E. M., Roux, B., & MacKerell, A. D., Jr. (2014). Recent advances in polarizable force fields for macromolecules: Microsecond simulations of proteins using the classical Drude oscillator model. *The Journal of Physical Chemistry Letters, 5*, 3144.

Jorgensen, W. L., & Tirado-Rives, J. (1988). The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society, 110*, 1657.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics, 79*, 926.

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society, 118*, 11225.

Jurecka, P., Cerny, J., Hobza, P., & Salahub, D. R. (2007). Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of noncovalent complexes compared with ab initio quantum mechanics calculations 80. *Journal of Computational Chemistry, 28*, 555.

Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., & Jorgensen, W. L. (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B, 105*, 6474.

Khutorsky, V. E., & Poltev, V. I. (1976). Conformations of double-helical nucleic acids. *Nature, 264*, 483.

Kitaigorodski, A. I. (1959). *Organic chemical crystallography*. New York: Consultants Bureau.

Kitaigorodsky, A. I. (1973). *Molecular crystals and molecules*. New York: Academic.

Kitaygorodsky, A. I. (1960). Calculation of conformations of organic molecules. *Tetrahedron, 9*, 183.

Kitaygorodsky, A. I. (1961). The interaction curve of non-bonded carbon and hydrogen atoms and its application. *Tetrahedron, 14*, 230.

Klauda, J. B., Venable, R. M., MacKerell, A. D., Jr., & Pastor, R. W. (2008). Considerations for lipid force field development. *Current Topics in Membranes, 60*, 1.

Lam, A. R., Rodriguez, J. J., Rojas, A., Scheraga, H. A., & Mukamel, S. (2013). Tracking the mechanism of fibril assembly by simulated two-dimensional ultraviolet spectroscopy. *The Journal of Physical Chemistry A, 117*, 342.

Langlet, J., Claverie, P., Caron, F., & Boeuve, J. C. (1981). Interactions between nucleic acid bases in hydrogen bonded and stacked configurations: The role of the molecular charge distribution. *International Journal of Quantum Chemistry, 20*, 299.

Laury, M. L., Wang, L.-P., Pande, V. S., Head-Gordon, T. L., & Ponder, J. W. (2015). Revised parameters for the AMOEBA polarizable atomic multipole water model. *Journal of Physical Chemistry B*. doi:10.1021/jp510896n.

Leach, A. R. (2001). *Molecular modelling: Principles and applications*. Harlow: Prentice Hall (Pearson Education).

Leach, S. J., Némethy, G., & Scheraga, H. A. (1966a). Computation of the sterically allowed conformations of peptides. *Biopolymers, 4*, 369.

Leach, S. J., Némethy, G., & Scheraga, H. A. (1966b). Intramolecular steric effects and hydrogen bonding in regular conformations of polyamino acids. *Biopolymers, 4*, 887.

Levitt, M., & Lifson, S. (1969). Refinement of protein conformations using a macromolecular energy minimization procedure. *Journal of Molecular Biology, 46*, 269.

Lewars, E. G. (2011). *Computational chemistry. Introduction to the theory and applications of molecular and quantum mechanics* (2nd ed.). Dordrecht/Heidelberg/London/New York: Springer.

Lifson, S., & Warshel, A. (1968). Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and *n*-alkane molecules. *The Journal of Chemical Physics, 49*, 5116.

Lii, L.-H., & Allinger, N. L. (1991). The MM3 force field for amides, polypeptides and proteins. *Journal of Computational Chemistry, 12*, 186.

Lii, L.-H., Chen, K.-H., Johnson, G. P., French, A. D., & Allinger, N. L. (2005). The external-anomeric torsional effect. *Carbohydrate Research, 340*, 853.

Liwo, A., Lee, J., Ripoll, D. R., Pillardy, J., & Scheraga, H. A. (1999). Protein structure prediction by global optimization of a potential energy function. *Proceedings of the National Academy of Sciences of the United States of America, 96*, 5482.

Lorecchio, C., Venanzi, M., Mazzuca, C., Lettieri, R., Palleschi, A., Thi, T. H. N., Cardová, L., Drasar, P., & Monti, D. (2014). Tuning the chiroptical and morphological properties of steroidal-porphyrin aggregates: A mechanistic, structural, and MM investigation. *Organic & Biomolecular Chemistry, 12*, 3956.

