Journal of Computer-Aided Molecular Design
https://doi.org/10.1007/s10822-018-0111-4

PERSPECTIVE

Biomolecular force fields: where have we been, where are we now, where do we need to go and how do we get there?

Pnina Dauber-Osguthorpe[2] · A. T. Hagler[1,3]

Received: 17 November 2017 / Accepted: 9 March 2018

Abstract

In this perspective, we review the theory and methodology of the derivation of force fields (FFs), and their validity, for molecular simulations, from their inception in the second half of the twentieth century to the improved representations at the end of the century. We examine the representations of the physics embodied in various force fields, their accuracy and deficiencies. The early days in the 1950s and 60s saw FFs first introduced to analyze vibrational spectra. The advent of computers was soon followed by the first molecular mechanics machine calculations. From the very first papers it was recognized that the accuracy with which the FFs represented the physics was critical if meaningful calculated structural and thermodynamic properties were to be achieved. We discuss the rigorous methodology formulated by Lifson, and later Allinger to derive molecular FFs, not only obtain optimal parameters but also uncover deficiencies in the representation of the physics and improve the functional form to account for this physics. In this context, the known coupling between valence coordinates and the importance of coupling terms to describe the physics of this coupling is evaluated. Early simplified, truncated FFs introduced to allow simulations of macromolecular systems are reviewed and their subsequent improvement assessed. We examine in some depth: the basis of the reformulation of the H-bond to its current description; the early introduction of QM in FF development methodology to calculate partial charges and rotational barriers; the powerful and abundant information provided by crystal structure and energetic observables to derive and test all aspects of a FF including both nonbond and intramolecular functional forms; the combined use of QM, along with crystallography and lattice energy calculations to derive rotational barriers about ϕ and ψ; the development and results of methodologies to derive "QM FFs" by sampling the QM energy surface, either by calculating energies at hundreds of configurations, or by describing the energy surface by energies, first and second derivatives sampled over the surface; and the use of the latter to probe the validity of the representations of the physics, reveal flaws and assess improved functional forms. Research demonstrating significant effects of the flaws in the use of the improper torsion angle to represent out of plane deformations, and the standard Lorentz–Berthelot combining rules for nonbonded interactions, and the more accurate descriptions presented are also reviewed. Finally, we discuss the thorough studies involved in deriving the 2nd generation all-atom versions of the CHARMm, AMBER and OPLS FFs, and how the extensive set of observables used in these studies allowed, in the spirit of Lifson, a characterization of both the abilities, but more importantly the deficiencies in the diagonal 12-6-1 FFs used. The significant contribution made by the extensive set of observables compiled in these papers as a basis to test improved forms is noted. In the following paper, we discuss the progress in improving the FFs and representations of the physics that have been investigated in the years following the research described above.

Keywords Force fields: force field derivation · Potential functions · van der Waals · Hydrogen bond: drug discovery · Molecular dynamics · Molecular mechanics · Protein simulation · Molecular simulation · Nonbond interactions · Combination rules · Polarizability · Charge flux · Nonbond flux · Polarizability flux · Free energy · Coupling terms · Cross terms · AMBER · Charmm · OPLS · GAFF · AMOEBA · SDFF · CFF · VFF · Consistent force field · Electrostatics · Multipole moments · Quantum derivative fitting · QDF

✉ A. T. Hagler
athagler@gmail.com

Extended author information available on the last page of the article

Published online: 30 November 2018

1 Introduction

The application of energy based computational methods to molecular systems spans more than half a century. From the first it was recognized that the value of the method depended entirely on the validity of the energy surface or force field (FF) employed. Many of the questions addressed over this time span are still relevant today. For example:

> Why is the hydrogen bond energy described simply in terms of electrostatic and van der Waals interactions instead of the more complex functional forms earlier assumed? What is the physical basis for coupling between valence terms and are "cross-terms" required to account for this coupling? Do cross-terms contribute significantly to intramolecular strain energy and barriers? Why are combination rules for nonbonded constants typically calculated by arithmetic and geometric mean rules and are these valid? Are there more accurate rules and what is the evidence for them? What is the role of small-molecule crystal data in the derivation and testing of force fields? What information does it provide over that accessible from liquid and gas-phase data? What role does vibrational spectroscopic data play in derivation of force fields? Is it important for force fields designed for proteins and biomolecules to account for the spectra of model compounds? What role can Quantum Mechanics play in helping derive biomolecular force fields? How should QM results be integrated with experimental observables in the derivation and testing of FFs? How should the spatial electron distribution be represented? How important are atomic multipoles? Can spatial electron density and atomic multipoles be determined experimentally by X-ray crystallography?

Other issues also touched upon include the representation and importance of atomic polarizability, charge transfer, anisotropic nonbond interactions, charge penetration and the large effect on atomic forces of charge flux (the shift in charge on deformation of molecular geometry such as bond stretching, angle bending etc.), as well as nonbond and polarizability fluxes.

The functional form describing the energetics of these interactions have been studied over the years in various force fields, as we elaborate below. A great deal has been learned about the energetics of intra and intramolecular interactions as implied by the questions listed above. Two of the most important results of these studies are the construction of the extensive experimental data sets employed and documented in the derivation and testing of the corresponding force fields, and the documentation of where these force fields failed to adequately account for the experimental result. These data still hold great value today as much work remains in FF development. Especially valuable are the problematic observables, the accounting of which would (and should) serve as a minimal requirement for proposed improved force fields (FFs).

The evolution of FFs, how the questions above have been addressed, the approximations made, their validity, and ultimately their widespread adoption for biomolecular simulations is the subject of this review. We will focus mainly on the progression and advancements that took place in the development of biomolecular force fields. The first generation of these FFs emerged for the most part from the early workers in the field, starting, by and large, in the early 1960s and culminating, roughly, at the turn of the century. At this stage a new generation of force fields began to emerge, with two distinct approaches taken for further development. The first involved the introduction of torsion error functions, either as a Fourier series or grid correction factors, to correct for discrepancies between FF results and observed properties. The second stage involved further research into the underlying physics of inter and intramolecular interactions, and the representation of this physics in the FF. In this paper, we review the early research and discoveries, while in the following paper we discuss the second stage of development.

2 Early history

From the beginning of the field of chemistry itself, chemists realized that understanding the structure of molecules could lead to the understanding of many aspects of chemical properties and reactivity. Thus in 1865, the German chemist August Wilhelm von Hofmann [1] introduced ball and stick models fashioned from croquet balls and sticks obtained by cutting up the mallets, to demonstrate structural concepts in a lecture at the Royal Society.

The idea of three dimensional structures for molecules emerged in the mid nineteenth century. Louis Pasteur's discovery of molecular chirality and spontaneous resolution in 1848 was based on experimental observations of some crystals of organic molecules, and led to the foundation of molecular chirality [2]. Independently, a key development occurred in 1874 with the publications of the Le Bel [3] and van't Hoff [4] theoretical concept of the asymmetrical tetrahedral carbon atom, and presentations of conformations using "ball and stick" diagrams.

The still familiar space filling and Dreiding models, built to scale, made their appearance in the mid twentieth century. They were invented to assist in visualization and manipulation of molecular structure, and to elucidate the implications of these structures on chemical properties. The initial specialized, expensive and cumbersome models were

then further developed for wider availability, robustness and affordability [5–7].

The next stage in the evolution was to introduce quantitative considerations, which occurred as described below in the decades between 1930 and 1960 [8–13]. The cornerstone of energy based methods such as molecular mechanics and dynamics is the analytical expression of the energy surface as a function of the molecular coordinates—aka the force field. Classical molecular mechanics (MM) techniques are based on an empirical representation of the Born–Oppenheimer approximation, according to which the ground state of molecules is a continuous function of the atomic coordinates.

2.1 The first force fields: spectroscopy and the lost coupling terms

The earliest articulation of potential energy functions as a mathematical expression of internal coordinates significantly predated molecular mechanics calculations with papers appearing as early as the 1930s and 1940s' in the field of spectroscopy [9, 14–16].

The first applications of these methods investigated the ability of the force fields to account for the vibrational frequencies of small common molecules such as ethanol and ethyl halides [17, 18]. In the early applications the FF was molecule specific. For these applications the overall potential energy of a molecule, $E_{Valence}$, included harmonic terms representing contributions from strain in bonds (b), valence angles (θ), as well as the critical 'cross terms' which represent coupling of these internals (usually only coupling between internals with a common atom or common bond were included [9, 14]). This was commonly referred to as a "Valence Force Field (VFF)" and is defined in Eq. (1), where E_b and E_θ are the strain energies associated with bond (b) and valence angle (θ) distortions from their reference values b_0 and θ_0 and K_{bb}, $K_{b\theta,}$ and $K_{\theta\theta}$ are the force constants corresponding to the coupling between these internals.

$$E_{Valence} = E_b + E_\theta + E_{Coupling} \tag{1}$$

$$E_b = \sum_b K_b(b - b_0)^2 \tag{1a}$$

$$E_\theta = \sum_\theta K_\theta(\theta - \theta_0)^2 \tag{1b}$$

$$E_{Coupling} = \sum_{bb'} K_{bb'}(b - b_0)(b' - b'_0) + \sum_{\theta\theta'} K_{\theta\theta'}(\theta - \theta_0)(\theta' - \theta'_0) + \sum_{b\theta} K_{b\theta}(b - b_0)(\theta - \theta_0) \tag{1c}$$

Alternatively, the Urey–Bradley force field [15, 16] was utilized in which the only coupling accounted for is bond–angle coupling as represented by a 1–3 atom–atom interaction.

$$E_{Urey-Bradley} = \sum_{ij} K_{ij}(r_{ij} - r_{ij}^0)^2 \tag{2}$$

where r is the '1–3' distance, between these atoms, and K_{ij} is the Urey–Bradley force constant. Note that these early applications treated tetrahedral molecules without torsion angles.

What might be considered one of the first "modern" FFs was due to Schachtshneider and Snyder [19, 20] who addressed the alkanes and focused on the transferability of parameters (force constants). In these applications in addition to the terms in Eq. (1), torsion strain was accounted for by a cosine dependence [21] [the same terms that were to be used in contemporary conformational studies, as defined in Eq. (3)]. A pictorial representation of the various terms (as well as terms which will be discussed later) is given below in Fig. 1.

Schachtshneider and Snyder [20] also confirmed the need for the complete set of coupling terms, and showed that even the Urey–Bradley (UB) enhancement was not sufficient to account for vibrational properties—which, we note, describe the curvature of the energy surface at a given configuration. To the extent these deviate from experiment, the energy surface itself will deviate at some distance as well. These authors were also the first to include torsion angles as coordinates in the energy function, which was afterwards the common practice in force fields in general.

Although in these applications no van der Waals or exchange repulsion terms were included (or needed), the force fields for the valence terms used at that time were more complete than those used in many MD applications today. In fact, the latter diagonal force fields are a subset of these, omitting potentially important coupling interactions—known to be crucial to these early spectroscopists.

2.2 Conformational studies

Studies involving conformational properties had a different emphasis than the contemporaneous spectroscopic studies of small molecules. These studies were focused on the configuration and energies of molecules, which required torsional terms. A Fourier expansion, i.e. a series of cosines with different periodicity (n) and phase (s) were the simplest suitable functions as shown in Eq. (3), where φ is the torsion angle, K_φ is the torsion force constant, s is the phase, and n is the periodicity.

$$E_\phi = \sum_\phi K_\phi(1 + s\cos n\phi) \tag{3}$$

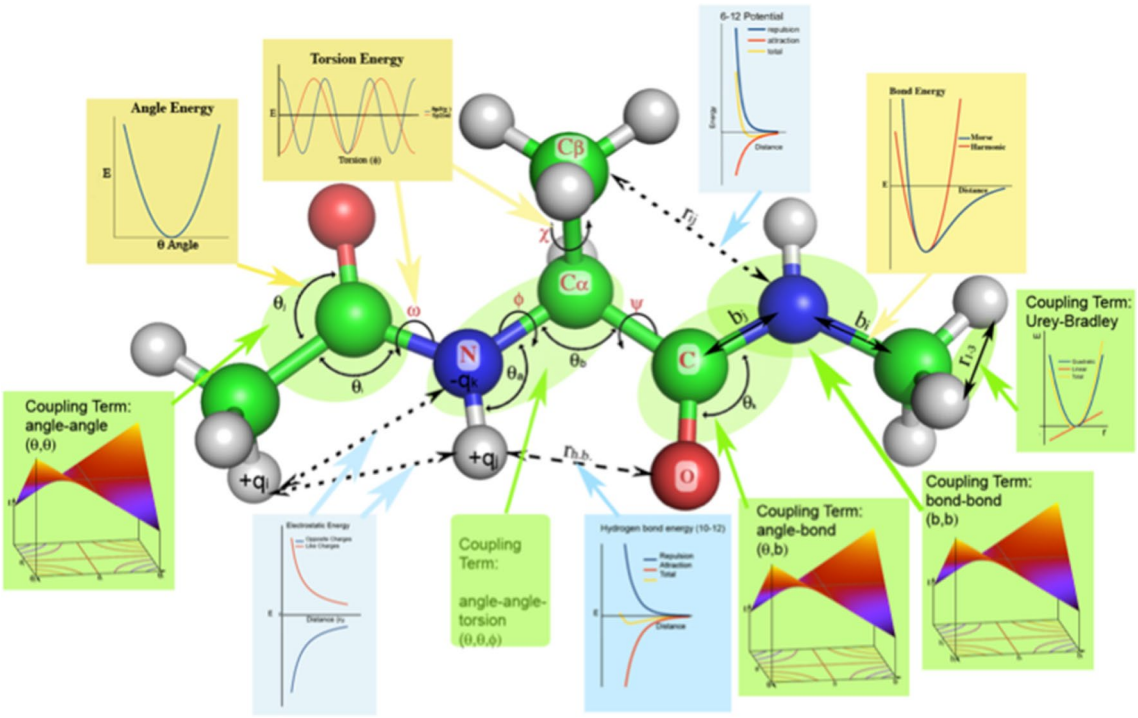


Fig. 1 Graphic representation of the main terms used in mathematical expressions defining the potential energy of a molecule (or 'force field'), demonstrated on the alanine dipeptide, the basic building block of peptides and proteins. The components depending on single internal coordinates—bonds (b), valence angles (θ), dihedral angles or torsions (φ, ϕ, ω, and χ)—are depicted in yellow rectangles. Non-bond interactions between atoms separated by more than two bonds, at a distance of r_{ij} apart, including van der Waals (e.g., 6–12), electrostatic, and explicit hydrogen bond (e.g. 10–12), are shown in blue rectangles. Cross terms, or coupling terms between two or more internals are highlighted by green ellipses and defined in the corresponding green rectangles: bond–bond (b–b)—with a common atom; angle–angle (θ–θ) and bond–angle (b–θ); and angle–angle–torsion (θ–θ–φ) with both angles being part of the torsion. Also shown is an alternative to the cross terms, the Urey–Bradley term which is a function of the interatomic distance, r_{1-3}, between atoms separated by two bonds. The standard nomenclature for atom names (N, C_α, C_β, C and O) and torsion angles (φ, ϕ, ω and χ) of an amino acid are also included and shown in red lettering

In addition, terms representing non-bonded interactions were introduced. The non-bonded energy, $E_{Non\text{-}bonded}$, was defined as a sum of pairwise interactions between atoms i and j:

$$E_{Non\text{-}bonded} = \sum_{i<j} E(r_{ij}) = \sum_{i<j} E_{vdW}(r_{ij}) + \sum_{i<j} E_{Coul}(r_{ij}) \quad (4)$$

$$E_{vdW}(r_{ij}) = E_{L-J}(r_{ij}) = \frac{A_{ij}}{r_{ij}^{n}} - \frac{C_{ij}}{r_{ij}^{m}} \quad or$$

$$E_{vdW}(r_{ij}) = \varepsilon_{ij}\left[\left(\frac{m}{n-m}\right)\left(\frac{r_{ij}^{*}}{r_{ij}}\right)^{n} - \left(\frac{n}{n-m}\right)\left(\frac{r_{ij}^{*}}{r_{ij}}\right)^{m}\right] \quad (4a)$$

Or

$$E_{vdW}(r_{ij}) = E_{Buckingham}(r_{ij}) = A_{ij}exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^{m}} \quad (4b)$$

$$E_{Coul}(r_{ij}) = \frac{q_i q_j}{r_{ij} D} \quad (4c)$$

Or

$$E_{dipole} = 1/D[\boldsymbol{\mu}_i \times \boldsymbol{\mu}_j / r^3 - 3(\boldsymbol{\mu}_i \times \boldsymbol{r})(\boldsymbol{\mu}_j \times \boldsymbol{r})/r^5] \quad (4d)$$

$$r_{ij}^{*} = 1/2(r_{ii}^{*} + r_{jj}^{*}) \quad \varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2} \quad (4e)$$

where E_{vdW} and E_{Coul} are the van der Waals, and Coulombic components of the energy of interactions between the atoms [22, 23]. The vdW component, composed of a repulsion term and an attraction term, can be defined by a Lennard-Jones function [24], Eq. (4a), or a Buckingham term [25], Eq. (4b). The attractive term represents the dispersion forces due to instantaneous dipole–induced dipole interactions. These interactions fall off as the inverse 6th power of the distance and thus m is commonly taken as six. The repulsion term represents forces due to exchange repulsion between the electrons. For the Lennard-Jones equation, values of n = 12, 9 have been used [26].

The parameters r_{ij}^* and ε_{ij} are the van der Waals radius and well depth parameters, respectively, as shown in Fig. 2.

Equation (4e) defines the commonly used arithmetic—geometric combination rules which define the values, r_{ij}^* and ε_{ij} for interactions of mixed unlike atoms *i, j* from the "like–like" parameters. Combination rules are usually used to reduce the number of independent parameters, but as discussed below this form is too simplistic and can be the source of non-trivial errors and more accurate rules are needed, and have been derived [27].

The electrostatic interactions have generally been represented by Coulomb interactions between atomic point charges q_i and q_j, (Eq. 4c), where r_{ij} is the distance between the charges, and/or by the interaction of dipoles (Eq. 4d, see also Fig. 1). This is also an approximation of questionable accuracy.

2.3 The birth of molecular mechanics and dynamics

Calculations which were the precursor to modern molecular mechanics and dynamics calculations were first introduced in 1946 by Hill [28] and Westheimer [29], and several applications were carried out shortly thereafter [10, 30]. Interestingly, Dostrovsky et al. [30] recognized and included anisotropy of the van der Waals interaction and the differences between intra and intermolecular interactions in 1946—aspects only recently "rediscovered" and still not accounted for in biomolecular simulations. Several of these studies were motivated by the desire to account for steric effects in reactions and all were carried out by hand. Westheimer described these efforts in a review in 1956 [31]. He recognized that in addition to spectroscopic properties, many of the chemical and physical properties of molecules could be calculated if we could generalize the spectroscopic force fields to a transferable set of energy expressions. Of course, the complexity of the calculation limited the application and made this a tedious undertaking. Similarly, Allinger [32] used a hybrid of model building and calculation in investigating cyclohexane conformations. He built the models by hand, measured dihedral angles directly by projection on a protractor and then calculated the energy using a torsional function as in Eq. (3). These limitations were overcome by Hendrickson, who was the first to exploit the advent of computers, and published his pioneering study of the conformation energetics of cyclic alkanes in 1961 [13]. He presented the major two components of his fundamental strategy (and to a large extent of all MM studies to follow), namely, find an appropriate mathematical function that expresses the total energy in terms of geometrical variables, and determine the set of variable values corresponding to a minimum energy (i.e. a stable conformation). It is interesting to read his evaluation of the problems facing scientists in this field at the time:

> This statement of approach to the problem serves to underscore the vast complexity of the calculations in any molecule of organic chemical interest, since the number of independent geometrical variables can be overwhelming while the calculation of the total energy of any single set of geometric parameters is itself extremely ponderous. In order to reduce the problem to workable dimensions, it is generally necessary to simplify it by certain assumptions of conformation or of parameter constancy, and frequently by simplifying or simply ignoring certain of the relevant energy functions themselves. These simplifications, however, too often so drastically alter the complexion of the problem as to render the answers either suspect or unreasonable. The basis of the present work is accordingly an effort to break through this barrier of undue simplification by employing machine calculation, thus allowing a far greater magnitude of mathematical effort in a reasonable time with untiring accuracy, and a consequent capability of a more intimate probing into these problems than is possible with hand calculation.

As we shall see his statement was a prophetic assessment of the development of the field, and remains emblematic of the dilemma facing researchers to the present. There is always an insatiable appetite to apply the techniques to molecular systems 'of interest', which are increasingly larger and more complex. Consequently, there is always a conflict between keeping the computational requirements practical and trying to eliminate severe approximations that give rise to misleading results. Though the increase in computer power since has been immense, the use of approximations of varying severity remains. This issue needs to be kept at the forefront of considerations in designing and evaluating new studies.

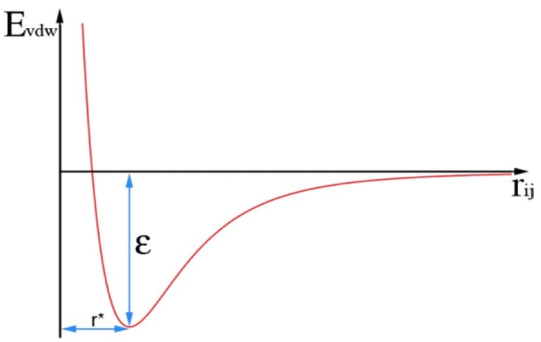


Fig. 2 Representation of Lennard-Jones potential

2.4 The introduction of Cartesian coordinates and first derivatives

The landmark calculations of Hendrickson [13] opened up new possibilities for practical expansion of molecular mechanics based research. Initially calculations were limited to hydrocarbons and utilized the simplest possible mathematical form. The first studies focused on improvement of computational techniques and of the parameters (constants) in the mathematical representation of the energy. Wiberg [33] recognized that by using Cartesian coordinates as the central structural variables, rather than internals, the calculation scheme is greatly simplified and generalized. He was also the first to implement a minimization technique, steepest descent, for finding the minimum energy structure, replacing the 'trial and error' procedure used previously by Hendrickson. In this method, an iterative process was implemented where numerical derivatives of the energy with respect to each of the Cartesian coordinates were calculated, and the Cartesians incremented by an amount proportional to the corresponding derivatives. Iterations were carried out until the energy converged. Limitations in computational power were a major consideration, and resulted in the introduction of geometrical constrains—fixed CH bond length for most of the minimization, and enforcement of symmetry.