Maciejczyk, M., Spasic, A., Liwo, A., & Scheraga, H. A. (2010). Coarse-grained model of nucleic acid bases. *Journal of Computational Chemistry, 31*, 1644.

MacKerell, A. D., Jr. (2004). Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry, 25*, 1584.

Mahoney, M. W., & Jorgensen, W. L. (2001). Quantum, intramolecular flexibility, and polarizability effects on the reproduction of the density anomaly of liquid water by simple potential functions. *The Journal of Chemical Physics, 115*, 10758.

Mason, R. (1969). The intermolecular potential and structure of crystals of aromatic molecules. *Molecular Crystals and Liquid Crystals, 9*, 3.

Matsuoka, O., Clementi, E., & Yoshimine, M. (1976). CI study of the water dimer potential surface. *The Journal of Chemical Physics, 64*, 1351.

McAllister, S. R., & Floudas, C. A. (2010). An improved hybrid global optimization method for protein tertiary structure prediction. *Computational Optimization and Applications, 45*, 377.

McGuire, R. F., Momany, F. A., & Scheraga, H. A. (1972). Energy parameters in polypeptides. V. An empirical hydrogen bond potential function based on molecular orbital calculations. *The Journal of Physical Chemistry, 76*, 375.

Momany, F. A., Vanderkooi, G., & Scheraga, H. A. (1968). Determination of intermolecular potentials from crystal data. I. General theory and application to crystalline benzene at several temperatures. *Proceedings of the National Academy of Sciences of the United States of America, 61*, 429.

Momany, F. A., Carruthers, L. M., & Scheraga, H. A. (1974). Intermolecular potentials from crystal data. IV. Application of empirical potentials to the packing configurations and lattice energies in crystals of amino acids. *The Journal of Physical Chemistry, 78*, 1621.

Momany, F. A., McGuire, R., Burgess, A., & Scheraga, H. (1975). Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *The Journal of Physical Chemistry, 79*, 2361.

Nada, H., & van der Eerden, J. P. J. M. (2003). An intermolecular potential model for the simulation of ice and water near the melting point: A six-site model of $H_2O$. *The Journal of Chemical Physics, 118*, 7401.

Nash, H. A., & Bradley, D. F. (1966). Calculation of the lowest energy configurations of nucleotide base pairs on the basis of an electrostatic model. *The Journal of Chemical Physics, 45*, 1380.

Némethy, G., & Scheraga, H. A. (1965). Theoretical determination of sterically allowed conformations of a polypeptide chain by a computer method. *Biopolymers, 3*, 155.

Némethy, G., Gibson, K. D., Palmer, K. A., Yoon, C. N., Paterlini, G., Zagari, A., Rumsey, S., & Scheraga, H. A. (1992). Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *The Journal of Physical Chemistry, 96*, 6472.

Ogata, K., & Nakamura, S. (2015). Improvement of parameters of the AMBER potential force field for phospholipids for description of thermal phase transitions. *The Journal of Physical Chemistry*. doi:10.1021/acs.jpcb.5b01656.

Oostenbrink, C., Villa, A., Mark, A. E., & van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry, 25*, 1656.

Pauling, L., & Corey, R. B. (1951). The pleated sheet, a new layer configuration of polypeptide chains. *Proceedings of the National Academy of Sciences of the United States of America, 37*, 251.

Pauling, L., Corey, R. B., & Branson, H. R. (1951). The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America, 37*, 205.

Pawar, S., Sawant, S., Nerkar, A., & Bhosale, A. (2014). In silico design, synthesis and pharmacological screening of novel 2-(6-substituted benzo [d] thiazol-2-yl) isoindoline-1, 3-diones as potential COX-2 inhibitors for anti-inflammatory activity. *International Journal of Pharmacy and Pharmaceutical Sciences, 6*, 353.

Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T. E., Laughton, C. A., & Orozco, M. (2007). Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophysical Journal, 92*, 3817.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., & Klaus Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry, 26*, 1781.

Pinilla, C., Irani, A. H., Nicola Seriani, N., & Scandolo, S. (2012). Ab initio parameterization of an all-atom polarizable and dissociable force field for water. *The Journal of Chemical Physics, 136*, 114511.

Pol-Fachin, L., Hugo Verli, H., & Lins, R. D. (2014). Extension and validation of the GROMOS $53A6_{glyc}$ parameter set for glycoproteins. *Journal of Computational Chemistry, 35*, 2087.

Polozov, R. V., Poltev, V. I., & Sukhorukov, B. I. (1973). Relation of the interactions of nucleic acid bases to the helical conformations of polynucleotides. *Studia Biophysica, 40*, 13.

Poltev, V. I., & Bruskov, V. I. (1978). On molecular mechanisms of nucleic acid synthesis fidelity aspects 1. Contribution of base interactions. *Journal of Theoretical Biology, 70*, 69.

Poltev, V. I., & Shulyupina, N. V. (1986). Simulation of interactions between nucleic-acid bases by refined atom-atom potential functions. *Journal of Biomolecular Structure & Dynamics, 3*, 739.

Poltev, V. I., & Sukhorukov, B. I. (1967). Theoretical examination of the physical nature of the intermolecular interactions determining the conformational state of polynucleotides. *Biophysics (Moscow), 12*, 879.

Poltev, V. I., & Sukhorukov, B. I. (1970). Semiempirical calculations of interaction energy of DNA nitrous bases. *Studia Biophysica, 24/25*, 179.

Poltev, V. I., Grokhlina, T. I., & Malenkov, G. G. (1984). Hydration of nucleic-acid bases studied using novel atom-atom potential functions. *Journal of Biomolecular Stucture & Dynamics, 2*, 413.

Ponder, J. W. (2015). TINKER – Software tools for molecular design. http://dasher.wustl.edu/tinker/

Pullman, A., & Pullman, B. (1968). Aspects of the electronic structure of the purine and pyrimidine bases of the nucleic acids and of their interactions. *Advances in Quantum Chemistry, 4*, 267.

Pullman, B., Claverie, P., & Caillet, J. (1966). Van der Waals-London interactions and the configuration of hydrogen-bonded purine and pyrimidine pairs. *Proceedings of National Academy of Sciences of United States of America, 55*, 904.

Rae, A. I. M., & Mason, R. (1968). The intermolecular potential and the lattice energy of benzene. *Proceedings of the Royal Society A, 304*, 487.

Ramachandran, G. N. (1990). This week's citation classic. *Current Contents, 10*, 119.

Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology, 7*, 95.

Ramachandran, K. I., Deepa, G., & Namboori, K. (2008). *Computational chemistry and molecular modeling: Principles and applications*. Berlin: Springer.

Raman, E. P., Guvench, O., & MacKerell, A. D., Jr. (2010). CHARMM additive all-atom force field for glycosidic linkages in carbohydrates involving furanoses. *The Journal of Physical Chemistry B, 114*, 12981.

Rasse, D., Warme, P. K., & Scheraga, H. A. (1974). Refinement of the X-ray structure of rubredoxin by conformational energy calculations. *Proceedings of the National Academy of Sciences of the United States of America, 71*, 3736.

Ren, P., & Ponder, J. W. (2003). Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B, 107*, 5933.

Ren, P., & Ponder, J. W. (2004). Temperature and pressure dependence of the AMOEBA water model. *The Journal of Physical Chemistry B, 108*, 13427.

Ren, P., Wu, C., & Ponder, J. W. (2011). Polarizable atomic multipole-based molecular mechanics for organic molecules. *Journal of Chemical Theory and Computation, 7*, 314.

Renugopalakrishnan, V., Lakshminarayanan, A. V., & Sasisekharan, V. (1971). Stereochemistry of nucleic acids and polynucleotides III. Electronic charge distribution. *Biopolymers, 10*, 1159.

Ripoll, D. R., & Scheraga, H. A. (1988). On the multiple-minima problem in the conformational analysis of polypeptides. II. An electrostatically driven Monte Carlo method – Tests on poly(L-alanine). *Biopolymers, 27*, 1283.