As in Hendrickson's study, the values assigned to the various parameters were derived from available experimental data. Because the calculations were carried out for only a few molecules, and these contained only two types of atoms, it was feasible to test the parameters using a trial and error approach.

2.5 The 1960s: peptides, proteins and the first "protein" potential functions in the new age of computers

The 1960s were an exciting time, in many ways giving birth to the field of modern computational biology. Scientists such as Flory, Liquori, Ramachandran, and Scheraga realized the possibilities opened up by the emerging computer age for the investigation of the conformational properties of proteins and other biomolecules.

The studies carried out by these researchers, involving conformational properties, had a different emphasis than early structural and spectroscopic calculations. In particular, due to the biological importance of peptides and proteins there was much interest in understanding the factors defining the conformation of these molecules. Examining the covalent structure of the peptide unit it was observed that conformation of peptides and proteins could be characterized in terms of the two main chain torsion angles of each residue as shown in Fig. 1, (originally called φ, φ′ later defined as φ, ψ [34]), while other internals were approximated as rigid.

Ramachandran [35, 36] was the first to apply computational resources to the investigation of the conformational energy surface of the peptide unit. He used geometrical considerations i.e., hard sphere contact distances between the various atoms, to determine three φ, ψ regions—fully allowed, outer limits and disallowed, based on the existence of favorable short distances between hydrogen bonding atoms or the steric clashes between any other atoms (Fig. 3a). For its time the task of calculating interatomic distances as a function of the main torsion angles was a large

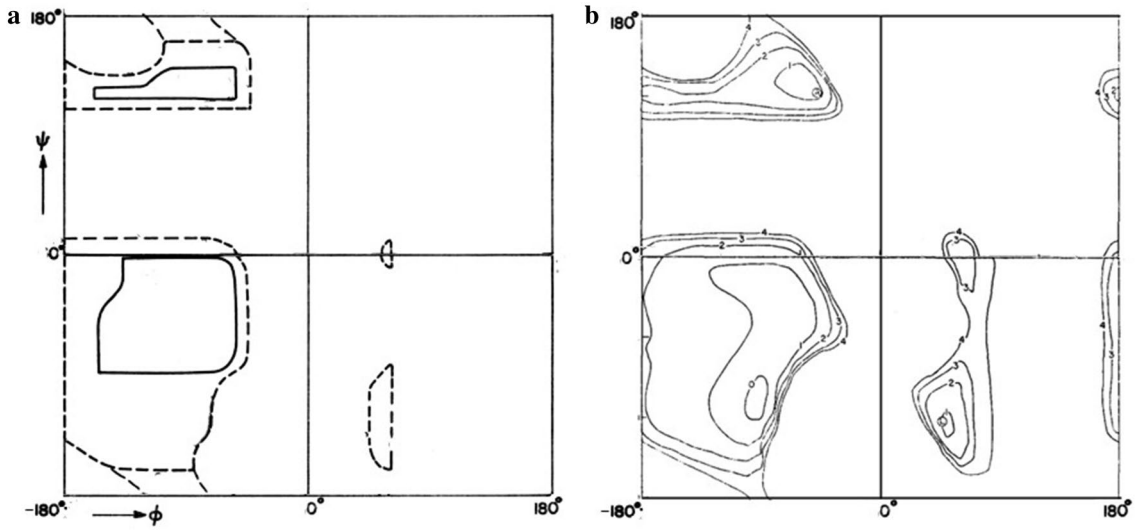


Fig. 3 φ, ψ maps for an alanine dipeptide. **a** Ramachandran [37] diagram: full lines (__) delineate fully allowed region, and dashed line (---) enclose 'outer limits' of allowed regions. **b** Scott and Scheraga [41] contour map of energy surface of poly-alanine: contours are given in kcal/mol, at intervals of 1 kcal

and challenging undertaking. In addition to working out the required analytical equations, the number of calculations needed to be carried out was large. In the first stages of this research the calculations were done using a desk calculator, but soon computers were exploited. Ramachandran's group used an Elliott-803 computer to start with [36] and then went on to IBM 1620, CDC 3600 and IBM 7090 models [37]. In a subsequent study of favorable helical conformations of a polypeptide Ramachandran considered formation of possible hydrogen bonds in addition to the hard sphere collisions. These considerations were also of a geometrical nature (i.e.—an N and an O were considered to be hydrogen bonded if they fell within 2.6–3.2 Å and the angle between N–H and N···O is less than 30°) [37]. Later studies [38–41] refined this approach, replacing the 'hard sphere' model with an explicit energy function, and presenting the results in the form of energy contour maps rather than sterically allowed contours (Fig. 3b).

In 1965, De Santis et al. [38], in what was essentially the first energy based calculations of peptide properties, introduced a van der Waals potential, and dipole–dipole interactions to account for hydrogen bonds, to calculate the energy landscape of helical peptides. The van der Waals interactions were represented by a mixture of Lennard-Jones and Buckingham functions as in Eqs. (5a, b). The minima in the maps they derived were shown to be in agreement with those derived from fiber diffraction data.

$$E_{L-J}(r_{ij}) = \frac{A_{ij}}{r_{ij}^{n}} - \frac{C_{ij}}{r_{ij}^{m}} \tag{5a}$$

$$E_{Buckingham}(r_{ij}) = A_{ij}exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^{m}} \tag{5b}$$

where the variables are defined above in Eq. (4).

The determination of intrinsic torsion parameters was problematic. This was addressed in 1965 by Scott and Scheraga [40] who introduced a method to calculate the intrinsic barriers, E_0, from knowledge of the nonbond interactions, van der Waal radii, and experimental microwave data. They defined the total energy for a rigid molecule as:

$$E_{\phi} = \sum_{\phi} K_{\phi}(1 + s\cos n\phi) \tag{6a}$$

$$E_{Non\text{-}bonded} = E_{Buckingham}(r_{ij}) + E_{Coul} = A_{ij}exp(-b_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^{m}} + \frac{q_i q_j}{r_{ij}} \tag{6b}$$

$$C_{ij} = (4\pi\varepsilon_0)^{-2}(3/2ehm_e^{-1/2}\alpha_i\alpha_j[(\alpha_i/N_i)^{1/2} + (\alpha_j/N_j)^{1/2}]^{-1} \tag{6c}$$

Here φ is the torsion angle, E_{φ} is the intrinsic torsion barrier (the barrier in the absence of nonbonded interactions), s is the phase and n is the periodicity. The nonbond term (Eq. 6b) is as given above in Eq. (4), where the dispersion constant C was determined by the Slater Kirkwood Eq. (6c). In the latter, ε_0 is the permittivity of vacuum, e and m_e are the electron charge and mass respectively, h is Planck's constant, and α_i and N_i, are the polarizability and number of effective outer-shell electrons of atom i, respectively.

Differences in the atomic electronegativities in bonds, along with molecular dipole moments were used to determine the charges, q_i, q_j in the coulomb term, while the value of B_{ii}, the exponent in the Buckingham Eq. (6b) was taken from literature values. The mixed exponent B_{ij}, was taken as the geometric mean of the homogeneous exponents. Finally, the coefficient A in the Buckingham potential was determined by requiring the 6-exp potential be a minimum at the sum of the van der Waal radii, which were taken from the literature.

This work was followed with an analysis of the energy surface of helical conformations of polyalanine and polyglycine [32] again using torsion functions, a Lennard-Jones nonbond potential and partial charges for electrostatics with parameters found previously, but they also added an explicit hydrogen bond potential based on a modified Lippincott–Schroeder potential [43–45], commonly used at the time [46] (Eq. 7).

$$\begin{aligned} E_{HB} = &-D^{*}\cos^2\theta_1 \exp\left[\frac{-n^{*}(R-r-r_0^{*})^2}{2(R-r)}\right] \\ &- D^{*}\cos^2\theta_2 \exp\left[\frac{-n^{*}(R-r-r_0^{*})^2}{2(R-r)}\right] \\ &+ A\exp(-bR) - 1/2A\left(\frac{R_0}{R}\right)^{m}\exp(-bR_0) \end{aligned} \tag{7}$$

where D^{*} is the strength of the hydrogen bond, θ_1 and θ_2 are the angles between the H–O bond and the "reference" hydrogen bonding direction taken to be 120°. R and R_0 are the actual and equilibrium N···O distances, taken to be 2.65 Å, and r and r_0 are the actual and equilibrium H···O lengths.

Based on the experimental values for the rotational barrier in acetic acid the barrier about ψ was chosen to be 0.2 kcal/mol. The barrier about φ was more ambiguous due to lack of analogous compounds, but by taking into account the relative double bond character of the carbonyl and amide bonds they concluded that the barrier should be ~0.6 kcal/mol. They tested values of both values by varying them from 0 to 1.5 kcal/mol and concluded that the torsional constants are small and did not have a significant effect on the relative stability of the left and right handed α-helices [41]. Their

map is given in Fig. 3b and compared to the hard-sphere results for the single peptide unit of Ramachandran (Fig. 3a).

Brant and Flory [39] used similar potential functions, Fourier torsion, Buckingham vdW and dipole–dipole interactions (Eqs. 8a–c), to calculate the dimensions of random coil polypeptide chains from the φ, ψ maps and showed that they agreed with experimentally determined values.

$$E_{\phi} = \sum_{\phi} K_{\phi}(1 + s\cos n\phi) \tag{8a}$$

$$E_{Buckingham}(r_{ij}) = A_{ij}exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^{m}} \tag{8b}$$

$$E_{dipole} = 1/D[\boldsymbol{\mu}_i \times \boldsymbol{\mu}_j/r^3 - 3(\boldsymbol{\mu}_i \times \boldsymbol{r})(\boldsymbol{\mu}_j \times \boldsymbol{r})/r^5] \tag{8c}$$

The first two terms in Eqs. (8a–c) are identical to those described above. In the dipole representation of the electrostatics, D is the dielectric, $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ are the peptide dipoles and r is the interdipole distance. They assigned values of 1 and 1.5 kcal/mol to the torsion barriers about φ, and ψ respectively, and similarly to the results obtained by Scott and Scheraga [41], found that their conclusions were not sensitive to variations in these constants.

The results of the studies described above suggested the existence of a small number of possible stable conformations of the peptide unit and were exciting as they gave insight into the underlying basis for the helical secondary structure observed in the then recently published structures of myoglobin and hemoglobin [47, 48], as well as other allowed residue conformations appearing in that protein crystal.

These works laid the foundations for energy based calculations of proteins and established the basic terminology to be used since, in all protein structure related research. Unfortunately, the torsion space representation is not as powerful as the Cartesian representation—or as conducive for MD simulations, and for the most part has been abandoned for modern Cartesian based simulations, though it still plays an important role in some docking applications.

3 CFF: the origin of modern force fields

At the end of the 1960s fundamental contributions to the development of molecular mechanics (MM) studies emerged from the laboratory of Shneior Lifson at the Weizmann Institute in Israel. Lifson was focused on the fundamental physical properties of organic molecules and felt it was necessary to rigorously account for these if accurate results were to be obtained. He introduced the concept of a 'consistent force field' (CFF) defining the scope and strategy for deriving and validating force fields [49, 50]. The term 'force field' was adopted from normal mode analysis of molecular vibrations as had been used in interpreting vibrational spectra. The term 'consistent' emphasized the philosophy of demanding that the functional form and parameter values simultaneously accounted for a wide variety of compound types (i.e. strained, cyclic etc.) and physical observables, in this way ensuring that the energy surface was accurate and transferable over a large domain of configurational space. The initial investigation carried out by Bixon and Lifson [51] used a diagonal FF, with contributions from bonds, valence and torsion angles and non-bonded terms:

$$E_{Valence} = E_b + E_{\theta} + E_{\phi} + E_{vdW} \tag{9}$$

$$E_b = \sum_{b} K_b(b - b_0)^2 \tag{9a}$$

$$E_{\theta} = \sum_{\theta} K_{\theta}(\theta - \theta_0)^2 \tag{9b}$$

$$E_{\phi} = \sum_{\phi} K_{\phi}(1 + s\cos n\phi) \tag{9c}$$

$$E_{vdW}(r_{ij}) = E_{Buckingham}(r_{ij}) = A_{ij}exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^{m}} \tag{9d}$$

To define the parameters, they followed a similar strategy to that of Wiberg, incorporating parameter values derived initially from individual observables, e.g., bond and angle force constants (K_b and K_a) were taken directly from spectroscopic studies, (which did not include torsion or non-bonded terms), and the zero strain values, $b_{0,}$ and θ_0 were taken as experimental values of bond lengths and valence angles. The values of the parameters were then modified one at a time to optimize the fit to the experimental data.

In further studies, with Warshel [26, 52–54] and Ermer [55], Lifson embarked on derivation of a CFF for hydrocarbons in a systematic way and introduced a novel methodology. The underlying hypothesis was that for each chemical family, a *limited* set of elementary energy functions is associated with the types of internal coordinates in the family, and should define simultaneously *all* types of observables including conformation, crystal structure, thermodynamic properties and vibrational spectra. Another way of stating this is that there is only one energy surface determined by fundamental physics, and that all the molecular properties (e.g. energy, structure, vibrations etc.) are determined by this surface. The fundamental physical laws are independent of property, phase etc. The functions have to be consistent not only with the data from which they were derived, but also with all other relevant experimental observables, and, importantly, transferable to molecules of the family which were not included in the fitting process. In principle, the functional forms are chosen so that they have a physical

rationale (e.g. strain of internals, interactions between non-bonded atoms etc.).

The 'trial and error' method for testing and optimizing parameters was abandoned in favor of a least-squares algorithm to give the best possible agreement with a large amount and variety of observed data. Analytical derivatives of the various calculated quantities with respect to the energy parameters helped to facilitate the computational procedures. In addition to the steepest descent minimization technique the Newton–Raphson technique (which involves exploitation of the Hessian) [56] was introduced, improving convergence near the minimum.

In the first papers with Warshel a Urey–Bradley functional form (UBFF) was used (Eq. 10):

$$E_{Valence} = E_b + E_\theta + E_\phi + E_{Urey-Bradley} + E_{vdW} + E_{Coul} \quad (10)$$

$$E_b = \sum_b K_b(b - b_0)^2 \quad (10a)$$

$$E_\theta = \sum_\theta K_\theta(\theta - \theta_0)^2 \quad (10b)$$

$$E_\phi = \sum_\phi K_\phi(1 + s\cos n\phi) \quad (10c)$$

$$E_{Urey-Bradley} = \sum_{ij} K_{ij}(r_{ij} - r_{ij}^0)^2 \quad (10d)$$

$$E_{vdW}(r_{ij}) = \frac{A_{ij}}{r_{ij}^n} - \frac{B_{ij}}{r_{ij}^m} \quad \text{or} \quad E_{vdW}(r_{ij}) = A_{ij}\exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^m} \quad (10e)$$

$$E_{Coul}(r_{ij}) = \frac{q_i q_j}{r_{ij}} \quad (10f)$$

where the variables are defined as above.

As a result of the streamlined methodology Warshel and Lifson could optimize the parameter values for the energy function used, and by analyzing any remaining deficiencies in accounting for the experimental data could evaluate the functional form as well as the parameters. Insight from these deficiencies yielded information on interactions unaccounted for, such as coupling terms that needed to be incorporated into the FF functional form. For example—they evaluated the merits of different forms of the van der Waals component of the non-bonded term (Eq. 10e). Values of n = 12, 9, were tested for the repulsion term, as well as an exponential (i.e. the Buckingham function [25]). The 9-6-1 potential and the exp-6-1 potential were equally successful and both were found to be superior to the 12-6-1 form. Since a polynomial function is more *efficient* computationally than an exponential, the 9-6-1 representation was adopted.

We note that the almost universally used (to date) 12th power was originally introduced simply for computational ease since it was just the square of r^6 dispersion dependence [57]. However, Warshel and Lifson found it couldn't account for intra and intermolecular interactions simultaneously.

The initial approximations were subject to reevaluation when needed. When difficulties were encountered in fitting some observables these approximations were analyzed and the missing interaction terms added, resulting in further improvements to the FF. For example a coupling term between torsions and the two angles contained in the dihedral angle (Eq. 11), was shown to be required to improve reproduction of vibrational frequencies [52].

$$E_{\theta\theta'\phi} = K_{\theta\theta'\phi}(\theta - \theta_0)(\theta' - \theta_0')(1 + s\cos n\phi) \quad (11)$$

3.1 Introducing coupling terms into an MM/MD FF

In a later work Ermer and Lifson [55] studied a series of non-conjugated olefins and introduced a valence consistent force field incorporating an extensive set of coupling terms (Eq. 12e), similar to those in traditional spectroscopic analysis, instead of the Urey–Bradley term (Eq. 10d). In addition, an out of plane term, Eq. (12d), was introduced to maintain the planarity of unsaturated groups. A total of 39 parameters were refined using a least squares procedure to fit experimental data comprised of 259 experimentally assigned frequencies, 44 structural observables; 10 *cis–trans* energy differences and excess values of heats of hydrogenation. Agreement with experiment obtained with the VFF was superior to that obtained with the UBFF [55].

$$E_{Valence} = E_b + E_\theta + E_\phi + E_{oop} + E_{Coupling} + E_{vdW} + E_{Coul} \quad (12)$$

$$E_b = \sum_b K_b(b - b_0)^2 \quad (12a)$$

$$E_\theta = \sum_\theta K_\theta(\theta - \theta_0)^2 \quad (12b)$$

$$E_\phi = \sum_\phi K_\phi(1 + s\cos n\phi) \quad (12c)$$

$$E_{oop} = \sum_\chi K_\chi \chi^2 \quad (12d)$$

$$\begin{aligned} E_{Coupling} = &\sum_{bb'} K_{bb'}(b - b_0)(b' - b_0') + \sum_{\theta\theta'} K_{\theta\theta'}(\theta - \theta_0)(\theta' - \theta_0') \\ &+ \sum_{b\theta} K_{b\theta}(b - b_0)(\theta - \theta_0) \\ &+ \sum_\phi K_{\theta\theta'\phi}(\theta - \theta_0)(\theta' - \theta_0')(1 + s\cos n\phi) \\ &+ \sum \chi\chi' K_{\chi\chi'}\chi\chi' \end{aligned} \quad (12e)$$

$$E_{vdW}(r_{ij}) = \frac{A_{ij}}{r_{ij}^{9}} - \frac{B_{ij}}{r_{ij}^{6}} \tag{12f}$$

Further investigations of the highly strained molecule tri-tert-butylmethane (TTBM) by Hagler et al. [58] with the Urey–Bradley and valence consistent force fields pointed to the need for the inclusion of an anharmonic bond stretching (Morse [58, 59]) potential in both force fields and the necessity of the stretch (CH)—bend (CCH) interaction term (which was previously omitted) in the valence FF [58].

3.2 Extension of the CFF: the first application to proteins

After the initial investigations of FF's for hydrocarbons, Warshel et al. [53] extended the CFF to include a partial set of parameters related to the amide group and its vicinity, the moiety characteristic of the peptide backbone. A Urey–Bradley FF was used, as in their original hydrocarbon FF [26] (Eq. 10), but with an added term for out-of-plane angles for the amide NH and CO groups, as in Eq. (12d). This was a much more limited study—due in part to the limited experimental data available for the amides as compared to the alkanes. The FF was optimized to fit the vibrational spectra and structures of only three simple amides (*N*-methyl formamide, acetamide, and *N*-methyl acetamide), and their deuterated analogues. It was then tested on lactams. No strain energies or crystal data was included and, unfortunately, because of the lack of low frequency vibrational data, the important energy parameters describing the φ and ψ torsion angles could not be determined. They were transferred from the work of Scott and Scheraga [41].

Finally, in a parallel application, not related to the use or derivation of force fields but rather to exploitation of the program to treat proteins, Levitt and Lifson [60] applied the software to the refinement of protein X-ray structures (Myoglobin and Lysozyme). This study could be considered as a feasibility study, reflecting the eagerness to apply molecular mechanics techniques to proteins. The goal was to refine the structure, but a rigorous CFF for proteins had not yet been derived. Instead, a very crude FF was invoked (e.g. *all* types of angle force constants in the protein were taken to be the same 30 kcal/rad^2), a diagonal quadratic FF was applied and the coordinates were restrained to their X-ray values by a quadratic restraining potential to correct defects introduced by the potential. As they stated 'No constraining force would be necessary if the force-field used was that experienced by the protein molecule'.

4 ECEPP

As noted above, the Scheraga lab became interested in the conformational analysis of polypeptides and proteins in the early 60s [41]. At this time investigations were only possible if significant approximations were made, e.g.—only dihedrals and non-bonded terms were included in the FF, bonds and angles were fixed, and amide bond were fixed as well in a trans conformation. Similarly, parameters were of an approximate nature due to limited availability of experimental information. In spite of this, the results were of qualitative interest, and served to emphasize the importance of the non-bonded terms. In an effort to refine the force-field parameters Scheraga invoked the use of crystal data of small organic compounds, and embarked on a series of studies to better refine these parameters [61, 62].

4.1 Invoking crystal structure, sublimation energy and charges from QM

In 1974 an extensive study was carried out, incorporating data from 23 crystals to derive parameters for hydrocarbons, amides, amines and acids [63, 64]. The partial charges and the attractive term of a hydrogen bond function, were computed by one of the first semi-empirical quantum mechanical methods (CNDO) [65] while the coefficients of the attractive term of the Lennard-Jones potential were computed by the Slater–Kirkwood method, Eq. (6c). The coefficients of the repulsive terms of both the Lennard-Jones and of the general hydrogen bond potentials were then obtained by varying these coefficients until the energy-minimized lattice constants agreed as closely as possible with the observed ones.

This study was extended to compute the structures and lattice energies of crystals of various amino acids [66]. Although some deviations suggested further improvements might be needed, the authors were pleased overall with the fit to experiment in most cases, and felt that this FF, termed ECEPP, (empirical conformational energy program for peptides), opened the way for conformational studies of polypeptides and proteins. Over the years ECEPP evolved culminating in ECEPP-2 being published in 1983 [67]. The improvements, for the most part were based on experimental information that became available in the intervening years. One of the simplifications introduced involved assigning only one (rather than two) set of nonbond parameters to all nitrogens, which resulted in the backbone nitrogen repulsion being softened. While the changes generally had little effect on the computed conformations of the blocked *N*-acetyl-*N*′-methylamide dipeptides of the naturally occurring amino acids, a better agreement was achieved with observed side-chain conformational distributions in peptides and proteins, and with infrared and Raman spectra. The availability of

more recent data on some sulfur-containing aromatic molecules, including direct determination of heats of sublimation, allowed a more precise determination of the nonbonded Interactions of these atoms. The new parameters improved the fit to both the unit cell parameters and the lattice energy.