Rojas, A. V., Liwo, A., & Scheraga, H. A. (2011). A study of the α-helical intermediate preceding the aggregation of the amino-terminal fragment of the β amyloid peptide (Aβ1–28). *The Journal of Physical Chemistry B, 115*, 12978.

Savelyev, A., & MacKerell, A. D., Jr. (2014a). Balancing the interactions of ions, water and DNA in 1314 the Drude polarizable force field. *The Journal of Physical Chemistry B, 118*, 6742.

Savelyev, A., & MacKerell, A. D., Jr. (2014b). All-atom polarizable force field for DNA based on 1317 the classical Drude oscillator model. *Journal of Computational Chemistry, 35*, 1219.

Sawant, R. L., Hardas, D. B., Pawa, K. K., & Shinde, A. K. (2014). QSAR analysis of structurally similar 1, 3, 4-oxadiazole/thiadiazole and 1, 2, 4-triazole derivatives of biphenyl-4-yloxy acetic acid as antiinflammatoryagents. *World Journal of Pharmaceutical Research, 3*, 1844.

Scheraga, H. A. (2008). From helix–coil transitions to protein folding. *Biopolymers, 89*, 479.

Scheraga, H. A. (2015). My 65 years in protein chemistry. *Quarterly Reviews of Biophysics, 48*, 117.

Scheraga, H. A., Pillardy, J., Liwo, A., Lee, J., Czaplewski, C., Ripoll, D. R., Wedemeyer, W. J., & Arnautova, Y. A. (2002). Evolution of physics-based methodology for exploring the conformational energy landscape of proteins. *Journal of Computational Chemistry, 23*, 28.

Scott, R. A., & Scheraga, H. A. (1966a). Conformational analysis of macromolecules. III. Helical structures of poly-glycine and poly-L-alanine. *The Journal of Chemical Physics, 45*, 2091.

Scott, R. A., & Scheraga, H. A. (1966b). Conformational analysis of macromolecules. II. The rotational isomeric states of the normal hydrocarbons. *Journal of Chemical Physics, 44*, 3054.

Shi, Y., Xia, Z., Zhang, J., Best, R., Wu, C., Ponder, J. W., & Ren, P. (2013). Polarizable atomic multipole-based AMOEBA force field for proteins. *Journal of Chemical Theory and Computation, 9*, 4046.

Shipman, L. L., Burgess, A. W., & Scheraga, H. A. (1975). A new approach to empirical intermolecular and conformational potential energy functions. I. Description of model and derivation of parameters. *Proceedings of the National Academy of Sciences of the United States of America, 72*, 543.

Sippl, M. J., Némethy, G., & Scheraga, H. A. (1984). Intermolecular potentials from crystal data. 6. Determination of empirical potentials for O–H$\cdots$O=C hydrogen bonds from packing configurations. *Journal of Physics Chemistry, 88*, 6231.

Snir, J., Nemenoff, R. A., & Scheraga, H. A. (1978). A revised empirical potential for conformational, intermolecular, and solvation studies. 5. Development and testing of parameters for amides, amino acids and peptides. *The Journal of Physical Chemistry, 82*, 2527.

Song, K., Hornak, V., de los Santos, C., Grollman, A. P., & Simmerling, C. (2008). Molecular mechanics parameters for the FapydG DNA lesion. *Journal of Computational Chemistry, 29*, 17.

Stillinger, F. H., & Rahman, A. (1974). Improved simulation of liquid water by molecular dynamics. *The Journal of Chemical Physics, 60*, 1545.

Tröster, P., Lorenzen, K., & Tavan, P. (2014). Polarizable six-point water models from computational and empirical optimization. *The Journal of Physical Chemistry B, 118*, 1589.

Van der Spoel, D., Lindahl, E., Hess, B., Groehof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry, 26*, 1701.

Van Gunsteren, W., Fand, H. J., & Berendsen, C. (1987). *Groningen molecular simulation (GROMOS) library manual*. Groningen: BIOMOS.

Vanommeslaeghe, K., & MacKerell, A. D., Jr. (2015). CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochimica et Biophysica Acta, General Subjects, 1850*, 861.