The improvements to the force field continued, and in 1992 another version, ECEPP-3, was published [68]. The most significant changes were improvements in the geometry and the interactions of prolyl and hydroxyprolyl residues, on the basis of crystallographic structural data. The structure of the pyrrolidine ring was revised to correspond to the experimentally determined extent of out-of-plane puckering of the ring. The parameters for nonbonded interactions involving atoms in Pro and Hyp were also modified, which resulted in improvements in the computed minimum-energy conformations of peptides containing the Pro–Pro and Ala–Pro sequences. In particular, it was demonstrated that an α-helix-like conformation of a residue preceding Pro became an accessible state. This had important implications for calculations of collagen-like polypeptides.

5 MM1 and MM2: coupling terms, hydrogen electron offset, anharmonic bond stretching and bending, and Taylor series for torsions

In the late 1960s, Allinger also noted the implications of Hendrickson's work [13] and launched a similar study [69] on a group of hydrocarbons. Allinger exploited the technology from the viewpoint of an organic chemist. His goal was to be able to accurately calculate the structures of a broad range of organic compounds and to be able to relate these structures to their physical properties. In this first paper Allinger reparametrized a simple diagonal force field for ~20 hydrocarbons and their isomers, and was able to calculate bond lengths, angles, and relative energies of different conformations in reasonable agreement with experiment. Continuing studies led him to develop an improved hydrocarbon FF—MM1 in which he introduced an angle-bend coupling term and offset the electron density of the hydrogen atom by 8%, so that the electron cloud of this atom is positioned 0.92 Å from the bonded atom [70]. Again a wide range of structural and energetic properties were calculated [71]. This work established an approach, similar to Lifson's CFF protocol, which he would follow for FF optimization over many decades: namely he invoked a large set of compounds, ~20 or more, and demanded that the FF account for a variety of fundamental physical properties of these compounds. When deviations arose over subsequent applications, they were collected, and then used to inform the derivation of an improved FF, usually involving additional terms to account for the physics of interactions or coupling previously omitted.

5.1 Systematically collecting and analyzing discrepancies to consistently improve FF functional form and force constants

In the subsequent years numerous studies were carried out where MM1 was applied to a diverse set of organic moieties [71]. By 1977 experience with deficiencies in the FF as reflected in problems arising in energetic and structural properties led to further significant modificationd of the force field and the release of MM2 [72]. The main modifications to the FF included: adding onefold and twofold (K_1 and K_2) terms to the torsion potential; removing the combination rule for C/H energy interactions so they were independent parameters; adding an anharmonic sixth power term, $k(\theta–\theta_0)^6$ to the quadratic angle deformation energy in the angle bending term; treating bond stretching anharmonically with a cubic term added to the simple quadratic bond stretching energy. The bond–angle cross term was retained, and all force constants including the nonbonded parameters were reoptimized. The resulting force field is given in Eq. (13).

$$E = E_b + E_\theta + E_\phi + E_{Coupling} + E_{Non\text{-}bonded} \tag{13}$$

$$E_b = \sum_b K_b\left[\left(b - b_0\right)^2 - 2\left(b - b_0\right)^3\right] \tag{13a}$$

$$E_\theta = \sum_\theta K_\theta\left[\left(\theta - \theta_0\right)^2 + 7 \times 10^{-8}\left(\theta - \theta_0\right)^6\right] \tag{13b}$$

$$E_\phi = \sum_\phi K_{\phi 1}(1 + \cos\phi) + K_{\phi 2}(1 - \cos 2\phi) + K_{\phi 3}(1 + \cos 3\phi) \tag{13c}$$

$$E_{Coupling} = \sum_{b\theta} K_{b\theta}(b - b_0)(\theta - \theta_0) \tag{13d}$$

$$E_{Non\text{-}bonded} = E_{Buckingham}(r_{ij}) = A_{ij}\exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^m} \tag{13e}$$

No electrostatic terms were employed for hydrocarbons, and dipoles rather than point charges were assigned to bonds of polar groups. Their rational for the latter was mainly practical—there are fewer dipole points than partial charge centers, and dipole interactions converge faster. In addition to an improved fit to structural properties of alkanes, the calculated heats of formation of 42 selected diverse types of hydrocarbons, yielded a standard deviation from experiment of only 0.42 kcal/mol with the improved FF.

They later extended their force field to additional functional groups, namely, thiols and thiaalkanes [73]

disulphides [74] sulphoxides [75], ethers and alcohols [76], acids and esters [77]. Although reasonable fits were obtained for all groups, there were limitations on the accuracy mainly due to more limited availability of experimental data. For example, they reported that though the structures and energies of alcohol and ethers [76] were reasonable, the agreement was not as good as that obtained for hydrocarbons.

6 CFF for peptides and proteins: deriving the Hagler–Lifson electrostatic-non-bond representation of the hydrogen bond

From the earliest conformational studies, it became clear that the non-bonded terms are of crucial importance. Therefore, the attention of several research groups turned to an invaluable source of experimental data, structural and thermodynamic properties of crystals. In the mid 1960s pioneering studies were carried out by Williams [78, 79] for aromatic and non-aromatic hydrocarbons, by Kitaigorodskii [80–84] for a variety of organic molecules and by Scheraga [61, 62] for benzene and polyaminoacids.

In the early '70s Hagler and Lifson embarked on a study to investigate the hydrogen bond and derive non-bonded parameters for amides and acids [85–88]. They exploited the wealth of data available from crystals as elucidated in the studies noted above. Though it is now generally accepted that the hydrogen bond is essentially electrostatic in nature and can be represented by Coulomb and van der Waals interactions this was far from the case before the mid 1970s. The most popular function used to describe the hydrogen-bond at the time was the Lippincott–Schroeder equation [41, 43, 44] alluded to above (Eq. 7). This was widely used, in various modified forms, especially with respect to angular dependence [22, 45, 46, 89].

We developed a novel procedure for obtaining parameters from crystal data based on minimizing the estimated deviation in lattice parameters from their experimental value as obtained by the Newton Raphson method [87]. This obviated the need to minimize the crystal structure, with lattice sums being calculated only once at the experimental structure, thus greatly enhancing the efficiency of the optimization. The method was then used to derive an intermolecular force field for amides [85, 86] by fitting a set of observables consisting of the unit cell vectors of eight amide crystals, six sublimation energies and three dipole moments. The molecules were kept rigid so that the nonbonded forces could be isolated. Thus, the force field optimized was of the form:

$$E_{Non\text{-}bonded} = \sum_{i<j} E(r_{ij}) = \sum_{i<j} E_{vdW}(r_{ij}) + \sum_{i<j} E_{Coul}(r_{ij}) + \sum_{k<l} E_{HB}(r_{kl}) \tag{14}$$

$$E_{vdW}(r_{ij}) = \frac{A_{ij}}{r_{ij}^{n}} - \frac{C_{ij}}{r_{ij}^{m}} \quad \text{or}$$

$$E_{vdW}(r_{ij}) = E_{L-J}(r_{ij}) = \varepsilon_{ij}\left[\left(\frac{m}{n-m}\right)\left(\frac{r_{ij}^{*}}{r_{ij}}\right)^{n} - \left(\frac{n}{n-m}\right)\left(\frac{r_{ij}^{*}}{r_{ij}}\right)^{m}\right] \tag{14a}$$

$$E_{Coul}(r_{ij}) = \frac{q_i q_j}{r_{ij}} \tag{14b}$$

where r_{ij} are the interatomic distances, q are partial atomic charges, and A and C are the repulsive and attractive parameters in the van der Waal energy potential. The attractive term was represented by the standard inverse 6th power distance dependence ($m = 6$), while values of $n = 12$, 9 were tested for the exchange repulsion term. E_{HB} represents a potential describing the hydrogen bond.

6.1 Redefining the functional form for the hydrogen bond: the origin of the electrostatic-van der Waals representation

Several hydrogen bond functions, which had been proposed in the literature, including the Lippincott–Schroeder function (Eq. 7), were tested in the optimization. Even though additional parameters were available in these representations of the hydrogen bond, none of them improved the fit. Thus, one of the most surprising results that emerged from this study was that no explicit function was needed to represent hydrogen bonding in these crystals. Instead, the features of hydrogen bonding resulted naturally from the usual non-bonded and electrostatic energy functions. What enabled the hydrogen bond interaction was the negligible radius derived for the amide hydrogen, which allows a short NH···CO contact distance, resulting in a strong electrostatic interaction. Although essentially universally adopted today, it was awhile before this was generally recognized in the field of protein simulations and as recently as the 1990s explicit HB functions were still being employed [90].

In addition, parameters for both 12–6 and 9–6 van der Waals functions were derived and both yielded similar fits to experiment. They could not be differentiated by this set of data. Interestingly, the non-bonded van der Waals radii, r^*, derived with no a priori assumptions with regard to minimum energy distances or radii, were consistently larger than those used at the time for conformational calculations. This was a natural manifestation of compression due to intermolecular forces on the observed contact distances.

6.2 On the power of crystal systems to assess the validity, and more importantly the deficiencies in force fields

The force field was further tested by full lattice minimization on these amides as well as two that were not in the training set. Minimization was carried out with respect to the lattice parameters and translation and rotation of all molecules in the unit cell, without imposing symmetry. Several conclusions were reached which helped clarify the physical nature of the derived FF. In all cases the newly derived FF performed as well or significantly better than others which included explicit H-bonding potentials such as the Lippincott–Schroeder [44] or Stockmayer [91, 92] potentials, while the representation was simpler and involved fewer parameters.

Of more importance, however, is the ability provided by these systems to analyze the discrepancies of the calculated structures from the well-defined experimental structure in these relatively simple systems. Here the origin of a deficiency in the FF representation may often be indicated by the simple nature of the structural deviation. Compared, for example, with protein systems where a deviation in a hydrogen bond length might be due to its functional representation, by packing considerations of the secondary structures, by solvent interactions, or by problems with counter-ions etc., the errors in these high resolution, simpler systems are much easier to analyze. They also provide a much more straightforward set of systems to test modifications of the interatomic interactions proposed to account for the deficiencies.

This situation may also be compared with calculated deviations in liquid properties, which don't give structural insight into the nature of the underlying defect. (We emphasize that this is not meant to say any one method of deriving or assessing potential functions is superior in all respects or sufficient onto itself—rather they are complementary and optimally, at this point in time, should all be used in concert, along with high level QM results available today, as per Lifson's protocol).

An example of the insight available from crystal packing analysis is the result obtained from minimization of the formamide crystal [86]. A comparison of the minimized structure with the experimental structure of the formamide crystal is given in Fig. 4a.

As can be seen from the figure the H···O=C angle has opened up from approximately 125° in the observed structure to ~ 140° in the minimized structure. These results suggest that the lone pairs may indeed be important and that the reason for the discrepancy in formamide may be due to the omission of the effect of these orbitals from the force field. Significantly, the deviation in angle is essentially independent of whether the 6–12 or 6–9 potential is used thus indicating it is not in some way due to the van der Waals interactions. In addition, the trend in the H···O=C angles for all the ten amide crystals minimized agrees with the hypothesis based on the analysis of formamide that the orbital position may play an important role in determining the H···O=C angles, further demonstrating the power of crystal structures to yield information on the accuracy and deficiencies in potential functions. This analysis has been borne out in more recent studies where it was found that the directionality of the H-bond can be accounted for by inclusion of higher order atomic moments, including the quadrupole moment, effectively reflecting the anisotropic atomic electron density, including the lone pair orbitals [93–98].

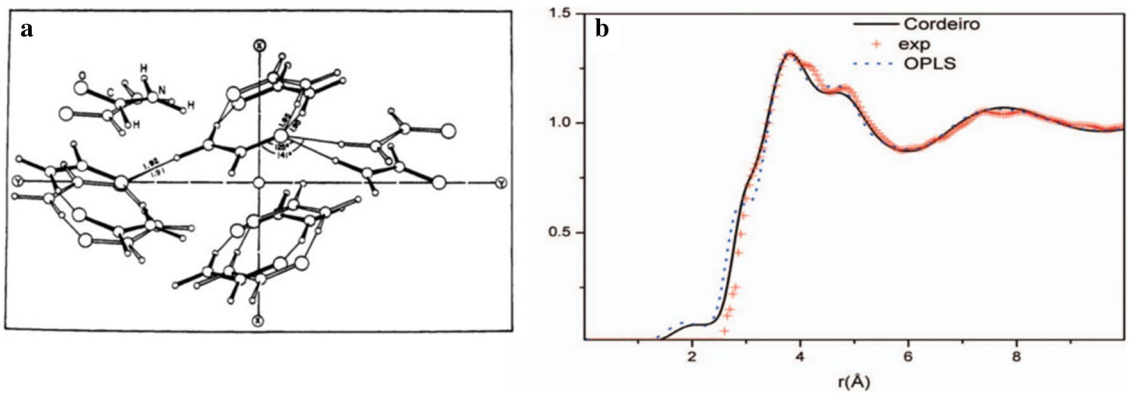


Fig. 4 The nature of information contained in liquid observables compared to that in the crystal. **a** A comparison of the observed and calculated crystal structures of formamide. The molecules in the observed structure have solid bonds, while the bonds of the molecules in the minimized structure are open. Hydrogen bonds are shown by single lines. **b** Radial distribution function G(r) of liquid formamide from X-ray diffraction and simulation (cross: experiment, solid line: Cordeiro model, dashed line: OPLS model). From Bako et al. [99]

6.2.1 Crystal versus liquid observables

The value of crystal data in deriving and validating FFs, as compared to liquid data can be appreciated by comparing the information intrinsic to the structural data in Fig. 4a with that available from the use of liquid formamide as an observable. In Fig. 4b the results of liquid formamide simulations of Bako et al. [99] are presented, where calculated radial distribution functions are compared with experimental.

Clearly, crystal data provides a far more precise probe into the nature of both intermolecular interactions and underlying deficiencies in FF representations. This is analogous to the revolution in understanding of function provided by elucidation of protein structures, as opposed to the previous information provided by solution properties of proteins such as by light scattering. We note that in addition to structure, liquid data invoked in FF optimization mainly consists of density and heats of vaporization. The analogous information provided by crystals consist of unit cell vectors, and cell volume, again the cell vectors providing more information than the single density observable from liquids. In addition, as noted elsewhere, if there are deviations in the liquid properties, the structural origins of these discrepancies are uncertain, whereas in crystals, if the sublimation energy is in error, one can interrogate the corresponding structure as e.g. in Fig. 4a, to gain insight into possible errors in the FF. Perhaps more importantly, if the energy and densities agree, one can readily ascertain that these correspond to the experimental structure in the case of crystals, while in liquids the thermodynamic observables though in agreement with experiment, may correspond to an erroneous structure.

As noted above, the best scenario is to test FFs against all possible properties, but if a choice must be made, far more information, at a lower computational cost, can be (and should be) obtained by invoking crystal structure and energetic observable for the testing and validation of FFs. These data are in general as available if not more so, as liquid data (perhaps with the exception of sublimation energies, which if the case, should certainly be supplemented by heats of vaporization.)

A similar methodology was then applied to the carboxylic acids [88, 100, 101]. The strategy adopted was to examine the transferability of the energy functions from amides to carboxylic acids, by seeking the smallest number of additions to the amide FF that would account for both carboxylic acids and amides. Crystal structure parameters, sublimation energies, and dipole moments, taken from a set of 14 mono- and di-carboxylic acids, were used to derive the force field, and again, both 9–6 and 12–6 Lennard-Jones potentials were considered. Surprisingly, it was found that the 110 observables could be fit with only a single adjustable parameter, the charge on the hydroxyl hydrogen. Optimizing additional force constants rather than transferring them from the amides did not lead to a significant improvement in fit. Importantly, the hydrogen bond of carboxylic acids like that in amides required no special function to describe it, being reasonably represented as a balance between van der Waals and Coulomb interactions between two polar moieties, and having negligibly small van der Waals parameters for the hydroxyl hydrogen. The difference between NH···O and OH···O hydrogen bond lengths and energies emerged naturally as a result of the different van der Waals radii of N and O.

Following the derivation of the FF Hagler et al. [101] introduced the concept of constructing a benchmark of crystal properties for evaluating alternative FFs and using this to assess deficiencies and advantages, of different analytical representations. In this pilot study a comparison of the CFF 6–9 and 6–12 functional forms and the MCMS FF derived by Momany et al. [63], which utilized a 12–10 hydrogen bond function, was carried out. The concept is of general importance for several reasons. The first is that those who would like to select one of the available FFs at any given time should be able to evaluate the alternatives against a common set of high resolution, high information content, data, appraising for example the tradeoff between accuracy and computational efficiency. This is important in that different FFs are often derived in different studies used different molecules, methods of derivation and evaluation of the functions, making comparison difficult. In addition, such a comparison can also be helpful in investigating the physical meaningfulness of various terms and approximations in the different FFs and the consequences on the calculated properties of these well-defined systems. Finally, a benchmark can help pinpoint artifacts of a particular force field and deviations common to various FFs. Analysis of these trends can suggest ways to further improve a FF. The benchmark constructed for this first study contained an extensive set of crystal structures, sublimation energies, and dimerization energies derived from 14 carboxylic acid and 12 amide crystals. (This set included all the data included in the derivation of the FFs as well as additional crystals as a further test of the functions [101]).

The utility of such an analysis is exemplified by the results summarized in Table 1 [101]. For example, as seen from Table 1, on the whole the 9–6 Lennard-Jones CFF gives the best overall fit to experiment; Consistent with our previous observations [85], inclusion of an explicit HB function (in the MCMS FF) did not result in a better fit to experiment in either the acids or amide. In general, the calculated amide crystal properties gave a better fit to the experimental structures than those of the carboxylic acid in all force fields, probably because the amide group forms more extended networks of hydrogen bonds in the crystals. Perhaps of most importance, there was a significant correlation of the degree of fit of most individual crystals between all force fields. The

Table 1 Root mean square deviations of properties calculated for carboxylic acids and amides by the various force fields

Property	Units	No. of terms	RMSD		
			12-6-1	9-6-1	MCMS
Acids					
Energy	kcal/mol	12	2.486	2.053	2.118
UCV length	Å	42	0.489	0.307	0.604
UCV angle	deg	17	3.456	2.856	4.465
Volume	Å	14	15.911	16.772	18.876
D < 4	Å	14	0.247	0.190	0.322
H···O dist	Å	16	0.062	0.072	0.058
O···O dist	Å	16	0.047	0.071	0.041
C–O···O angle	deg	16	11.071	9.881	14.048
O···O=C angle	deg	16	7.843	7.760	11.786
H···O=C angle	deg	16	12.362	12.144	17.985
180-O–H···O	deg	16	8.491	7.732	11.710
Amides					
Energy	kcal/mol	6	1.574	1.930	8.446
UCV length	Å	36	0.208	0.235	0.261
UCV angle	deg	14	1.824	1.261	2.385
Volume	Å	12	7.057	17.797	13.951
D < 4	Å	12	0.145	0.145	0.164
H···O dist	Å	30	0.049	0.059	0.056
N···O dist	Å	30	0.055	0.055	0.076
C–N···O angle	deg	22	3.337	3.575	4.071
N···O=C angle	deg	22	5.931	5.502	9.257
H···O=C angle	deg	30	5.830	5.609	7.329
180-N–H···O	deg	30	4.396	3.894	4.093

latter indicates common flaws and strengths in the representations and thus especially gave insights into deficiencies in the force field, which could lead to their improvement, either, for example, in functional form, or possibly on limitations to transferability.

An example is the β-oxalic acid crystal. The deviations in the sublimation energy of this crystal were −3.3, −5.1, −5.1 kcal/mol in the 9–6, 12–6 and MCMS potentials respectively. These deviations of between 15 and 20% were significantly larger than the other crystals where the deviations were around 2 kcal/mol or ~10% (Table 1). Thus, this pointed to oxalic acid as a test for putative improvements in FFs, such as inclusion of lone pair orbitals, polarization, anisotropic charge distributions etc. Another characteristic which came out of the comparison was the nature of deviations in H-bond geometry. Although the RMS deviations in H-bond length obtained with the CFF and MCMS were comparable (Table 1), inspection of the individual crystals showed dramatically different behavior. Thus, the calculated H-bond distances resulting from the MCMS potential were essentially all the same, independent of crystal packing. Thus, the variation in H-bonds in the individual crystals was not reproduced and it was clear from this result that the 12–10 H-bond potential was too stiff. The deviations observed with the CFF, on the other hand were "random" and, based on the formamide analysis, may indicate a need for a better representation of the electrostatics, or other the factors enumerated above.

6.3 Extrapolation to deviations to be expected in protein simulations

Finally, this type of analysis provides some indication of the "errors" to be expected in applying these potentials to other systems. Though it is difficult to make a direct extrapolation, since the potentials were derived from crystal systems and these are in many cases relatively simple it would seem clear that when applying such potentials to other systems, for example, proteins, one should not expect to get a better representation of the energy and structure than the deviations obtained for the small molecules. Thus, one should expect, minimally, root mean square deviations in interatomic distances, and energies of the order of 0.3–0.4 Å and ~10% (or ~2 kcal/residue) due to the potentials alone. Other complicating factors such as solvent effects, deviations in secondary structure packing, effects of less well parametrized functional groups, counter ion effects, etc. would likely worsen the deviations.

6.4 Inefficiencies in FF development: the H-bond, a case study

Interestingly, though the Hagler–Lifson (H–L) representation of the H-bond (vdW-electrostatic with a small to negligible polar hydrogen exchange repulsion) is now essentially universally adopted, explicit representations persisted for some years [102]. An example is the 12–10 H-bond representation introduced by McGuire et al. based on semiempirical CNDO/2 calculations [103]. Its use extended through the 1980s [104], and was only abandoned in the ECEPP potential in 2006 [105].

The 12–10 form was also adopted in 1984 in the AMBER FF [106], only to be abandoned for the H–L representation in the 2nd generation AMBER FF in 1995 [107]. The 12–10 H-bond potential was also used in the early applications of CHARMM [108], and later in 1983 an orientation dependent function, similar to the Lippincott–Schroder H-bond potential was introduced [109]. These forms were both abandoned in 1996 [110], replaced by the H–L formalism.

These derivations of H-bond functions which were subsequently abandoned clearly represent a misspent effort and, along with subsequent applications using these forms, now called into question, constitute a significant expenditure of resources that could have been better deployed. This is not

because the H-bond functions were flawed, but rather that the Lifson paradigm, which would prevent such dead-end diversions, were violated. All analytical representations are approximations and systematic improvement needs to be implemented. Examination of deficiencies in fitting experimental and ab initio data inform on shortcomings in the representation of the physics, and can be used to improve this representation as shown by Lifson and Warshel [52] and Allinger, who continually improved the physics accounted for on going from MM2 [72] to MM3 [111] and then to MM4 [112]. The key is systematically improving FFs without taking spurious spurs which lead to dead ends requiring U-turns to the original point of departure. Lifson's paradigm (and Allingers methodology) ensures this, i.e. by parameterizing against a wide variety of compounds and physical observables, including all used in deriving previous iterations of the FF. For example, the AMBER FF of 1984 [106] modified the H–L [86] nonbond parameters, and introduced a 12–10 H-bond potential, based on qualitative considerations, and semi-empirical and minimal basis set HF calculations, and tested them by comparing with PSILO φ–ψ maps. However, the original data set of crystal structure and energies, as well as a subsequent benchmark [101]. Far more powerful probes of these interactions were not included nor were the liquid amide densities and vaporization energies exploited by Jorgensen. Lifson's paradigm leads to inclusion of both of these observable sets, as well as the additional observables introduced in the AMBER study. Inclusion of these observables would have revealed the flaws in this potential representation, allowed immediate correction, and avoided applications of the flawed form and the need for subsequent reparametrizations.

Similar considerations hold for the other functional forms incorporated in the FFs cited above, all of which have since been discarded. The goal is to make systematic incremental improvement in each revision of a FF, without dead ends such as discussed above, resulting from inadequate training and validation sets, such as in Allingers continued improvement of his MMx FFs. This is enabled by including the broad range of data prescribed by Lifson. This includes *all* data used to derive the previous versions of the FF that is being improved. It is basic good scientific practice to demonstrate that a force field that is purported to improve on, and be superior to another FF, whether a previous version or competing version, reproduces the observables used in parametrizing and validating the original FF better than that FF.

Unfortunately, a similar situation has persisted in FF derivation over the past decades, as described in the second part of this perspective [113]. Though it takes a small investment in preparing and calculating the broad set of observables described above, the return is huge, preventing the wasted effort inherent in the subsequent reparametrizations, and applications carried out with the soon to be discarded FFs.

7 Early insight from quantum mechanics on the functional representation of potential energy surfaces

It became clear rather early on that quantum mechanics was a valuable resource and could provide complementary information about the nature of both intra- and intermolecular energy surfaces. The first applications of quantum mechanics to biological systems were carried out by the Pullmans beginning in the late 1950s when they began studying nucleic acid bases [114, 115] by the Hückel method developed in 1930 [116]. In 1963 they turned their attention to amino acids, calculating partial charges by the Del Re method [117, 118]. As often occurs in science, converging developments reach a tipping point and different groups initiate similar studies almost simultaneously. An example of this occurred in1969, arising from the confluence of the emerging interest in proteins and peptides in the early to mid 1960s, with developments in semi-empirical quantum methods such as Extended Hückel Theory (EHT) and Complete Neglect of Differential Overlap (CNDO) in the mid 1960s [65, 119, 120]. Thus, before 1969 there had been no publications on the applications of QM to peptide conformation, while just 3–4 years later there were at least 30 from different labs in disparate parts of the world.

The first study was reported by Roald Hoffman on calculations of phi–psi energy maps for Glycine and Alanine residues using his EHT [121]. At roughly the same time 1969 Rossi et al. [122] applied EHT to the calculation of the relative energies of tetra and pentapeptides of Gly in the α-helical and extended conformations, while Kier and George [123] exploited EHT to calculate the preferred conformations, of glycine, alanine, proline, and phenylalanine residues. In the following 2 years Govil published a series of five papers in which he applied EHT to the conformational preferences of trans and cis Ala and Gly peptides [124, 125], the conformational preferences of side chains [126] and both EHT and CNDO to the effect of H-bonds and chain length on secondary structure [127], as well as structural preferences of Ser and Thr [128]. The Scheraga lab also embarked on a series of QM investigations of various aspects of peptides in 1970, going beyond simply the calculation of phi–psi maps. In collaboration with Hoffmann they applied EHT and CNDO/2 methods to compute charge distributions, dipole moments, energies for internal rotation, and electron spectra of a series of amide and peptide analogs [129]. In subsequent papers, they applied the semiempirical QM techniques to the effect of hydrogen bonding on charge distributions, and for the first time determined analytical potentials from QM data. They investigated the barrier to rotation about the peptide bond [42], barriers to rotation in model compounds of aspartic acid [130], the conformation of Ala and Gly di- and

tri-peptides [131], and the hydrogen-bond potential (resulting in the $r^{-12} - r^{-10}$ form) [103]. Their CNDO/2 studies of hydrogen-bonded amide complexes was one of the first demonstrations of the cooperativity of the hydrogen-bond [42].

During the same period the Pullman lab exploited the (then) recently developed Perturbative Configuration Interaction using Localized Orbitals (PCILO) method [132] to carry out similar studies on the energy surfaces of the dipeptide model. In the first papers by Maigret et al. the phi–psi maps of Ala and Gly dipeptides [133] and Ala preceding a Pro residue [134] were calculated. These were then followed in the next few years by similar PCILO calculations of essentially all 20 amino acids [135]. Numerous refinements of semi empirical methods appeared over the years, beyond the scope of this discussion. They were reviewed by Beveridge and Pople in their 1970 book [65], and later developments, extending to the evolution of Hartree Fock and DFT algorithms are also recounted in a recent fascinating history of the Quantum Chemistry Program Exchange (QCPE) by Boyd [136].

7.1 The genesis of modern rigorous ab initio methods applications to biomolecular systems: Clementi

While the major development and applications of QM to organic systems in the 1960s and early 1970s involved semi-empirical methods, Clementi, one of the pioneers in the development and application of molecular orbital techniques, was focusing his efforts on more rigorous approaches [137–139]. In 1967, while at IBM, he developed a program for carrying out self-consistent field molecular orbital (SCF-MO) calculations invoking the now standard approximation of expressing the molecular orbitals as a linear combination of atomic orbitals (LCAO), which were themselves represented as a set of Gaussian functions [140]. This program was deposited the same year, with the QCPE, the major platform for the exchange of computational chemistry programs at the time [136]. Clementi had an early interest in the properties of biomolecules. His first studies in the late 1960s early 1970s involved the electronic structure and hydrogen bonding of the nucleic acid bases [141, 142]. He also had an early and abiding interest in the properties of water and its interactions in molecular complexes, carrying out some of the first studies on these systems [143–145]. Clementi, in the first application of exploiting extensive QM calculations to the derivation of intermolecular potential functions, went on to determine and apply intermolecular force fields for the interactions of water–water and water–peptide systems from ab initio calculations, in parallel with the continued development of the MO techniques. In 1973 his lab derived a water–water potential by fitting the calculated ab initio energies of 216 different configurations of the water dimer [143]. They invoked a large basis set, especially for that time, consisting of 13s, 8p, 2d, and 1f functions on the oxygen and 6s, 2p, and 1d functions on the hydrogen in order to obtain results close to the Hartree–Fock (HF) limit. Analysis of the results indicated that polarization played a major role in the energetics of the dimer, especially at intermolecular distances in the neighborhood of the van der Waal contact.

It is noteworthy that several analytical potential models were tested by fitting the energies of these 216 configurations, and the best results were obtained with a charge distribution corresponding to the Bernal–Fowler model [8] (see Fig. 9a), and the analytical form used

$$E = A\ \exp(-br) + q_i q_j / r_{ij} \tag{15}$$

corresponded to a Buckingham (exponential) repulsion (Eq. 5b) along with a coulomb electrostatic term (Eq. 4c—with D = 1). No dispersion interaction was included. Monte-Carlo (MC) calculations of the liquid represented by 27 water molecules in a cubic box were carried out with this potential and the Barker–Watts program [146]. Comparisons with experimental H–H, H–O, and O–O pair correlation functions [147] showed reasonable agreement. This was followed in 1975 by the inclusion of dispersion forces obtained by a perturbation technique [148]. Interestingly it was found that one needed not only the inverse 6th power of distance to describe the dispersion interactions, but also inverse 8th and 10th powers, corresponding to higher order instantaneous multipole interactions. Though long known to exist, the need and magnitude for such higher order terms, has not been explored further in the biomolecular simulation community.

7.2 The NCC and NCC: vib water potential—one of the most rigorously derived and validated water potentials to date

This potential was again used to carry out MC of the liquid, this time including 343 water molecules, and again compared to experimental results, including neutron scattering and X-ray scattering intensities at different scattering angles [148]. Over the following years Clementi continued to improve the representation of the water potential, carrying out ever more extensive quantum mechanical calculations, including a configuration interaction study of 66 water dimer complexes. This culminated in 1986 with the derivation of a flexible water potential (NCC) constructed from the configuration interaction potential (MCY potential) of Matsuoka et al. [145] with the intramolecular potential derived by Bartlett et al. [149]. The latter is a quartic potential containing all possible bond angle coupling/cross terms up to fourth order, derived by fitting the energies of 22 symmetric and 14 asymmetric configurations of the water molecule [150]. This potential was used in an MD simulation in which numerous

properties of the liquid were calculated and compared with experiment—including internal energy, heat capacity, pressure, geometry, radial distribution functions and vibrational frequencies among others [150]. These studies were followed by Corongiu and Clementi whose development of water potentials and application to probing an extensive array of experimental thermodynamic and kinetic properties spanned some 20 years. Both the QM calculations on which they were based, and the derived potentials became increasingly rigorous, ultimately resulting in the NCC-vib flexible and polarizable potential [151]. These potentials were used in numerous studies of both liquid water and ice, probing an impressive array of thermodynamic and kinetic properties and comparing with experimental data—see papers by Corongiu and Clementi [151–154] and ref therein.

Simultaneously with the water studies the Clementi group began to derive potentials for the interaction of water with biomolecules, using similar techniques. In 1977 they presented a derivation of the Lennard-Jones 6th and 12th power coefficients and partial charges to describe water–amino acid interactions for 21 amino acids [155]. These were obtained by fitting the parameters in a L–J plus Coulomb potential to the SCF-LCAO-MO energies of1690 water amino acid complexes. In the same year they went on to exploit these potentials to calculate contour maps of the interaction energy of a single water molecule with Lysozyme [155]. Clementi also used his platform to calculate and apply potentials for the interaction of water with nucleic acids obtaining energy contours for these systems as well [156–158]. Finally in 1992 he and Corongiu extended the method to the derivation of a nonbonded force field for protein–protein, protein–nucleic acid and nucleic acid–nucleic acid nonbonded interactions [159]. To accomplish this 30,000 SCF interaction energies of model compounds were fit to an optimized analytical expression. Interestingly the form of the atom–atom interaction (where the total energy was the sum over all atom pairs in the interacting model compounds) was found to be:

$$E_{ij} = \frac{C_{10}}{r_{ij}^{10}} + \frac{C_4}{r_{ij}^4} + \frac{C_2}{r_{ij}^2} + \frac{C_0}{r_{ij}} \ldots \tag{16}$$

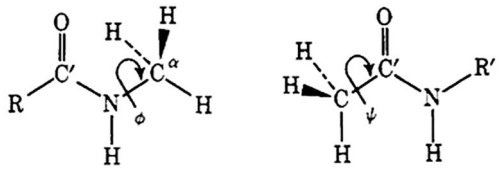
Molecular dynamics calculations of several representative protein and nucleic acid systems were carried out with the newly derived potential.

In a quirk of scientific research, for some reason Clementi's work (encompassing several 100 relevant publications) has been largely overlooked by practitioners in the biosimulation and even water communities. It is a puzzling phenomenon as this was pioneering work in the derivation of nonbonded functions from high level ab initio calculations, always at or above state of the art for the time, and many at a level beyond those currently being used for similar purposes. The body of work on the MCY and NCC-vib flexible and polarizable water potential, is likely the most extensive testing of a water potential against experimental properties by MD and MC carried out to date, and the latter may well be one of the "best" water potentials available [160]. We can only speculate that perhaps this is due to the fact that the overwhelming majority of his publications appeared in the International Journal of Quantum Chemistry, and J. Chem. Phys. rather than the computational chemistry and biochemistry journals where most of the papers in this field appear, and/or conferences where his work was presented were similarly focused on QM research as opposed to those devoted to the computational biology and drug discovery community. Whatever the reason it is unfortunate. Several reviews of this work have been published and the reader is referred to these for additional references and discussion [161, 162].

Though not focused on either potential function development or biomolecules, John Pople was perhaps the individual most responsible for the wide application of ab initio Hartree–Fock and post HF methods to these fields, through his Gaussian series of programs [136]. His work focused on establishing the concept of "model chemistry" [163], essentially analogous to the philosophy for QM method development as Lifson's Consistent Force Field philosophy for empirical force field development; namely rigorously determining the difference between the chemistry derived by a given model and experiment, for many molecules in many families, and then applying the model to studies of new systems [164]. In 1998 he was awarded the Nobel Prize in chemistry "for his development of computational methods in quantum chemistry". His Nobel address also provides an informative description of the evolution of the field of quantum mechanics [165]. In 1969, during the development and application of the semi empirical models described above, Hehre et al. [166] outlined the foundation for one of the most widely used ab initio MO programs (the Gaussian series), to be used by generations to come. They showed that a sum of Gaussian-type orbitals obtained by a least-squares representation of Slater-type atomic orbitals (STOs) used in SCF-MO calculations could account for ionization energies, atomic populations, and dipole moments which converged rapidly, with increasing size of the Gaussian expansion, to the values obtained for the pure STOs. The first program in the series, Gaussian 70 [167], was released to the QCPE in 1973 [136]. As noted above, Pople was not focused on biological applications, although along with Del Bene he carried out some of the first studies of water complexes, studying basis set dependence and demonstrating the non-additivity of hydrogen bonding [168–170].

7.3 Application of SCF-LCAO-MO to peptide conformation

In the mid to late 1970s Hagler et al. [171] carried out the first applications of ab initio Hartree Fock calculations to the peptide conformational problem aimed toward solidifying the basis for protein force fields. As had been noted by Scott and Scheraga and Brant and Flory there was only limited data on model compounds for rotational barriers about φ, and essentially no data for the energy accompanying rotation about ψ. In order to address the lack of information on the nature of the energetics about these critical torsion angles we applied ab initio calculations along with experimental crystal structure determination, and lattice energy calculations to a total of 19 amide crystals of the types shown in schematics **1** and **2**, to determine the barriers to rotation about these angles [171].



In addition, as discussed below, we applied the same strategy to the study of conformational polymorphism, isomorphism, and the relation between crystal forces and molecular-conformation [172–174]. These systems contain valuable information not only about nonbond force fields but also report on the intramolecular energetics. It would be well worth revisiting systems such as this to further test and derive biomolecular force fields.

Below, we diverge a bit and go into some detail in this discussion as the energetics of rotation about φ and ψ, are key components in the simulation of the conformational energetics and folding of proteins and have continued to vex the developers of protein force fields for many years. In addition, though the analysis of packing and packing effects on conformation have been studied extensively in the crystallographic community [175–178], the practitioners of biomolecular simulations and force field development, for the most part, have yet to fully exploit this powerful information source.

7.4 The φ–ψ torsion barriers determined from X-ray structural analysis, quantum mechanics and lattice energy calculations

X-ray crystal structure analyses of 12 molecular conformations indicated that the position of the minimum of φ (C′–N–C″–H), schematic **1**, in the solid state corresponded to a methyl C–H anti to the C′–N bond (φ = 180°). In ψ (H–C″–C′–N) the minimum was found to be 0°, i.e., methyl C–H syn to the C′–N bond, based on analysis of ten molecular structures of the type given in schematic **2**. Variations from these rotational minima appeared to be induced by crystal forces. In order to exploit these data to quantify the rotational energetics, ab initio molecular orbital and empirical force field calculations of the rotational potential surface, along with lattice energy calculations of the effect of crystal forces on the conformation were carried out.

The experimental data discussed here and in the original paper [171] can provide valuable information to help reduce the uncertainty, errors, and wide variability of proposed rotational energy surfaces for these dihedrals (see e.g. Hornak [179], and supplementary material therein, Feig [180] and Bochevarov [181]).

For those structures in which the hydrogen positions weren't given in the literature, (which was not uncommon in the '70s), we located them by first refining the positional and anisotropic thermal parameters of the C, O, and N (and other heavy) atoms with reported structure factors. This was followed by computation of electron density difference maps on which the hydrogen atoms were evident (Fig. 5) and further refinement with isotropic temperature factors for these hydrogen atoms [171]. (We were not able to locate the hydrogen atoms in the orthorhombic form of acetamide by these same techniques indicating it is freely rotating.)

Hydrogen atoms were usually not located by room temperature X-ray diffraction studies to the precision required to determined precise molecular geometry in the 1970s (e.g., scatter in bond lengths was often of the order of 0.1 Å and in bond angles as much as 10°, an order of magnitude larger than for heavier atoms). However, for the purposes of this work, these

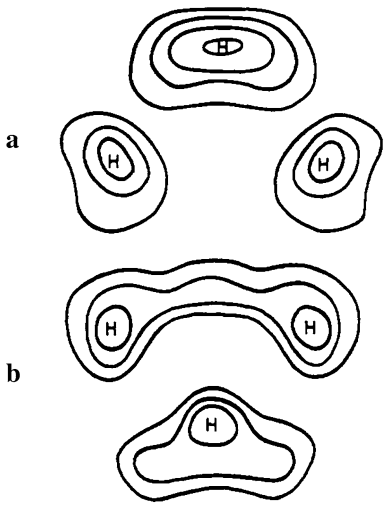


Fig. 5 Electron-density difference maps through the plane of the three hydrogens in *N*-methylacetamide. **a** *N*-methyl hydrogens; **b** C′-methyl hydrogens

deviations were not so serious when the hydrogen positions are used to determine torsion angles for two reasons. The first is that the methyl torsion was determined from three hydrogens and thus statistical errors were reduced by the averaging process. Secondly, the objective in determining torsion angles here was not the precise determination of structure but rather of the overall conformation (e.g., the extent of the deviation of the methyl torsion angle from the staggered form). Thus, determination of this angle to within 3°–5°, which is attainable by these techniques, was highly satisfactory for these purposes.

The values of the torsion angle about φ and ω for a series of 12 *N*-methyl amides, contained in ten crystal structures (two of which contain two molecules per asymmetric unit) are given in Table 2. The torsion angles, φ, as given in Table 2 are such that for conformation **1**, in which one of the C–H bonds is anti to the C′–N bond, φ = 180°. Of these 12 molecules, six were found to display φ angles of 180° ± 15°, while four of the torsion angles deviated by as much as 30° from this conformation (φ ≈ 150°) and only the two molecules of *N*-methylformamide (in the crystal complex with oxalic acid) presented torsion angles of ~ 120° where the C′–N bond is eclipsed by a methyl C–H bond.

Variation of ψ in crystals. The values of the torsion angles about ψ for ten molecules contained in nine crystal structures are given in Table 3. The torsion angles for ψ are defined such that for conformation **2**, in which one of the C–H bonds is eclipsed by the C′–N bond, ψ = 0°. This set of molecules also bears out that significant differences in the conformation of the methyl group can arise from differences in crystal forces. The same molecule, acetamide, is contained in four different crystal structures and is found to have significantly different values of ψ in these different crystal environments. The inherently favored position of the methyl in these condensed state hydrogen bonded systems appears to be ψ = 0° with the largest value of ψ being − 30° found in rhombohedral acetamide.

Since the *N*-methyl (N–CH_3) is fairly far removed from the R group (see **2**), we expected that the effect of this group on the intramolecular torsion potential should be small. Thus, it appears that the differences in conformation observed in Table 2 are primarily due to the effect of the different intermolecular environments on the *N*-methyl torsion angle in the different crystals. The crystal structure of *N*-methyl-α-chloroacetamide, which contains two molecules per asymmetric unit having significantly different φ angles (φ = 173° and 156°, while the two values of ψ are 179° and 166°) as well as the various values of ψ observed in the different acetamide complexes given in Table 3 provide support for this argument.

7.4.1 Qualitative conclusions on the nature and magnitude of the φ–ψ barriers in peptides

The results presented above would seem to indicate that both φ and ψ have inherently preferred conformations in hydrogen bonded complexes. For φ, the minimum appears to be that conformation in which one of the methyl C–H bonds is anti to the C′–N bond, i.e. φ = 180° as defined here. This is borne out by both the crystal data and the electron diffraction results on *N*-methylacetamide (NMA) [182]. The barrier to rotation would appear to be small (≤ 1 kcal) since it is observed that crystal forces can cause significant variation in the methyl torsion. The small to negligible values of the barriers for both φ and ψ, are further supported by the series of gas phase electron diffraction studies carried out on NMA [182], *N*-methylformamide (NMF) [183] and Acetamide

Table 2 Observed torsion angles φ and ω in *N*-methylamides (angles in deg)[a]

Molecule	φ (H–C^{α}–N–C′)	ω (C–C′–N–C)	R
N-methyltetrolamide	164 (4)[b]	0	–C≡CCH_3
N-methylacetamide	180 (0)	0	–CH_3
N-methylpropiolamide	169 (2)	6	–C≡H
N-methylcinnamide	156 (3)	5	–CH=CHC_6H_5
N-methylsorbamide	147 (3)	2	–C=CHCH=$CHCH_3$
N-methyl-α-chloroacetamide	173 (2)	1	–CH_2Cl
	156 (6)	1	
N-methylformamide	125 (7)	42	–H
	118 (6)		
N-methylbenzamide	167 (2)	1	–C_6H_5
N-methyldipropylacetamide	179 (7)	0	–C(H)(CH_2 CH_2 CH_3)$_2$
N-acetyl-DL-phenylalanine-*N*-methylamide	144 (12)	5	

[a]For references to these structures see Hagler et al. [171]

[b]The standard deviation in torsion angle as determined from the scatter of the three individual H–C–N–C′ torsion angles

Table 3 Observed values of ψ in ten amides and peptides (angles in deg)

Molecule[a]	ψ (H–C–C′–N)	R′
Acetamide (rhombohedral)	29 (9)[b]	–H
Acetamide (oxalic acid complex)	15 (4)	–H
	14 (6)	
Acetamide (allenedicarboxylic acid complex)	0 (3)	–H
Acetamide (–5,5′diethylbarbituric acid complex)	5 (4)	–H
N-methylacetamide	0 (0)	$–CH_3$
N-acetylglycine	Free rotation	$–CH_2COOH$
Acetanilide	13 (0)	$–C_6H_5$
N-acetyl-L-tryptophan methyl ester	1 (7)	
N-acetyl-DL-phenylalanine-*N*-methylamide	3 (4)	

[a]See Hagler et al. [171] for references to these structures

[b]The standard deviation in torsion angle as determined from the scatter of the three individual H–C–C′–N torsion angles

[184], by Kitano and Kuchitsu in 1973 and '74. As noted above, the gas electron diffraction results for NMA indicated that the methyl C–H bonds was anti to the C′–N bond (φ = 180°), while the methyl conformations in *N*-methyl formamide and acetamide were indeterminate, possibly indicating free rotation and again an extremely small barrier height, less than kT (≤ 0.7–0.8 kcal at experimental temperatures).