Wang, J., Cieplak, P., & Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry, 21*, 1049.

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general AMBER force field. *Journal of Computational Chemistry, 25*, 1157.

Wang, J., Wang, W., Kollman, P. A., & Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics & Modeling, 25*, 247.

Wang, L.-P., Chen, J., & van Voorhis, T. (2013a). Systematic parametrization of polarizable force fields from quantum chemistry data. *Journal of Chemical Theory and Computation, 9*, 452.

Wang, L.-P., Head-Gordon, T., Ponder, J. W., Ren, P., Chodera, J. D., Eastman, P. K., Martinez, T. J., & Pande, V. S. (2013b). Systematic improvement of a classical molecular model of water. *The Journal of Physical Chemistry B, 117*, 9956.

Warme, P. K., & Scheraga, H. A. (1974). Refinement of the X-ray structure of lysozyme by complete energy minimization. *Biochemistry, 13*, 757.

Warme, P. K., Momany, F. A., Rumball, S. V., & Scheraga, H. A. (1974). Computation of structures of homologous proteins; α-lactalbumin from lysozyme. *Biochemistry, 13*, 768.

Warshel, A., & Lifson, S. (1970). Consistent force field calculations. II. Crystal structures, sublimation energies, molecular and lattice vibrations, molecular conformations, and enthalpies of alkanes. *The Journal of Chemical Physics, 53*, 582.

Watson, J. D., & Crick, F. H. C. (1953). A structure for deoxyribose nucleic acid. *Nature, 171*, 737.

Weiner, P., & Kollman, P. (1981). AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *Journal of Computational Chemistry, 2*, 287.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr., & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society, 106*, 765.

Westheimer, F. H., & Mayer, J. E. (1946). The theory of the racemization of optically active derivatives of diphenyl. *The Journal of Chemical Physics, 14*, 733.

Wiberg, K. B. (1965). A scheme for strain energy minimization. Application to the cycloalkanes. *Journal of the American Chemical Society, 87*, 1070.

Williams, D. E. (1966). Nonbonded potential parameters derived from crystalline aromatic hydrocarbons. *The Journal of Chemical Physics, 45*, 3770.

Williams, D. E. (1967). Nonbonded potential parameters derived from crystalline hydrocarbons. *The Journal of Chemical Physics, 47*, 4680.

Williams, D. E., & Weller, R. R. (1983). Lone-pair electronic effects on the calculated ab initio SCF-MO electric potential and the crystal structures of azabenzenes. *Journal of the American Chemical Society, 105*, 4143.

Yan, J. F., Momany, F. A., Hoffmann, R., & Scheraga, H. A. (1970). Energy parameters in polypeptides. II. Semiempirical molecular orbital calculations for model peptides. *The Journal of Physical Chemistry, 74*, 420.

Yilmazer, N. D., & Korth, M. (2015). Enhanced semiempirical QM methods for biomolecular interactions. *Computational and Structural Biotechnology Journal, 13*, 169.

Yin, Y., Sieradzan, A. K., Liwo, A., He, Y., & Scheraga, H. A. (2015). Physics-based potentials for coarse-grained modeling of protein-DNA interactions. *Journal of Chemical Theory and Computation, 11*, 1792.

Yu, W., Lopes, P. E. M., Roux, B., & MacKerell, A. D. (2013). Six-site polarizable model of water based on the classical Drude oscillator. *The Journal of Chemical Physics, 138*, 034508.

Zgarbova, M., Otyepka, M., Sponer, J., Mladek, A., Banas, P., Cheatham, T. E., & Jurecka, P. (2011). Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *Journal of Chemical Theory and Computation, 7*, 2886.

Zgarbova, M., Luque, F. J., Sponer, J., Cheatham, T. E., Otyepka, M., & Jurecka, P. (2013). Toward improved description of DNA backbone: Revisiting epsilon and zeta torsion force field parameters. *Journal of Chemical Theory and Computation, 9*, 2339.

Zhurkin, V. B., Poltev, V. I., & Florentiev, V. L. (1980). Atom-atom potential functions for conformational calculations of nucleic-acids. *Molecular Biology (Moscow), English Translation, 14*, 882.