N-methylformamide in its crystal complex with oxalic acid is an exception where the methyl C–H bond is found to be syn to the C′–N bond rather than anti. The gas-phase conformation of the methyl in this compound was indeterminate and the conformation in the crystal was hypothesized to be due to the strong amide-acid packing forces. This exception is an example of a system which should provide a valuable test of proposed inter- and intra-molecular potential functions.

Similar considerations hold for the barrier to rotation about ψ. Here the preferred conformation appears to consist of a methyl C–H syn to the C′–N (i.e. ψ = 0) in the solid state. In this case, however, as indicated by the apparent free rotation of the methyl group in *N*-acetylglycine in the crystal and in one of the methyl groups in diacetamide in the gas phase, the barrier to rotation appears to be even smaller (e.g., of the order of a few tenths of a kilocalorie).

7.5 Quantum mechanics: bond-torsion and angle-torsion coupling crucial to calculating φ–ψ torsional energies

In order to assess these observations quantitatively, both with respect to the approximate size of the rotation barrier, the effect of substituent (R), geometry, and the magnitude of the effect of the crystal forces, we carried out ab initio and empirical force-field calculations of the rotation barrier and lattice energy calculations of crystal energy as a function of methyl rotation. Ab initio calculations were carried out for both minimal and split valence basis sets, which were tractable at the time. Selected results for the split valence basis set are given in Table 4. As seen in this Table, the minimum in the rotation about φ occurs at 180° in agreement with the experimental observations, and limited gas phase data. In the case of ψ however the calculations predict a minimum at ψ = 60° rather than at 0° observed in the crystal. Both barriers are extremely small with the barrier for φ of the order of 0.2–0.7 kcal and for ψ being negligible, of the order of 0.1–0.2 kcal, corresponding to essentially free rotation at room temperature, again in agreement with the experimental observations.

Table 4 Barriers to rotation about φ and ψ: effects of substituents, and geometry

Molecule	φ°	E (φ) – E (180°) (kcal/mol)
N-methylformamide (C′–N, 1.38 Å)[a]	120	0.33
N-methylformamide (C′–N, 1.29 Å)	120	0.13
N-methylacetamide (ψ = 0)	120	0.73
N-methylacetamide[a] (NMF geometry)	120	0.25
N-methylpropiolamide	120	0.25
Molecule	ψ°	E (ψ) – E (0°) (kcal/mol)
N-methylacetamide (φ = 180°)	60	− 0.06
Acetamide	60	− 0.15

As expected from the structures of the molecules, the nature of the R group did not make a difference in the location of the barrier though it did affect the barrier height somewhat (as much as 0.5 kcal in NMA vs. *N*-methylpropiolamide, Table 4). However, analysis of the calculations indicated that in fact the barrier is very sensitive to the amide geometry (e.g., the OC′N and C′NC angles) rather than to any direct interaction of the *N*-methyl group with the R group. Thus, the OC′N and C′NC angles are slightly smaller in NMA than they are in either NMF or NMP and it is this difference that leads to the differences in the barrier heights in these molecules. The effect of the geometry (as opposed to inductive or direct interactive effects with substituents) was confirmed by calculating the rotational potential for φ in NMA with the geometry of NMF by QM. These results corroborated that the direct effect of the R group on the barrier is indeed very small (−0.1 kcal) and that it is the variation in geometry either induced by the different R groups or by differences in crystal forces that accounts for the rest of the variations in the barrier heights. It was also found that there is a coupling between the amide bond length (C′–N) and the barrier. Thus, when the barrier in isolated NMF is calculated with the bond length observed in the solid state rather than the gas phase (C′–N = 1.38 vs. 1.29 Å) the torsion barrier was reduced by 0.2 kcal. (These observations once again emphasize the importance of coupling between internals, and intramolecular potentials as these directly affect conformational energetics. In this case the results also document intramolecular coupling with hydrogen bonding as well, as the latter is responsible for the change in bond length.)

7.5.1 Lattice energy calculations: intermolecular forces induce strain about φ–ψ, and modulate the intrinsic rotational profiles through geometric distortions

The experimental observations [171] reproduced in Tables 2 and 3 demonstrated that although the methyl groups have an inherently preferred orientation in φ and ψ, the orientations are compromised somewhat by the requirements to optimize the crystal packing energy. In fact, the conformation about φ and ψ in a given compound was shown to vary in different crystal environments. In most cases the deviation from the minimum was less than 30°, but in a few cases this was achieved and even exceeded (Tables 2, 3). If the torsion energy is of the form $V(\phi) = 1/2K_{\phi}(1 \pm \cos 3\phi)$, a deviation of 30° from the minimum corresponds to an increase of $1/2K_{\phi}$ (or half the barrier height). This puts an upper bound on the value of the barrier, as it is limited to the energy available from lattice forces to induce the observed deformations.

To quantify the lattice energy available to induce deformations about these torsion angles we carried out an analysis of the crystal packing energy as a function of rotation about these methyl groups. If this is very large, i.e. much greater than the barrier to rotation, the crystal forces alone would determine the methyl orientation. On the other hand, if the lattice energy was insensitive to methyl rotation, the methyl orientation would be determined solely by the intramolecular "inherently preferred" position. From the experimental observations, we expected that the change in crystal energy with methyl rotation would be of approximately the same magnitude as the intramolecular energy (e.g., the barrier height). Thus the calculation of minimum crystal energy changes as a function of methyl rotation can provide us with corroboratory evidence as to the magnitude of the rotational barrier as estimated by the quantum calculations, and the limited gas phase experimental determinations [182].

Calculations were carried out on *N*-methylacetamide (R=CH_3 in schematic **1**), which had the advantage of being relatively small, yet still contains both C′- and *N*-methyl groups, thus enabling us to vary both φ and ψ. Each of these angles was varied by 30° increments, and for each set of φ, ψ angles the lattice energy (E_{Lat}) was minimized with respect to all crystalline degrees of freedom (i.e., unit cell parameters and molecular orientation). This was carried out with three force fields (6–9 [86], 6–12 [86], and MCMS [63]), in order to avoid, to the degree possible, potential dependent conclusions. The results are given in Table 5.

As was expected, the variation in packing energies as a function of methyl orientations was indeed small. Despite significant differences in the total lattice energy (especially in the case of the MCMS potential) this behavior was independent of the FF used. The largest energy difference were of the order of ~0.6–0.7 kcal as calculated with the 6–9 [86] and MCMS [63] potentials corresponding to the difference between the structures (φ, ψ) = (120, 60) and (150, 0).

The magnitudes of the variations in energy for these conformational polymorphs were compatible with data available at the time. These included Kitaigorodskii's estimate of energy difference between polymorphs as well as

Table 5 Lattice energy of *N*-methylacetamide as a function of methyl rotations (φ and ψ)

φ (deg)	ψ (deg)	E_{Lat} 6–9 potential (kcal/mol)	E_{Lat} 6–12 potential (kcal/mol)	E_{Lat} MCMS potential (kcal/mol)
120	0	−15.38 (0.62)	−16.58 (0.36)	−11.04 (0.69)
120	30	−15.71 (0.29)	−16.66 (0.28)	−11.33 (0.40)
120	60	*−16.00 (0.00)*	*−16.94 (0.00)*	*−11.73 (0.00)*
150	0	**−15.37 (0.63)**	−16.49 (0.45)	**−10.97 (0.76)**
150	30	−15.50 (0.50)	**−16.47 (0.47)**	−11.06 (0.67)
150	60	−15.88 (0.12)	−16.77 (0.17)	−11.50 (0.23)
150	90	−15.76 (0.24)	−16.73 (0.21)	−11.34 (0.39)
180	0	−15.67 (0.33)	−16.79 (0.15)	−11.40 (0.33)
180	30	−15.74 (0.26)	−16.70 (0.24)	−11.32 (0.41)
180	60	−15.84 (0.16)	−16.64 (0.30)	−11.31 (0.42)

other contemporaneous determinations, which invariably gave energies on the order of 1 kcal/mol between crystal forms [185, 186]. The energy difference between polymorphs found by these early workers has been confirmed by many studies over the intervening years [175, 177, 178]. Thus in a review of crystal packing Dunitz and Gavezzotti recently stated in a summary of the research on polymorphs [177]. "Energy differences between crystal polymorphs are extremely small, a matter of a few kJ/mol. Typically, for any given molecule, a computational search produces many different periodic arrangements within a narrow window of lattice energies, with unknown energy barriers between them."

In addition, subsequent more rigorous quantum results with significantly larger basis sets have confirmed the conclusions based on results in Table 4. Thus for example Guo and Karplus carried out extensive calculations of NMA in various H-bonded complexes with water and formamide [187], Menucci and Martinez [188], and Villani et al. [189], also studied NMA and its H-bonded complexes in detail. These studies replicated the experimentally observed difference between gas phase [182] and solid state [190] geometry and barrier heights, and all concluded that the barriers to rotation in NMA are in the range 0.2–0.9 kcal/mol in agreement with our earlier results.

7.6 The bottom line: *Kφ* and *Kψ* are of the order of several tenths of a kcal/mol

Based on the above, there is little doubt that the lattice energy differences between the rotational conformers of NMA are reasonably recapitulated by the results given in Table 5. In addition, the difference in energies obtained for these conformational polymorphs were in excellent agreement with the ab initio calculations. As noted above, since the packing forces affect the rotational state about φ and ψ (Tables 2, 3) which display deviations from the minimum of as much as 30°, the barrier can't be much different then given by the ab initio energies of approximately 0.2–0.7. That is, for a 30° distortion the strain energy induced is half the barrier height (~ 1/2 $K\phi$), and the barrier must be, at most, the same magnitude as the lattice energy available to induce the strain. Thus. the barrier about ϕ can be at most a kcal or so, and most likely of the order of 0.5–0.7 kcal, while it is at most several tenths of a kcal for ψ, consistent with the several structures which exhibit free rotation about this angle.

7.7 Lifson's CFF paradigm, Allinger's approach, and the saga of the torsional potentials about φ and ψ

As seen from the determination magnitudes of the barriers about φ and ψ, there exists information rich experimental data which provide fairly tight constraints to validate any proposed torsion potential about these angles. Despite this, as noted above, these critical torsions have been problematic for practitioners in the field. This has led to numerous reparametrizations and to artifacts in the preponderance of secondary structures in protein simulations [179–181, 191]. We suspect that many of the problems arose due to the use of a limited number of experimental or QM data resulting in the determination of these parameters. In addition, many of these studies optimized ϕ, and ψ to fit high level quantum calculations of oligopeptide conformational energies and dipeptide maps, as well as conformational populations in proteins—constraining all other nonbond and electrostatic parameters to previously determined values. Thus, any errors in the multiple intra and intermolecular parameters would have to be absorbed in the torsion parameters for the systems being optimized. These errors in torsion potential in turn would cause discrepancies to arise when different or additional systems are treated, which has been the historical experience. Finally, these parameters are being forced to account for limitations in the functional form, further exacerbating the situation. In hindsight it would appear that much of the necessity to reparametrized these torsion potentials might have been alleviated, by adopting the Lifson approach [49]. Namely, if a wide variety of experimental and quantum observables were included, for example the extensive QM calculations on NMA [187], dipeptide φ, ψ maps, oligopeptide energetics, and crystal data, including that cited above and peptide crystals structures available in the literature, it is extremely unlikely, if not impossible that the results would differ by more than a few tenths of a kcal from those reported above. This may well have led to a focus on other deficiencies and perhaps to earlier focus on improvements in representing the spatial electron densities, polarizability, and coupling required to adequately represent the peptide energy surface, much as the Allinger approach discussed above.

Another problem, known from the gas phase and crystal determinations in the '60s and determinations in our work [171] and Guo's [187] is the change induced in the geometry of the amide group by hydrogen bonding in the condensed phase, and the effect of these geometrical changes on the torsional potential. Thus, the gas phase dipeptide and oligopeptide quantum results could be, and were, misleading. Basically, either a "condensed phase" potential was required or a coupling between hydrogen bonding and internals (at least C′–N bond length), and most likely a cross term between torsion and bond length would be required to account for this behavior. It is likely that a fit of all parameters in the FF by least squares to the entire large and diverse set of QM and experimental observables referred to above would have led to a "consistent force field" [49], elucidating the need for these coupling terms, analogous to the way the need for an angle–angle–torsion was found by Warshel and Lifson [52].

8 The hydrogen bond: electronic basis for omitting the van der Waals interaction of polar hydrogens

The early appreciation of the applicability of QM to biological systems was described above. In the early years of force field development, we, along with others, recognized the additional opportunity provided by QM algorithms to complement experimental data and provide information about the description of both intramolecular strain, and electronic structure and interactions beyond simple derivation of partial charges [192].

In 1976 we exploited the availability of these algorithms to carry out an ab initio study of the spatial electron density in amides, acids and peptides [193, 194]. The primary motivation for the study was to examine the validity of some of the more common assumptions made in conformational analysis. In particular, the spherical atom approximation implicit in the functional form of the atom–atom van der Waals interaction, the representation of the continuous electron distribution in the molecule by monopoles placed on the atomic center and the transferability of potential constants. In addition, and perhaps of most importance, was our observation on the nature of the hydrogen bond and the finding that the amide and acid hydrogens were essentially incorporated into the van der Waals sphere of the nitrogen and oxygen atoms. That is, O–H and N–H donors "as seen by acceptors" could be represented by a single Lennard-Jones center located at the heavy atom [85, 86, 88].

To this end, contour maps and three dimensional contours of constant electron density were calculated for both total and difference electron density maps of several amide and acid compounds [192–194]. An example of the spatial electron density surrounding the amide group in *N*-methylacetamide (NMA) at two contour levels is presented in Fig. 6. while 2-D contour maps of the molecular electron density, as compared to the density of superposed spherical atoms are given in Fig. 7.

The results shown in Fig. 6, (along with others in the study [193]) indicated that the total molecular electron density was virtually indistinguishable from the electron density resulting from superposition of spherical atoms placed at their coordinates in the molecule. Thus, aside from a general withdrawal of electrons from the peripheral regions of the molecule, the general features associated with molecular formation, such as the bonding density and carbonyl oxygen lone-pair orbitals, are not evident in Fig. 6a. In fact, the overall shape of the molecule is very similar to that of the density produced by superposed spherical atoms as given in Fig. 6b. The similarity between the actual (calculated) molecular density and that obtained from superposed spherical atoms is due to the fact that molecular formation involves a relatively small redistribution of electrons (see below). The size and shape of a molecule are related to the volume excluded to other molecules or parts of the same molecule. This factor, is represented by the spherically symmetric Lennard-Jones or Buckingham terms (as in Eq. 4). These features provide an understanding of why the use of spherical atoms as a first approximation, to represent the excluded-volume effect in Eq. (4) seem to give reasonable results. It should be noted however, that more recent studies indicate that rigorous results require inclusion of anisotropy in this interaction [175, 195–199].

8.1 Electron migration from amide and acid hydrogens

An interesting situation emerged for the N–H group. While in the higher density shape plot (Fig. 7b) the distinct shapes of the nitrogen and hydrogen are visible, the shape of the molecule emerging from the van der Waals surface is quite diffuse, and only a single, slightly distorted, spherical shape can be discerned, enveloping both the nitrogen and

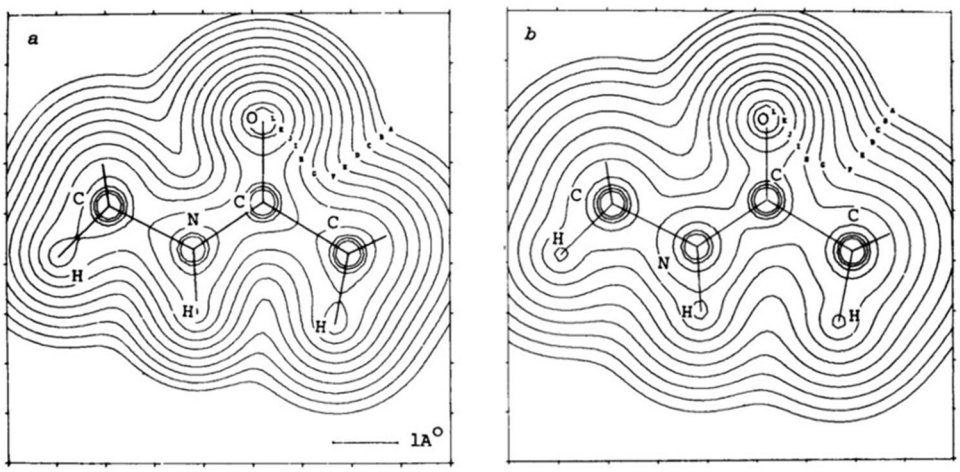


Fig. 6 Contour map of the electron density in the amide plane of *N*-methylacetamide as obtained with the split valence 6-31G basis set. **a** Total density as calculated from molecular wave function. **b** Density obtained from superposition of spherical atoms

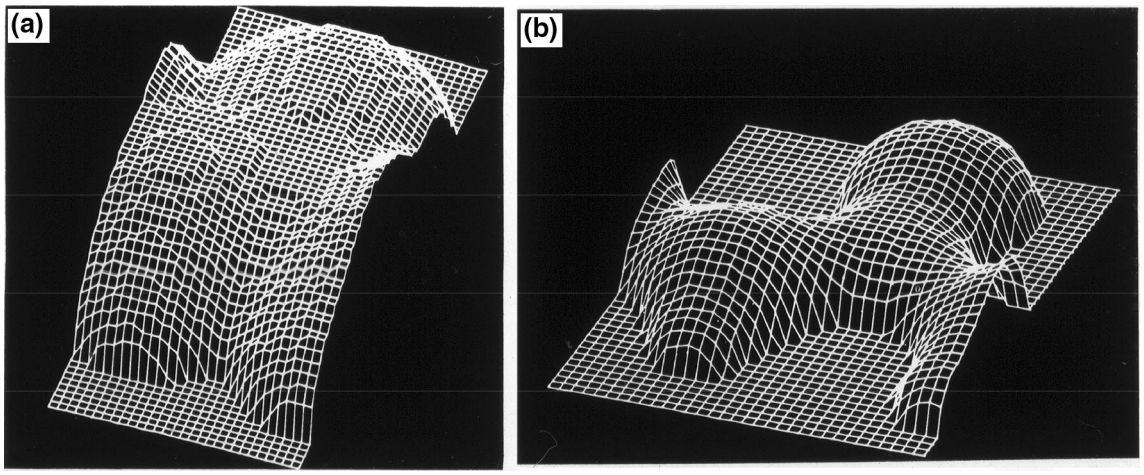


Fig. 7 The overall three-dimensional shape of the amide group in *N*-methylacetamide. The carbonyl group is on the right with the oxygen at the top of the figure, while the N–H is trans, directed toward the lower border of the plane. **a** Surface of constant electron density of 0.027 electrons/Å^3, roughly corresponding to van der Waals radius. **b** Surface of constant electron density of 0.75 electron/Å^3. The single spherical contour about the N–H atoms at the van der Waals distance is consistent with the omission of the vdW interaction of the amide hydrogen in the Hagler–Lifson formulation of the H-bond potential [86]

hydrogen. The N–H group appears to be one large atom, whose excluded volume may be represented by a single central repulsion. This is consistent with, and rationalizes the results of Hagler and Lifson [86] in the crystal work where the least-squares analysis revealed that the amide hydrogen parameters were indeterminate, having standard deviations several times larger than the parameters themselves. It was found that the van der Waals radii of the hydrogen could be ignored and incorporated into the nitrogen when representing the hydrogen bond, with no degradation in the ability to fit the structure and energetic properties of the amide and acid crystals.

8.1.1 Electron density and electrostatic interaction

The situation is very different for the electrostatic properties. Rather than being dependent on the total electronic distribution, these interactions arise from the small reorganization of electrons that accompany molecular formation. Thus, if the molecular electron density were truly represented by superposition of spherical atoms it would have a zero dipole moment and be apolar. In order to investigate the electron migration upon bond formation and the resulting nature of the charge distribution, difference electron density maps of several amides and acids were examined [193, 200]. These maps are calculated by subtracting the sum of the component atomic densities from the total molecular density. Figure 8 depicts the difference density map for NMA (*N*-methylacetamide). Here one can clearly see the common features usually associated with molecular formation such as migration of electrons into the bond region between the atoms in the difference density. In addition, the lone pair orbitals on the carbonyl oxygen are a dominant feature. On examination, it becomes clear that the total migration of electrons on molecular formation is small, especially when compared to the total number of electrons.

This may be seen quantitatively in Table 6, which summarizes the electronic population of the bonding and lone pair orbital densities in NMA, as obtained by numerical integration of the corresponding densities in the difference map. From this it can be seen that the integrated positive difference density in the lone pair orbitals is only ~ 0.25 electrons. Even this difference density as seen at low density, coalesces into a single contour surrounding the oxygen, which, although not spherical, deviates from the usual "rabbit ears" schematic representation of the orbitals, and constitutes a smaller perturbation on the spherical oxygen.

These plots emphasize the severe approximation implicit in atom centered partial atomic charges. Clearly if the electrostatic forces are a major contribution to the properties being calculated (as is almost always the case in biological systems), this is a major source of potential error. Interestingly models of water have attempted to incorporate the features of the deformation density from the beginning. Thus the 1933 model of Bernal and Fowler [8] (see below) invoked an off atom center of negative charge to calculate the properties of ice and water. In 1951 Rowlinson [201] incorporated not only multipole moments but also induction (polarization) into his treatment of the lattice energy of ice. Essentially all water potentials since, and they number in the hundreds, have incorporated off atom descriptions of the electrostatic interactions of water (with the notable exceptions of the SPC [202] and simple "TIP" models [203]). In addition crystallographers and others focused on prediction

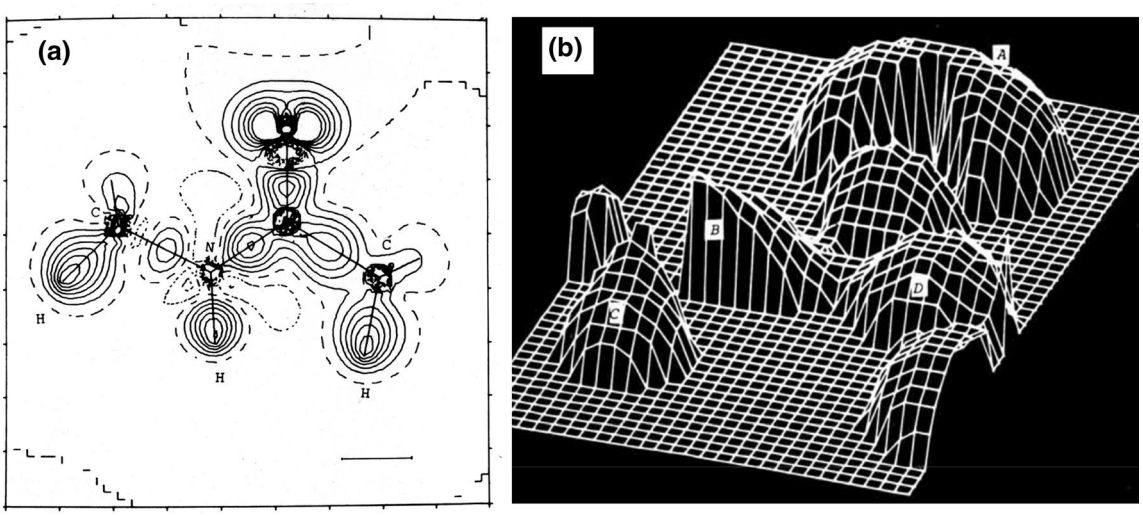


Fig. 8 *N*-methylacetamide difference map of the total molecular density as calculated with the extended 6-31G basis set minus the density of the spherical atoms. **a** Contour map in the plane of the amide group. Solid, dashed, and dotted lines represent positive, zero, and negative difference. **b** The overall three-dimensional shapes of the positive difference density. The lone pair orbitals are labeled A, while the C′–N bonding density is B, the N–H density, C and the C′–C density, D

Table 6 Electronic population and perpendicularity (to molecular plane) of the bonding and lone pair orbital in *N*-methylacetamide

Bond/lone pair	Population (electrons)	Height (Â)
O= (orbital)	0.25	0.97
C=O	0.12	0.6
C–N	0.20	1.1
N–C_α	0.05	0.93
C–C′	0.17	0.77
N–H	0.13	1.25
C–H (NH methyl)	0.22	0.77
C–H (C=O methyl)	0.28	0.99

of crystal packing modes also noted the importance of higher order models of electrostatic interaction rather early on [93, 204], including multipole interactions, and Gaussian representations among others [177, 205].

Despite the importance of these more realistic treatments of electrostatic interactions the overwhelming majority of simulations of protein–ligand systems and other biomolecules, including applications in drug discovery continue to use the crude partial atomic charge model. In large part this is due to the significant computational cost of more realistic treatments. There are indications that this situation may be changing however, and though still only sparingly applied, recent force field development, such as the SDFF [206–208], and the Amoeba potential functions [98, 209] incorporate atomic multipoles such as those used in the water and crystal engineering force fields.

9 Genesis of protein simulation software: divergent evolution of Lifson's CFF—CHARMM, AMBER and GROMOS

In the early '70s, the CFF program 'emigrated' to the US and became the basis for the first protein simulation packages. Outside the Lifson group itself, the first to adopt the CFF was the Karplus group at Harvard. Karplus spent a semester visiting Shneior Lifson at the Weizmann Institute, where he was first introduced to the potential of empirical energy functions. The CFF program was then brought by Warshel to his lab and this code was subsequently developed into the CHARMM package (Chemistry at HARvard Molecular Mechanics) [109]. Paul Weiner, who had been studying as a postdoc with Karplus at Harvard, brought the program and associated FF to the lab of Peter Kollman at UCSF. This version of the CFF evolved into Assisted Model Building with Energy Refinement (AMBER) [210]. Finally another branch of the descendants of the CFF emerged when van Gunsteren, who studied with Karplus, brought the program with him to his and Berendsen's lab in Groningen, the Netherlands, and it evolved into GROningen Molecular Simulation (GROMOS).

These groups were all excited by the idea of applying molecular mechanics and dynamics calculations especially to proteins and other biological molecules. As such their primary early focus was on adapting the program to the treatment of these large biological systems, rather than on the improvement or rigor of the underlying potential functions, and there was an intensive push to develop the techniques and software to enable such application. This led to

McCammon, Gelin, and Karplus' classic paper on the first MD simulation of bovine pancreatic trypsin inhibitor (PTI) "Dynamics of Folded Proteins" in 1977 [211]. (This also gives some perspective on the advances in computational power in the intervening years—the trajectory for this first application consisted of 9 ps of dynamics of PTI (58 residues) without hydrogens or solvent, as contrasted to the microsecond simulations of large proteins in aqueous solution made possible by the hardware technology of today.)

9.1 CHARMM

The CHARMM FF used in these early protein calculations included only harmonic terms for bonds and angles, sinusoidal function for torsions, 6–12 NB, and an electrostatic term with an optional dielectric (Eq. 10), and a complex angle dependent H-bond term [108]. The force constants were obtained to a large extent by using literature data by Schleyer [212], Lifson and Warshel [26, 54] (for hydrocarbons), and by Lifson et al. [53] for amides. Some parameters were taken from Gelin's thesis. A complete set of parameters (excluding charges) as used for simulations of the protein bovine pancreatic trypsin inhibitor is given by McCammon et al. [108]. In addition, the very impactful approximation (in retrospect) was made that hydrogen atoms could be ignored—the so-called united atom approximation. This approximation, reduces the number of atoms by roughly 30%, and the resulting calculation time per iteration by about 50%, and also reduces the total number of iterations that are required to achieve a given convergence criterion, so it was quite tempting to use it for investigation of large biological systems. However, although widely adopted at the time, there was no evidence that this approximation was justified. It also introduced complications with respect to retaining the chirality of carbon atoms (when the H_α is omitted). A special 'out of plane' term had to be introduced, preventing the C_α from becoming planar and then racemizing.

9.2 AMBER

The initial functional form and parameters of the AMBER force field [210], and the use of the united atom model, were taken directly from CHARMM [109]. A few years later the parameters were modified, but the functional form was retained. The Kollman group first modified the parameters to fit observables of small model molecules [106]. The parameters for equilibrium bond lengths and angles were taken from microwave and X-ray data. The force constants were obtained by considering the fit to experimental frequencies of 6 small model compounds. Similarly, a FF for nucleic acids was also developed. An extension of this work to an all-atom force field (i.e. inclusion of all hydrogen atoms) was presented in 1986 [213]. They did not attempt to obtain a good fit to vibrational frequencies, recognizing that this is unachievable without coupling terms. Instead they stated the goal was a 'wish to ensure that the general features of the vibrational spectra are reproduced'. Therefore, deviations of 100–150 cm^{-1} or more were not a cause for concern.

9.3 GROMOS

The GROMOS FF was also kept to a very simple form, with harmonic terms and parameters for the internals also mostly transferred from CHARMM [214]. Later non bond parameters [215] were chosen similar to those given by Hagler, Lifson and Dauber [100, 101]. These were chosen because of their compatibility with the SPC water model they developed (see below), which adopted the Hagler–Lifson formulation of the hydrogen bond [86], i.e. negligible polar hydrogen van der Waals, larger donor atom, and similar partial charges.

9.4 OPLS

Following the derivation of the simple 3 and 4 point TIPS potentials for water in the early 80s [203, 216, 217], Jorgensen's group pioneered the approach of using Monte Carlo simulations of liquids to derive parameters for small organic molecules, culminating in a FF for peptides and proteins [218, 219] (OPLS—optimized potentials for liquid simulations). In the first application Jorgensen derived potentials for amides by carrying out Monte Carlo simulations of liquid formamide, *N*-methylacetamide and dimethylformamide. Densities and heats of vaporization were fit to experimental values. The structure of amide dimers, in configurations observed in the crystal, as well as amide water complexes were also calculated to ensure that the potentials provide reasonable descriptions of the energy and structure of these complexes [218]. This was followed by a very extensive study in 1988, in which protein nonbond parameters were obtained and tested by Monte Carlo simulations of 36 pure organic liquids comprised of alcohols, sulfides ketones, etc. representing the side chain functional groups of amino acids [219]. Aqueous solutions of organic ions representative of the side chains of the charged amino acids were also simulated. Calculated heats of vaporization, and volumes were compared with experimental results for the pure liquid, and heats of hydration were compared for the solutions. The resulting FF was also tested by minimization of crystals of five small cyclic hexa- and pentapeptides containing Ala and Gly and on the small protein Crambin. In the latter calculations crystalline waters were also included [219].

The united atom approximation was utilized in these studies, omitting all non-polar hydrogens.

Bond stretch, angle bend, and torsional terms were adopted from the AMBER united-atom force field. A simple 6–12 Lennard-Jones potential and Coulombic electrostatic

were used, and the Hagler–Lifson [86, 100] hydrogen bond potential involving only van der Waals and Coulombic interactions, with a negligible repulsion on the donor hydrogen, was also adopted in these studies. The resulting fit of calculated and experimental properties used in the parameterization was also adopted as the primary measure for the FF's performance. It was shown that on the whole the modified Amber force field with the OPLS nonbond parameters gave a better fit to the peptide crystal systems and Crambin than the original Amber FF [219].

This very extensive work demonstrated the power of using liquid observables for parameter derivation. A downside of these systems is that unlike the situation with water there are often only two observables per liquid—i.e. the density and heat of vaporization. Thus, there is a practical limit on the ratio of observables to parameters. Crystals on the other hand contain nine or more observables per system, including unit cell vectors, orientation of the asymmetric unit, intermolecular geometries such as hydrogen bond distance and angle etc (see Fig. 4). Thus, by exploiting such systems for FF optimization, observable to parameter ratios of ~ 10:1 or greater can be achieved. The optimal solution is to adopt Lifson's principle and fit or test parameters against all available experimental data. Jorgensen et al. [219] in fact went a way toward adopting this principle by testing the force field against the peptide crystal systems.

9.5 Calculation of a drug receptor complex: first valence force field for proteins (CVFF)

While the Lifson and Warshel CFF [53] formed the basis for applications to peptides and proteins, it was not complete, as it did not address the need for parameters for the functional groups in the side chains of amino acid residues. In addition, for amides only the Urey–Bradley FF was employed. In the early 1980s we were interested in carrying out one of the first MM and MD studies of the energetics of a drug receptor system, the antineoplastic agent methotrexate (MTX) and the antibiotic trimethoprim (TMP), bound to their target receptor, dihydrofolate reductase (DHFR) [220, 221], (a project suggested by Joe Kraut at UCSD). In order to accomplish this it was necessary to extend the valence force field introduced by Ermer and Lifson [55] to account for the backbone and for all the functional groups found in protein side chains. This was carried out in a preliminary study, by fitting geometrical and vibrational data, rotational barriers and dipole moments of a limited set of 25 model organic compounds incorporating the moieties of the protein side-chains (amides, acids, amines, alcohols, ethers, thiols, sulfides, disulfides, aromatic and hetero-aromatic compounds). A full valence force field including cross terms was derived and published [221] (though the fit to the model compounds were not published at the time).

The parameters for aliphatic C′s and H′s were transferred from Lifson's work. All other parameters were adjusted to fit experimental data. Typical deviations from experiment for these compounds were on average about 0.01–0.02 Å for bond lengths (aside from the amide bond where the coupling between bond length and hydrogen bonding is not accounted for), 3° for angles and 30–40 cm^{-1} for the vibrational frequencies. A summary of a representative fit to a set of amides, including pyrrolidine is given in Table 7.

Though based on a minimal set of model compounds, the FF established a starting point for simulating peptide and protein systems, it was superior to contemporaneous FFs, most importantly because it was an "all atom" FF including all hydrogen atoms. In addition, it accounted for coupling between angles and torsions, important for accounting for rotational barriers and conformational equilibria.

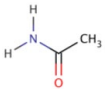
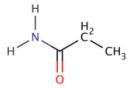
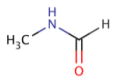
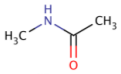
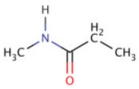
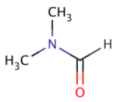
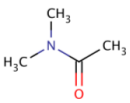
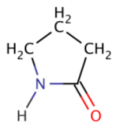
As noted above, we exploited this force field to carry out the first calculations of a drug receptor system, the complexes of TMP and MTX, with their receptor DHFR. All hydrogen atoms were included explicitly, as were all crystallographically determined water molecules within 7 Å of the ligand, as well as those that formed hydrogen bonds to at least two protein residues. The energetics, and free energy of ligand binding were analyzed for the simulated structures. Strain energy induced in the ligand, the corresponding entropy loss due to shifts in vibrational frequencies, and the role of specific residues in ligand binding were examined. Water molecules, even those not in direct contact with the ligand, were found to have significant interaction energies with the ligand. This constituted the first simulation in which the same force field used for small molecules, incorporating cross-terms and the explicit inclusion of all hydrogen atoms, had been fused for a protein system.

The same FF was also used in some of the first drug design investigations where the receptor structure wasn't known. In these studies, molecular dynamics simulations were carried out to investigate the conformational space available to a peptide hormone and its analogs, e.g. lysine vasopressin, as part of a strategy to elucidate the structural requirements for binding and transduction [222–224].

9.5.1 Computer resources yesterday and today

It is interesting to note that the minimization of the all-atom DHFR complex (10,800 iterations), required about 15 h of CPU time on the Cyber 205 supercomputer, one of the fasted machines available at the time. To put this in perspective the Cyber 205 (~ $18,000,000) processor was ~ 50–80 MHz with 32 MB Ram and 16 GB of memory, while the current iPhone 7 has two 64-bit 2.34 GHz cores and up to 320 GB of memory, and 2–3 GB RAM, so the calculation would take only ~ 30 min on a $600 iPhone today.

Table 7 Comparison of calculated and observed geometrical and vibrational properties of amides

Molecule		Δb (Å)	Δθ (deg)	$\Delta\nu^{a}$ (cm^{-1})	Δμ (debye)
Formamide		0.035	1	54	0.57
Acetamide		0.040	3.2	33	0.57
Propionamide		0.03	1.4	32	0.19
N-methyl-formamide (NMF)		0.013	0.7	45	0.44
N-methyl-acetamide (NMA)		0.015[b] (0.012)	2.3[b] (1.5)	21	0.26
N-methyl-propionamide (NMP)				20	0.41
N,N-dimethyl-formamide (DMF)		0.023		30	0.38
N,N-dimethyl-acetamide (DMA)				17	0.07
Pyrrolidone				20	0.58

9.6 United atom approximation

As we discussed above the united atom model was being widely utilized at the time of these calculations. We took the opportunity afforded by these simulations to compare the performance of this approximation with that of the full-atom representation of the DHFR-ligand complex [225]. Both the minimized all-atom and united-atom structures were compared to the experimental structure in order to assess the effect of omitting nonpolar hydrogens. A superposition of each minimized structure onto the experimental structure using a least-squares fit of the heavy atoms of the secondary structures, i.e., the alpha helices and beta strands was carried out. The root mean square (RMS) deviation between several structural segments of interest was then calculated. The results obtained are reproduced in Table 8.

As might be expected the united atom model introduced significant deviations over those obtained when all hydrogens were included. The fit to the experimental structure at the active site is of special importance and here we saw that the deviations were almost twice as large for the united-atom model (Table 8). As noted, though at the time the united atom was almost universally used because of the computational resources it saved, it has now, rightly so, been abandoned and essentially all simulations include all hydrogens (as well as surrounding water molecules). This serves as a cautionary example as often the outcome of approximations that seem to be required to allow faster achievement of results may correspond to "throwing the baby out with the bathwater".

Table 8 RMS deviations for heavy atoms and C, atoms between the minimized and the observed structures (in Å)

Region	All atoms	United atoms
All heavy atoms	1.23	1.48
Active site	**0.58**	**1.03**
α Helices	1.10	1.50
β Strands	0.82	1.02

10 Water models

Initially, molecular mechanics studies focused on gas phase calculations, i.e. each molecule was considered in isolation. It was also possible to simulate crystals and neat liquids. However, for applications to systems of biological importance it is necessary to include the effect of solvent, as all biological activity takes place in an aqueous environment. Several treatments of water have been discussed above and as noted, literally hundreds of potential models have been developed for this uniquely important molecule. A thorough review of these potentials is well beyond the scope of this perspective and the reader is referred to several recent papers and the references therein [226–231]. Here we review briefly the early work in the field and the evolution of some of the most popular water force fields still used in the majority of simulations today.

Models for water were suggested as early as the first half of the twentieth century, but their validity for solvent simulations was tentative due to limited experimental data. Numerous models have been put forward over time, differing in complexity, and the number and placement of interaction centers. In general, only one center, the oxygen atom, was used for the non-bonded repulsion/dispersion term, positive partial charges were located on hydrogen atoms, while different number and locations for the sites of negative charge have been offered. A schematic representation of various water models is given in Fig. 9.

10.1 1930s and 1950s, Bernal and Fowler and Rowlinson models form the basis of current water models

A simple model with a pairwise potential energy function was first derived in an elegant study by Bernal and Fowler in 1933 [8]. Considerations of a variety of experimental data including the structure of ice, the diffraction pattern of liquid water, as well as the dipole moment, led them to a model consisting of three point charges: positive charges on each of the hydrogens and a negative charge on the HOH bisector 0.15 Å from the oxygen (Fig. 9a), with the van der Waals interaction centered on the oxygen atoms. This was shown to lead to reasonable agreement with the sublimation energy of ice along with other physical properties of water [8]. In 1950 Rowllinson [201] suggested an atomic model which included a 6–12 non-bonded function for the oxygen atoms and a tetrad of point charges: a positive charge on each hydrogen atom and a pair of negative charges near the oxygen atom, about 0.5 A in the direction perpendicular to the plane of the molecule (Fig. 9c). This was shown to account for the lattice energy and intermolecular spacing in ice. Interestingly, Rowlinson included the effects of polarization and induced dipoles in his calculation.

Similar models were suggested in the 1970s by Stillinger and coworkers. The first, the BNS potential of Ben-Naim and Stillinger [232] was exploited by Raman and Stillinger to carry out the first molecular dynamics study of water [233], exploiting the MD technique first introduced by Alder and Wainright in 1959 and'60 [234, 235]. It is worth digressing

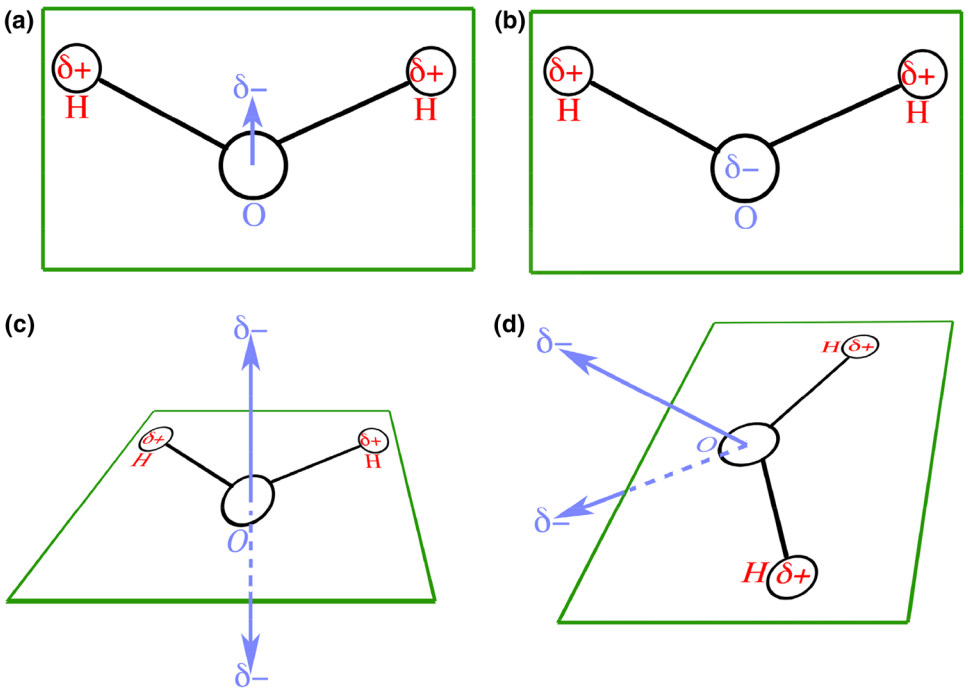


Fig. 9 Schematic of charge sites for several early water models. **a** Bernal and Fowler [8], TIPS2 [203], TIP4P [203]; **b** SPC [202], TIP3 [216], TIP3P [203]; **c** Rowlinson [201]; **d** Ben-Naim and Stillinger (BNS) [232] and Stillinger (ST2) [233, 236]. The plane of the molecule is indicated by a green parallelogram, oxygens are blue and hydrogens are red. δ's are the partial charges (+ for charges on hydrogens and − for charges on or near the oxygen)

for a moment to read Alder and Wainright's description of their methodology, to gain some perspective on the technological advances made since the first molecular dynamics simulation was carried out.

> As the calculation of the motions of the many particle system proceeds, the history of each particle is recorded on *magnetic tape*. A separate process later analyzes this very detailed data for various desired quantities. The reason for such a two-step procedure is that the size of the computing machine's memory does not allow the analysis to take place at the same time the motions are being calculated. However, a few easily calculated quantities such as the pressure, collision rate, and potential energy are obtained during the first phase of the calculation. This makes it possible to judge when the system has reached equilibrium. *To conserve tape* the positions and velocities of all the particles are put on tape only infrequently, but for each collision the positions and velocities of the two colliding particles are written on the tape.
> The calculation can also be *monitored by means of a cathode-ray tube* which is attached to the computer and which forms a picture of the system after each time step. On the face of this tube are plotted in plane-projection the positions of the centers of the particles. *A camera focused on the face of the tube* for a large number of time steps will record the trajectory of each particle as a succession of dots. A slightly different monitoring routine projects an identifying symbol at the position of each molecule so that the beginnings and ends of molecule migrations can be located [234]

Raman and Stillinger then slightly modified the BNS model, resulting in the ST2 potential. Both models differed from the Rowlinson model in the positioning of the pair of negative charges—instead of placing them directly above and below the oxygen atom, they were placed in a tetrahedral geometry (Fig. 9d). The major difference between the original BNS potential and the ST2 potential was the positioning of the negative charges which originally were placed 1 Å from the oxygen in the BNS potential and then repositioned to 0.8 A in the improved ST2 potential [236]. The latter model was widely accepted and utilized in a variety of simulations of biological molecules in aqueous environments.

In 1981 Jorgensen [216] sought to derive a simpler water potential that was transferable to solute–solvent systems and economical to calculate. He also included van der Waals interactions for the oxygen but included charges centered only on the hydrogens and oxygen. Not only do the single atom centered charges require less computer resources, but the potential can be combined straightforwardly with solute systems using standard combination rules. The parameters were chosen to yield reasonable structural and energetic results for both gas-phase dimers and pure liquids and was termed TIP3, Transferable Intermolecular Potential, with three centers [216] (Fig. 9b).

At the same time, and for the same reason, i.e. "…for molecular dynamics simulations of hydrated proteins a simple yet reliable model for the intermolecular potential for water is required" Berendsen et al. [202] derived the SPC (Simple Point Charge) model for water. Like the TIP3 potential they also included a 6–12 potential for the oxygen and three point charges, each positioned on the atomic centers (Fig. 9b). In this case the values of the parameters were optimized by running molecular dynamics simulations with varied parameter values, and requiring that both potential energy and pressure agreed with experimental values. They also calculated the density, radial distribution function, and diffusion constant, compared the results to the corresponding experimental values, and concluded that the potential satisfactorily accounted for the properties of liquid water.

10.2 The TIP models

Jorgensen et al. [203, 217] subsequently suggested several other models—TIPS2 [217], TIP3P [203], TIP4P [203] and TIP5P [237]. They then carried out a thorough benchmark study aimed at comparing these models and the other existing models BF, ST2, and SPC [202]. TIP3P was a 3-point model, based on their earlier version, TIP3, but with an improved set of parameters. Although the three point models were convenient from a computational point of view there were some flaws in reproduction of experimental results. TIPS2 and TIP4P models used 4 points to represent the molecule: A dispersion/repulsion center on the oxygen and 2 positive partial charges on the hydrogens, as in the 3 point models, but the oxygen charge was moved towards the hydrogens, along the HOH bisector. TIP4P and TIPS2 differ only in their parametrization. Both of these models were based on the 4 point model first proposed by Bernal and Fowler in [8]. TIP5P adopted the five-site geometry of the ST2 form. The parameters for the TIP potentials were derived and tested by performing Monte Carlo simulations for liquid water in the NPT ensemble at 25 C and 1 atm. Comparisons were made with experimental thermodynamic and structural data including radial distribution functions, self-diffusion constants, density, energy, heat capacity, and isothermal compressibilities. All models were successful to a degree except for the original BF model, which significantly overestimated the density and gave poor structural results. They concluded that overall the SPC, ST2, TIPS2 and TIP4P models gave reasonable structural and thermodynamic descriptions of liquid water, but the four point models were in general superior to the 3 point models especially in reproducing density and the radial distribution function.

10.3 Further water potential developments

This is, as mentioned above, an extremely brief summary of water potential development, which continues to this day. For example, both the SPC and the TIPS potentials continued to evolve, serving as a basis for several efforts to further improve them. This includes the introduction of polarizable force fields through fluctuating charges (FQ) by Rick et al. resulting in the TIP4P-FQ, and SPC-FQ models [238], and the further extension of this to couple Lennard-Jones parameters to the atomic charges in the TIP4P-pol and SPC-pol force fields [239]. In 2004 Horn et al. reparametrized the TIP4P FF for use with Ewald sums (TIP4P-Ew) [240], and in 2005 Abscal and Vega reparametrized it by including the properties of ice polymorphs (TIP4P/2005) [241]. Bauer et al. further modified the TIP4P-FQ potential in 2009, by coupling the "hardness parameter" in the charge equalization method to the atomic charge resulting in a charge dependent polarizability, the TIP4P-QDP model [242], and then introduced a further coupling between the charges and the LJ parameters, TIP4P-QDP-LJ [243]. One of the most recent potentials in this progression, for use in biomolecular simulations is the development of the classical polarization Amoeba model [228] for use with the Amoeba Biomolecular force field [209].

The analytical description of the water potential surface now spans some 80 years. In many ways, the evolution of this field brings to mind Simon and Garfunkel's famous lyrics from The Boxer, "After changes upon changes, we are more or less the same." Thus the first potential for water presented in 1933 by Bernal and Fowler was a "4-point" model containing off-atom charges [8]. This was followed by Rowlinson's potential in 1951, which refined this model to a "5-point" potential and included polarizability [201]. Mainly, due to the limitations imposed by computational resources, development of water potentials turned to *less* sophisticated (and accurate) representations of molecular interactions for many years, with simple fixed charge atom centered models being optimized. Starting roughly in the 1980s additional charge centers were reintroduced, and later polarization effects began to be taken into account once again. This summary barely scratches the surface of this rich field [244], but hopefully gives a feel for the efforts to develop adequate potential surface for accounting for the solvation effects of this important molecule in biomolecular simulations.

11 Second generation FFs

11.1 Enhancement of the MMX force fields: MM3 and MM4

In the years following the publication of MM2, Allinger carried out a unique study, cataloging and publishing a list of some 25 discrepancies in properties calculated with MM2, found in his lab or noted in the literature [111, 245]. This is a practice which would benefit, and lead to systematic improvements of all force fields and perhaps encourage a systematic comparison of FF performance. Indeed, toward the end of the 1980s Allinger reformulated the MM2 FF to address these itemized deficiencies, which suggested problems not only with the force constants, but also with the functional form. In addition he adopted the Lifson CFF approach [26, 55, 85] and extended the requirements of the force field to fit not only structure and thermodynamic data but also account simultaneously for crystal packing [246] and vibrational properties [247].

Correction of the deficiencies in the performance of MM2 required significant elaboration of the functional form. While MM2 was approximately a Class I diagonal quadratic FF, with the exception of angle–bond interaction coupling, and a sixth power angle bending dependence (see Eq. 17b), MM3 [111, 247, 248] and MM4 (below) are Class II FFs which take into account the extensive coupling between internal deformations, similar to those used by Schachshneider [20], Ermer and Lifson [55, 249] Lifson and Stern [250] and Dauber-Osguthorpe et al. [225] The MM3 force field [111, 247] for hydrocarbons was able to account for the structures and energies, including heats of formation, conformational energies, and rotational barriers, of a wide variety of hydrocarbons. These included a large number of strained and cyclic compounds [111]. In addition the vibrational frequencies of eight simple hydrocarbons [247] and the crystal packing of nine saturated and aromatic hydrocarbons [246, 248] were reproduced. The force field was later extended to a large number of important classes of organic compounds in a series of subsequent studies. These included, among others, functional groups found in amino acid side chains, e.g. alcohols and ethers [251], amines [252], thiols sulfides and disulfides [253, 254], carboxylic acids and esters [255] and aromatic heterocycles [256], as well as molecules representative of the peptide backbone [257].

11.1.1 Coupling terms and anharmonicity

Coupling terms that were found necessary to account for deficiencies in structures and vibrational frequencies were tested and included where helpful. Extra polynomial terms were added to the bond and angle terms (Eqs. 17a, b), mainly to avoid spurious effects on strained bonds or angles due to the negative contributions from odd polynomial terms. The stretch-bend coupling term (Eq. 17e) was maintained as in MM2, and bond–torsion (Eq. 17f) and bend–bend terms (Eq. 17g) were added. A torsion–bend term was considered in some studies after the publication of MM2 but found to be too small to matter in MM3. The Nonbonded functions

were made softer overall by decreasing the exponent in the Buckingham potential from 12.5 to 12 (Eq. 17h).

Electrostatic interactions were added to unsaturated hydrocarbon carbons and hydrogens (which enabled a stable perpendicular packing of aromatic rings) though the charges on saturated carbons were maintained as zero. Similarly, electrostatic energies were included for compounds containing nitrogen and oxygen. Bond dipoles rather than point charges were used as before, except when there was a full formal charge. However, a significant change compared to MM2 was the omission of lone pairs on these atoms. It was found that the observables (especially frequencies) could be accounted for as well without these, while eliminating the technical problems created by having the 'virtual atoms' [251, 252]. An explicit, angle dependent hydrogen bond function (Eq. 17i) was added in later studies [258].

11.1.2 An explicit hydrogen-bonding function?

The addition of the H-bond function is a curious choice in the development of the MMn force fields. Allinger has done some of the best work in the field, characterized by testing against perhaps the most extensive set of experimental data used to date in force field development. In fact, it would behoove those of us who continue to develop advanced force fields to "go back" and test against the wide variety of MM3 and MM4 data, and ensure that the new force fields fit these at least as well as Allinger did. The choice of the explicit hydrogen bond potential however departed from this paradigm, and instead Allinger relied solely on quantum mechanical data, and only the parameters of the H-bond potential were optimized. Thus, the parameters in the H-bond potential were forced to absorb any existing deviations in the FFs inability to fit the QM data (which also might arise because the FF was derived from experimental data). Nor was the FF back tested against crystals, as had been done previously (or on liquid systems). These are some of the most powerful experimental observables for assessing H-bonding, as we have seen above. Given this, it would have been desirable to apply the proposed, more complex function to the experimental data establishing that an explicit HB function was indeed necessary [87, 259],and that additional parameters were justified.

The resulting MM3 FF is given in Eq. (17).

$$E = E_b + E_\theta + E_\phi + E_{Coupling} + E_{Non\text{-}bonded} + E_{HB} \tag{17}$$

$$E_b = \sum_b K_b\left[(b-b_0)^2 - 2.55(b-b_0)^3 + 1.5(b-b_0)^4\right] \tag{17a}$$

$$E_\theta = \sum_\theta K_\theta\left[(\theta-\theta_0)^2 - 0.014(\theta-\theta_0)^3 + 5.6 \times 10^{-5}(\theta-\theta_0)^4 - 7.0\times10^{-7}(\theta-\theta_0)^5 + 9\times10^{-10}(\theta-\theta_0)^6\right] \tag{17b}$$

$$E_\phi = \sum_\phi K_{\phi1}(1+\cos\phi) + K_{\phi2}(1-\cos2\phi) + K_{\phi3}(1+\cos3\phi) \tag{17c}$$

$$E_{Coupling} = E_{b\theta} + E_{b\phi} + E_{\theta\theta'} \tag{17d}$$

$$E_{b\theta} = \sum_{b\theta} K_{b\theta}(b-b_0)(\theta-\theta_0) \tag{17e}$$

$$E_{b\phi} = \sum_{b\phi} K_{b\phi}(b-b_0)(1+\cos3\phi) \tag{17f}$$

$$E_{\theta\theta'} = \sum_{\theta\theta'} K_{\theta\theta'}(\theta-\theta_0)(\theta'-\theta'_0) \tag{17g}$$

$$E_{Non\text{-}bonded} = E_{Buckingham}(r_{ij}) = A_{ij}\exp(-B_{ij}r_{ij}) - \frac{C_{ij}}{r_{ij}^m} \tag{17h}$$

$$E_{HB} = \varepsilon_{HB}\left\{184000\exp\left[-12(R_{YH}/r)\right] - F(\beta,R_{XH})\times2.25(r/R_{YH})^6\right\}/D; \quad F(\beta,R_{XH}) = \cos\beta(R_{XH}/R^0_{XH}) \tag{17i}$$

where E_b and E_θ are the strain energies associated with bond (b) and valence angle (θ) distortions from their reference values b_0 and θ_0; $E\varphi$ is the energy of rotation about the dihedral angle φ. $E_{coupling}$ and $K_{coupling}$ are the energy and force constants of coupling interactions between internals. The hydrogen bond function, E_{HB} has the form of a Buckingham vdW function, where ε is an H-bond constant, r is equilibrium hydrogen bonding distance, R_{YH} is the hydrogen bonding distance Y···H, cosβ is the cosine of angle X–H···Y, R_{XH} is the X–H bond length, R^0_{XH} is the reference X–H bond length and D is the dielectric constant. F introduces an angular dependence to the hydrogen bonding energy.

11.1.3 MM4

The MM3 FF was continuously improved, and in the mid 1990s Allinger embarked on a major reparametrization, developing a new version, MM4. As done previously deficiencies were noted and collected over the intervening years and enumerated in the first paper in the MM4 FF series [112]. The first version of MM4 was presented in 1996 for hydrocarbons [112, 260, 261], and further functional

groups, including alkenes and conjugated molecules [260, 261], amides [262], amines [263], ethers, alcohols, and carbohydrates [264] were added over the next decade. A variety of deficiencies were addressed, including the lack of accuracy in vibrational frequencies, errors in rotational barriers in congested molecules, inclusion of vibrational energy in heats of formation, and some refinement in structure, among others. In addition modifications were made to account for electronegativity and hyperconjugation effects (such as the Bohlmann and anomeric effects) [261]. The general definition for the cross terms used in MM4, $E_{Coupling}$, is shown in Eq. (18a). The first three terms, the torsion–stretch, bend–stretch and bend–bend were carried over from MM3 (Eqs. 17e–g) and again were found to be critical. The new terms added in MM4 are defined in Eqs. (18b–e) [265].

$$E_{Coupling} = E_{b\theta} + E_{b\phi} + E_{\theta\theta'} + E_{bb'} + E_{\phi\theta} + E_{\theta\phi\theta'} + E_{\phi\phi'} \tag{18a}$$

$$E_{bb'} = \sum_{bb'} K_{bb'}(b - b_0)(b' - b'_0) \tag{18b}$$

$$E_{\phi\theta} = \sum_{\phi\theta} [K_{\phi\theta 1}(1 + \cos\phi) - K_{\phi\theta 2}(1 + \cos 2\phi) + K_{\phi\theta 3}(1 + \cos 3\phi)](\theta - \theta_0) \tag{18c}$$

$$E_{\theta\phi\theta'} = \sum \theta\phi\, K_{\theta\phi\theta'}(\theta' - \theta'_0)(\theta - \theta_0)\cos\phi \tag{18d}$$

$$E_{\phi\phi'} = -\sum_{\phi} K_{\phi\phi' 1}(1 - P_1)(1 + \cos 3\phi_1) K_{\phi\phi' 2}(1 - P_2)(1 + \cos 3\phi_2) \tag{18e}$$

$$E_{oop} = E_{BDCA} + E_{BADC} + E_{BCAD} = K_{BD}(1 - \cos 2\omega_1) + K_{BC}(1 - \cos 2\omega_2) + K_{BA}(1 - \cos 2\omega_3) \tag{18f}$$

The improper dihedral angles defining the out of plane energy are denoted ω. The torsion–bend, (not regarded as important in MM3) was required to explain the effect of torsion on angle behavior and hyperconjugation effects for conjugated molecules [260, 265] and for carbohydrates [264]. A stretch–stretch and torsion–bend were added as well. For the first time, a third order term, bend–torsion–bend was also introduced [262] as described by Warshel and Lifson [52]. These cross terms were needed in particular to improve the fit to vibrational frequencies.

The definition of the out of plane was also modified [265]. In MM3 it was defined as a bend, while in MM4 it was defined as an improper torsion angle as depicted in Fig. 10 below. For a planar arrangement where atom B is bonded to atoms A, C and D, the out of plane energy is

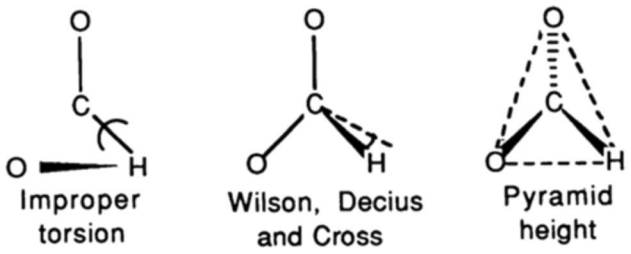


Fig. 10 Out of plane coordinate

defined as the sum of three energy terms for the three possible improper torsions,

An explicit hydrogen bond function was used again, but an extra functionality was introduced to include not only the X–H···Y angle, but also the angle with the lone pair, i.e. the H···Y–Z angle where Z corresponds to the lone pair on the acceptor, Y [266].

11.2 Conversion to all-atom FFs

11.2.1 AMBER

As noted above, unlike the MMx and CFF series of FFs, AMBER, CHARMm, OPLS, and Gromos, all invoked the united atom approximation in their initial incarnations. With the exception of Gromos all FFs have since abandoned this rather drastic construct. Amber has undergone several stages of development following introduction in 1981 [210], and reparametrization in 1984 [106]. The first major refinement involved the restoration of all nonpolar atoms in 1986 resulting in the introduction of the all-atom FF [213]. This FF has subsequently undergone many iterations of further modification, which continues to the present [113]. Several key milestones in this second stage of development were the PARM94 [267], PARM99 [268] and PARM99SB [179] FFs discussed below.

PARM94: In general, "a minimalistic strategy" was used in deriving the second generation PARM94 AMBER FF [267], both with respect to choosing the analytical form and with respect to assigning numeric values to the parameters. The function is a simple harmonic, diagonal functional form, with no cross-terms nor Urey–Bradley term to take into account coupling between internals, and, for the most part, a single Fourier term for the torsions. In general, parameters were chosen independently by fitting the properties of a few simple compounds which determine that parameter, rather than using a least squares optimization to optimize all parameters simultaneously (which could contribute to yielding a more consistent set).

Bond and angle reference values were taken directly from X-ray data, rather than deriving them in such a way that the minimized values fit the experimental ones. They point out that deviations in these internals are not very important for conformational studies and that in any case errors introduced by this approach were smaller than "the inaccuracies remaining in the dihedral and non-bonded (charge and VDW) parameters." However, it is known that angle deformation has an important effect on the torsional potentials and plays a key role in lowering rotational barriers, so this assumption with regard to angle reference values may not be fully justified [54, 55, 112, 269].

The force constants for bond stretching and angle bending of pure single bonds and double bonds were obtained from vibrational frequencies of a few simple model compounds such as ethane, benzene, and *N*-methylacetamide, and all others were assigned by interpolation using the observed bond length as a measure of the single/double bond character. For dihedrals, a single threefold, V_3, term was used for bonds with sp^3 hybridization. When partial sp^2 hybridization is involved, a V_2 term was used with a force constant interpolated between single and double bond according to the bond length.

Special emphasis was given to the non-bonded part of the potential, with the objective of reproducing condensed phase properties. A new method was introduced to determine charges by fitting ab initio electrostatic potentials (ESP). The derivation differed from the original Weiner et al. [210] method by applying restraints to the charges obtained on ill-defined "buried" atoms. The restraints kept the charges on buried non-polar atoms to a minimum and were observed to minimize the variability of derived charges with conformation [270].

The new van der Waals parameters were derived from liquid simulations. Most of these parameters were taken from the OPLS potential, but additional Monte Carlo simulations were needed to derive parameters for carbon atoms because at the time OPLS only provided united-atom parameters. In general, simple rules dictated the assignment of vdW parameters—the atom type and hybridization determined the type of parameters. Exceptions were polar hydrogens, which had zero size, and hydrogen-bound sp^3 nitrogens and oxygens that required larger atom sizes to compensate for the zero size of the hydrogens. The new charges and vdW parameters had the consequence that a special HB function (such as the 10–12 used in the original Amber FF) was no longer needed. Thus the Hagler–Lifson [85, 86] formulation of the hydrogen bond was adopted, where the HB interactions were accounted for simply by the nonbond and electrostatic interactions with a negligible vdW parameter for the hydrogen bonding hydrogen. Finally, a scaling factor was applied to charge interactions between atoms separated by three bonds (1–4 interactions).

11.2.2 Series of reparametrizations of φ, ψ torsions to account for conformational properties of peptides

Unfortunately it emerged that application of the force field to peptides revealed problems in accounting for their conformations, in particular in the (φ, ψ) maps of glycyl and alanyl dipeptides [267]. Here, the authors "were forced to add explicit dihedral parameters in order to reproduce the ab initio quantum mechanical energies for these models." As the authors state, "In the studies described above, the major weakness was the necessity of adding dihedral potentials for the φ and ψ of peptides … without obvious physical justification." In fact these potentials were later found to be much too steep and introduced artifacts into peptide conformational studies requiring reparametrization, as discussed below [179]. This has been a chronic problem in all the diagonal quadratic FFs including AMBER, CHARMm, and OPLS and there have been ongoing revisions of these FFs involving modification of these torsion constants continuing to the present day.

This is a known problem with optimizing a few select parameters to account for deficiencies in a limited set of observables without at least further testing on a wide range of data. The selected parameters are being required to account for all errors, even if they are due to problems in different parameters in the force field, and can (and do) lead to artifacts. This is one of the reasons to strongly pursue Allingers' method of accumulating numerous deficiencies/deviations found from applying the FF to numerous other systems, and then consistently reoptimizing all parameters on an enhanced set of data including all previous observables—with improvement of the functional form if required.

PARM99: Another round of Amber FF improvements (PARM99) was published in 2000 [268]. In this enhancement the torsion energies were further optimized by adding additional Fourier components and defining additional torsion types to the Parm94 force field (i.e., reducing transferability). For example, the torsional parameter (X—C_{sp^3}—C_{sp^3}—X) which had been transferable to all torsions independent of "X" was replaced by three specific types: CCCC CCCH and HCCH. Two additional Fourier terms were also added to the functional form of the torsion energy expression. To derive the new set of parameters they employed the conformational energies of a set of 34-molecules, which had been studied by both high-level ab initio calculations (GVB/LMP2) and experiment, and an additional set of 55-molecule for which high quality experimental data were available. The combined set of compounds was comprised of a diverse set of organic compounds representing ten classes, including model compounds for peptides, nucleic acids, and carbohydrates.

An automated algorithm "parmscan" was introduced in this work, in order to systematically scan torsion constants and find the best Fourier series and force constants so as to reproduce the energy difference for the 82 conformational pairs comprising the training set. In this way, a new set of torsional functions and force constants were obtained. The results were compared to those obtained for the same set of 82 conformational energy differences as calculated by MMFF, MM3, and CHARMm and it was shown that the new FF gave the smallest RMS deviation. A concern with the methodology is that in some cases the "… minimized conformations of conformer pairs deviate too much from reference structures" and in such cases the conformers were restrained to the reference structure [268]. Thus although the force field reproduced conformational energy differences, in some cases it wasn't able to reproduce the corresponding conformation.

11.2.3 One of first inclusions of polarizability (post Rowlinson)

In addition to the modified torsion parameters, the effect of accounting for atomic polarizability, and improving the representation of the charge distribution by adding lone pair orbitals to the nitrogen and oxygen atoms were tested. This was done by assigning atomic polarizabilities resulting in induced atomic dipoles by the field of the rest of the partial charges and other induced atomic dipoles as in Eqs. (19a–e),

$$E_{pol} = -1/2 \sum_i \mu_i E_i^0 \tag{19a}$$

$$\mu_i = \alpha_i E_i \tag{19b}$$

$$E_i = E_i^0 + \sum_{j \neq i} T_{ij} \mu_i \tag{19c}$$

$$E_i^0 = \sum_{j \neq i} q_i \frac{\mathbf{r}_{ij}}{r_{ij}^2} \tag{19d}$$

$$T_{ij} - \frac{1}{r_{ij}^3} \left(3\mathbf{r}_{ij} \frac{\mathbf{r}_{ij}}{r_{ij}^2} - 1 \right) \tag{19e}$$

where α_i is the polarizability of atom *i*, q_j is the charge; $\mathbf{r}_{ij}$ is the vector between atoms *i* and *j*. Eqs. (19b) and (19c) are iterated to self-consistency during the minimization and dynamics simulations.

The results of this "enhancement" were disappointing. "Without changing the torsional parameters, the use of more accurate charges and polarization leads to an increase in average absolute error compared with experiment, but adjustment of the parameters restores the level of agreement found with the additive model" [268]. Thus, even with the addition of supplementary parameters describing the polarization, and additional charge centers, *no* improvement in accuracy could be achieved.

This is a puzzling result, and a hint to the cause may be divined from the initial result, where the fit actually worsened with the introduction of the new terms when the remainder of the force constants were not reoptimized. As noted above, errors in the energy representation must be absorbed by the limited set of parameters being optimized. Thus, the original parameters were a compromise between the "best intrinsic parameters for the given internal, or charge representation", and the "effective" parameters needed to absorb errors in other terms (such as omission of polarization, realistic description of charge density, inaccurate combining rules, etc.). When the erroneous terms or omitted interactions are corrected, the effective parameters now introduce errors. Thus, when the torsions were reoptimized, the fit improved, although still not better than the original non-additive force field. This may well be because *all* parameters weren't reoptimized, including nonbond and internal force constants. These also were/are "effective parameters" accounting for the same errors. In fact, it is likely necessary to reoptimize the entire force field against all previous observables, and additional data, when adding additional interaction terms. The problem is that the effective terms required to fit deviations don't have the same functional form (distance or coordinate dependence) as the missing or erroneous potential terms. Thus, they can be forced to fit a limited set of data over a restricted conformational space, but inevitably artifacts will be introduced in regions outside the "training set" space. This problem has persisted in subsequent attempts to improve FFs by adding polarizability, as discussed in the following paper assessing the applicability and validity of modern FFs [113]. That is, though clearly important in determining the energetics of molecular interactions of polar molecules, as Wang et al. found [268], attempts to account for polarizability have yet to lead to significantly better agreement with experimental or QM energetics or structure [113].

Besides the organic molecules in the test set, Wang et al. [268] once again re-optimized the φ, ψ torsional parameters in peptides. High-level quantum calculations were used to obtain a training set consisting of six alanine dipeptide conformational energies, and 11 alanine tetrapeptide conformational energies. The latter were then used to test the parameters. Significant improvements were obtained over the results previously obtained with Parm94 [267]. However once again only the φ, ψ torsional parameters were optimized, in the context of all other fixed interaction terms, to fit a limited set of data. From the above discussion we would

expect this to lead to further artifacts, and this was borne out by Hornak et al. in their subsequent analysis [179].

PARM99SB: as described above, parameters for the φ, ψ torsions were treated separately and were somewhat problematic. In 2006 [179], Hornak once again set out to reoptimize these parameters to address limitations which had arisen over the years, such as over-stabilization of α-helices [179]. However, the same procedure of keeping all other force constants fixed was again employed. Three terms in the expansion of the dihedral angle potentials were retained and the parameters were optimized against high level ab initio calculation of blocked glycine and alanine tetrapeptides (–Gly_3–, –Ala_3–). The energies of 28 glycine and 51 alanine tetrapeptide conformers were calculated at the level QM LMP2/cc-pVTZ (-f) level, and were used in the fitting. The RMSD error in energies obtained from the parametrization was 1.17 kcal/mol for the glycine tetrapeptides, and 1.31 kcal/mol for the alanine peptides. The resulting modified force field termed Parm99SB, was then exhaustively tested against a diverse set of experimental data and compared with a variety of Amber FF variants. The comparisons included simulations of peptide conformational preferences, which were compared to distributions in the PDB and perhaps more importantly to results from spectroscopic data of di, tri and tetra peptides in solution. In addition, "decoy analysis" of longer peptides was carried out to make sure that the low energy conformers indeed corresponded to the experimental structures. Finally, the FF was tested in proteins by comparing calculated dynamics parameters of lysozyme and ubiquitin to NMR relaxation experiments. The results showed that Parm99SB successfully accounted for these properties, showing significant improvement over the other Amber FFs, and correcting the known deficiencies in them. Subsequent studies however, revealed remaining deficiencies leading to continued iteration of this process of torsion optimization, one which has to date proved fruitless as the deficiencies are almost certainly in the inability of the simple quadratic diagonal; 12-6-1 potential to account for the physics of biomolecular interactions [113].

11.3 CHARMM

In the same time period the developers of CHARMM [271] also recognized the need for explicit hydrogens and improvements in the force field. In this reparametrization of the force field the functional form used in the original FF was retained [109], namely a Urey–Bradley harmonic, diagonal force field with a 6–12 nonbond potential (Eq. 10). As in Amber it was recognized that no explicit hydrogen bond term was required and the Hagler–Lifson [85, 86] formulation of the H-bond was adopted [271] instead of the previously employed explicit forms [108, 272]. In some specific cases including aliphatic carbons, amide nitrogens and oxygen atoms, scaling of the 1–4 Lennard-Jones term was applied. The resulting FF was labeled CHARMM22 [271] (as it was introduced in the 22nd version of the CHARMM program).

11.3.1 An extensive training and validation set comprising diverse observables

Parameter optimization and testing was thorough and involved a very extensive set of ab initio and experimental data. Optimization was carried out in cycles, starting from previous CHARMM nonbond parameters. With these values the intramolecular valence parameters, (bond, angle, out of plane, Urey–Bradley, and torsion) were optimized using the structure and vibrational frequencies of one or two model compounds. Given these intramolecular parameters the non-bond parameters were optimized to fit condensed phase observables. It was found that automated least-squares procedure often led to a set of "unphysical" parameters that fit the training set, in part because of correlation, indicating that the parameters were underdetermined. As a result, the steps were iterated manually until convergence was achieved.

Optimization of the parameters for the intramolecular deformations of the protein backbone exploited two molecules, *N*-methylacetamide and the alanine dipeptide (Fig. 1). For NMA (schematic **1** R=CH_3) the experimental data included geometries, vibrational spectra, as well as thermodynamic data for liquid NMA and for NMA in aqueous solution. The dihedral potentials for φ and ψ were parametrized mainly based on high-level ab initio calculations of 15 conformational energies of the alanine dipeptide with special attention paid to the C^7_{eq}, C^7_{ax}, and C^5 conformations, as well as some information from survey results in proteins. The potential parameters for side chains were not presented, but were used in the simulation of the various training and test systems.

Good agreement was obtained for the geometries of the two compounds, though deviations of as much as 10° in dihedral angles were observed for the alanine dipeptide. There was a discrepancy in the C′–C_α–N angle, which was corrected by scaling of the 1–4 nonbonded interaction between nitrogen and oxygen (note this is caused by the torsion dependence of this angle [271], a manifestation of angle-torsion coupling). Deviations in the vibrational frequencies of NMA were typical of those that can be obtained by diagonal quadratic force fields with deviations of 100 cm^{-1} not uncommon (e.g. the CCN bend deviated by140 cm^{-1}), as opposed to the 30–50 cm^{-1} or less typical of force fields including coupling terms [55, 261, 273], and as seen in Table 7. A good fit was also obtained to the conformational energies of the alanine dipeptide. However, a red flag appeared in that the resulting dihedral parameters included systematic errors in the observed torsion angles of myoglobin. Thus, adjustments were made in order to lower

the energy of the α_R conformation. As the authors point out in regard to discrepancies in φ and ψ, these deviations may be due to inherent limitations in the functional form of the potential (e.g. a lack of coupling terms, polarization effects, atomic multipoles etc. as discussed in following paper [113]).

Following this modification, the force field was further tested against additional high level ab initio results for the energies of seven conformers of the alanine dipeptide. Overall good agreement (within ~1 kcal/mole) was achieved for the low energy regions of the φ, ψ map C^7_{eq}, C^7_{ax}, C5, α_R, and β, which are the most important regions for accounting for protein properties. However unacceptably large energetic discrepancies of ~6 and 3 kcal/mol were found for the α_L (67.0, 30.2), and β_2 (−130.9, 22.3) conformers, respectively. Again, this may be pointing to inadequacies of the simple diagonal harmonic functional form of the force field. As the authors pointed out, high energy regions may be visited during MD trajectories, so this is something to be aware of.

The parameters were also tested against the vibrational frequencies of the alanine dipeptide obtained by both (scaled) ab initio and experimental frequencies. Overall comparison of the empirical and ab initio data, excluding NH wagging frequencies, for the C^7_{Eq} and C^5 structures showed rms differences of 33 and 56 cm^{-1}. The aforementioned NH wagging frequencies showed very large deviations of as much as 400 cm^{-1} indicating a problem with the related force constants. The fact that this frequency is well accounted for in NMA may indicate the importance of coupling terms which are not included. Comparison with the experimental vibrational frequencies was limited by the lack of adequate assignments. Five typical vibrational modes were compared and in general deviations of the order of 100 cm^{-1} were obtained, consistent with the results for NMA.

11.3.2 Testing the CHARMM22 FF against crystal observables

The non-bonded parameters (partial charges and LJ parameters) were also determined using NMA. "Observables" included ab initio interaction energies and geometries of complexes of NMA with water and the NMA dimer, as well as the dipole moments, experimental heat of vaporization, heat of solvation, and molecular volume of liquid NMA. The energies and geometries of the NMA–water complexes were found to be in excellent agreement with the QM results, though the intermolecular distances were too short by ~0.2 Å. Heats of vaporization, solvation, and molecular volumes were also in excellent agreement with experimental values. The force field was further tested by MD and minimization studies of NMA and Ala dipeptide crystal structures. Good agreement was found for the NMA crystals, but some discrepancies were pointed out for the dipeptide crystal. While the peptide bonds deviated significantly from planarity in the crystal structure, they remained essentially planar in the simulations. The authors surmised that the potential may have been too steep near the planar structure. It was also suggested that expansion in polar–nonpolar distances in the dipeptide crystal might indicate limitations in the corresponding parameters. As the authors concluded, "Despite these limitations, the current parameters adequately reproduce the NMA crystal structure and lead to a reasonable reproduction of the L-alanine dipeptide crystal, although areas for improvement are evident." The results also further emphasize the value of crystal simulations. The insight provided by the discrepancies in planarity of the peptide, and polar-nonpolar distances on possible improvements to the FF are simply not available from liquid simulations.

The resulting force field was then exhaustively tested against a variety of crystal systems, three linear tripeptides, six cyclic peptides and three proteins, crambin, bovine pancreatic trypsin inhibitor (BPTI), and carbonmonoxy myoglobin. In general, the results for these test systems gave comparable fits to those systems used in the parametrization and previous systems simulated. As with the other systems problematic areas were pointed out, for example it was noted that in the tripeptide crystal simulations, water molecules undergo large shifts leading to significant changes in nonbonded interaction distances. As the authors stated, "this may indicate limitations in the force field with respect to interactions between water molecules and charged species." The MD simulations of the protein crystal systems resulted in RMS deviations of the order of 1 Å demonstrating that overall the force field gives satisfactory results in accounting for protein systems, though as has been demonstrated by more recent results, much longer trajectories are required to validly assess FFs from deviations from initial experimental structures [274, 275].

11.3.3 Documenting discrepancies for further improvements in the FF

Perhaps as important as providing a reasonable, well characterized force field for protein simulations, was the rigorous and extensive comparison with well characterized experimental systems. The elucidation of a variety of deficiencies in the force field, provide a powerful definition of the need, and lay the groundwork for further improvements in the functional form.

11.4 OPLS

Jorgensen et al. also abandoned the practice of using united atoms for hydrocarbon groups, and upgraded their force field to an explicit all-atom representation (OPLS-AA) [276]. A

thorough reparametrization was carried out using a large set of quantum mechanical and experimental data. As part of the upgrade they derived torsional and non-bonded parameters for a set of typical organic functional groups occurring in proteins. Bond and angle parameters were adopted mainly from the AMBER all-atom FF, with some adopted from CHARMM. A 3 term Fourier expansion was used to represent the torsional energy in the FF expression. The torsion parameters were determined by fitting to rotational energies obtained from QM calculations at the RHF/6-31G*//RHF/6-31G* level for about 50 organic molecules and ions representing functional groups found in proteins. They obtained a fit to within 0.2 kcal/mol for conformational energies. A good fit of structural data was also obtained—the average differences between the QM and OPLS-AA results were 0.01 Å for bond lengths, 2° for bond angles, and 1° for dihedral angles [276].

A 6–12 function was used for the vdW energy and a coulomb interaction between point charges centered on the atoms represented the electrostatic energy. Atoms separated by only 3-bonds had their vdW and coulomb interactions scaled by half. The nonbonded parameters were derived using Monte Carlo simulations to compute thermodynamic and structural properties for 34 pure organic liquids, including heat of vaporization and density as well as radial distribution functions where available. The families simulated included alkanes, alkenes, alcohols, ethers, acetals, thiols, sulfides, disulfides, aldehydes, ketones, and amides. In a further enhancement, the Monte Carlo simulations included sampling of all internal and intermolecular degrees of freedom. The torsion potentials were recalculated with the optimized nonbonded parameters, and if necessary an iterative process was implemented to obtain a consistent set of parameters. The calculated heats of vaporization and densities had average errors of 2% compared with experimental data.

Attempts were made to transfer parameters between similar groups as much as possible, but with some exceptions. An important example is the torsion parameters for ϕ and ψ of amino acids, which were not transferred from the model compound, NMA, but rather derived specifically from the dipeptide *N*-acetyl-*N*′-methylalaninamide. These special parameters enabled a better fit of the relative energies of four low energy conformers of the dipeptide to the corresponding energies obtained from high level LMP2 QM results [276].

11.4.1 OPLS-AA/L: reparametrization of torsion parameters

Due to the limitations that were found on further applications of OPLS-AA, and the increasing need for a reliable FF for biomolecules, an extensive reparametrization of the torsion parameters of OPLS was embarked upon in 2001 [277]. A very large set (more than 2000 points) of high level QM calculations at the level of LMP2/cc-pVTZ (-f)//HF/6-31G** was carried out to determine the relative energies of numerous conformers on the energy surface of the dipeptide analogs of all 20 amino acids. A least squares procedure was used to optimize the torsion parameters to fit the QM conformational energies. All other energy parameters were transferred from the previous OPLS-AA force field. In addition, an attempt was made to change as few of the OPLS-AA torsional Fourier coefficients as possible, and to maximize transferability. The new optimized FF was termed OPLS-AA/L, ("L" representing LMP2).

In the first stage, parameters for φ, ψ were derived from calculations on the alanine dipeptide, and tested on the alanine tetrapeptide. The parameters were then used as a basis for deriving parameters for all other amino acids—φ, ψ parameters were transferred from Ala and parameters for all χ′ s were derived. Nonbonded parameters were transferred from the previous force field with the exception of the sulfur containing amino acids for which new parameters were derived. The validity of these new parameters was tested by reproducing heats of vaporization, densities of liquids and gas-phase dimerization energies [277].

As noted above, the first system fit was the alanine dipeptide. It was immediately obvious that neither the standard OPLS-AA nor the OPLS-AA/L were able to reproduce all six ab initio alanine dipeptide minima. This is similar to the results of MacKerell et al. [271] where in that case the α_L and β_2 conformational energies could not be reproduced among the observed low energy Ala conformers. The same conformers are problematic in this case, though here the deviation in the conformational energy of β_2 6.57 versus the QM energy of 2.75 kcal is worse than the deviation of the α_L energy (3.16 vs. 4.31 kcal).

In order to test the transferability of the new φ, ψ torsion parameters the conformational energies of ten alanine tetrapeptide conformers [278] were calculated with no further parameter optimization. A good fit was obtained—the RMS energy difference between the OPLS-AA/L conformational energies and the QM energies was only 0.56 kcal/mol, significantly lower than the 1.47 kcal/mol deviation obtained with OPLS-AA.

The torsion parameters for all the amino acid side chains, χ, were then re-optimized using the same methodology. Generally, a good fit was obtained with an overall RMS deviation of 0.47 kcal/mol for the dipeptides of the neutral amino acids and 0.94 kcal/mol for the charged AAs (compared with 0.81 and 2.20 for the previous OPLS-AA). As with the Ala dipeptide, though the overall fit was good the exhaustive sampling of the energy surfaces allowed limitations to be characterized. Thus, for example, errors of the order of 2 kcal/mol were found for several conformers of ASN, and GLN, and deviations in the dihedral angles by as much as 20°–30° were observed in a few cases.

11.4.2 Documentation of limitations of simple functional form

Another limitation arose in the treatment of the aliphatic dipeptides, it was found that it was not possible to derive a single set of torsional parameters that would work well for leucine, valine, and isoleucine, and a different type had to be introduced for the latter. However, a good fit was obtained as a result overall. After the refitting of the torsion parameters the RMS error in energy and the dihedral RMS deviation from the ab initio data dropped from 0.43 to 0.27 kcal/mol and from 10.3° to 6.5°, respectively. The latter is not yet as good a fit as one might hope for and in this case, it is tempting to speculate that the inclusion of coupling interactions might improve the transferability.

Overall the reparametrization succeeded in significantly improving OPLS-AA and providing a reasonable force field for protein simulation, while as with MacKerell et al. [271] pointing to limitations in its abilities that can be exploited to further improve the functional form.

A secondary objective of the study was to explore the capabilities and characterize limitations of the simple functional form. The authors point out that the diagonal harmonic, static point charge functional form for the energy has inherent limitations introducing some errors in energetics that can't be accounted for [277]. They cite polarizability specifically although as we have seen interactions not included in this simple form include atomic multipoles and coupling between internals among others [199].

11.5 GROMOS

Towards the turn of the century, the van Gunsteren group started to release new or improved versions of their FF, GROMOS. The overall strategy involved parameterization based primarily on reproducing the thermodynamic properties of pure liquids, free energies of hydration, and solvation in nonpolar systems. Curiously, the united atom model continued to be employed [279, 280], unlike essentially all contemporary force fields, which had determined that the cost in accuracy overrides the savings in computational resources accompanying this major approximation [105, 112, 225, 267, 271, 276, 281].

It was observed that the trans-gauche ratio in alkanes was too large in the prior Gromos FF and an optimization and refitting of the torsion and 1–4 non-bonded parameters was undertaken [282]. This was accomplished by fitting the heat of vaporization, pressure and trans-gauche ratio for liquids of, n-butane, n-pentane and n-hexane. Shortly thereafter it was noticed that the calculated pressure of hexane and larger aliphatic chains was too low. This led to a second reparametrization, this time of the CH_2 and CH_3 van der Waals parameters [280]. In general, good agreement with experiment was obtained for liquid densities, heats of vaporization, and free energy of hydration for a set of linear, branched and cyclic alkanes. Several notable exceptions were pointed out. For example, for ethane, the united atom model used could not be parametrized to reproduce the density and heat of vaporization simultaneously. Also it was found that although it was possible to account for both the pressure and volume of cyclohexane and larger rings, this was not true for the smaller ones [280]. Whether these discrepancies are due to the use of the united atom approximation is yet to be determined.

11.5.1 Documentation of problems with transferability and other deficiencies in functional form

These studies were followed by extending the reparametrization to include polar groups motivated by the observation that free energies of hydration of such compounds was consistently underestimated in Gromos and other contemporary force fields. Only the nonbond parameters were addressed, the valence parameters were left unchanged [283]. The parameters were optimized to reproduce the thermodynamic properties of 28 pure liquids of small polar molecules containing "biomolecular" functional groups and the solvation free enthalpies of 14 amino acid analogs in cyclohexane and water. Perhaps the most noteworthy result was that in spite of many attempts at optimization it was impossible to simultaneously reproduce the properties of the pure liquids, the free enthalpies of solvation in cyclohexane, and the free enthalpies of hydration. A different set of charges from those derived from pure liquids was required to reproduce the experimental free enthalpies of hydration. This was rationalized in terms of the higher degree of polarization in aqueous solutions [283], which was suggested by others too and seems eminently reasonable. These studies provide a valuable benchmark for future testing of force fields that are derived to include atomic polarizability (as well as testing of existing force fields).

Incremental improvements to the force field continued to be made, though the united atom approximation was still invoked. It was noted that short α-helices were found to be less stable than expected. This suggested that the dihedral-angle parameters of the backbone transferred from the earlier version of the force field were no longer appropriate [284]. To address this the torsional angle terms were reparametrized by fitting a large set of crystal structures, and, in addition, the non-bonded interaction between the peptide nitrogen and oxygen was adjusted to be less repulsive [284]. These modifications were tested against the structural properties of four proteins: hen egg-white lysozyme, fox1 RNA binding domain, chorismate mutase and the GCN4-p1 peptide. The results revealed large residual deviations, with RMSDs of 2.4–6 Å, the latter being for the GCN4-p1

peptide, which is only 16 residues long. It is also worthwhile noting that as Beachy et al. [278] point out in discussing validation of force fields by fitting experimental protein structures.

"This procedure is not a reliable test of the ability to rank widely different compact structures in the gas phase or in solution, as it simply examines whether or not the X-ray structure (or a close approximation of it) is a local minimum on the force field surface. And, because of packing considerations and the local stability of secondary structure elements in the gas phase due to hydrogen bonding, it almost always is. The energy of an α-helix or β-sheet conformation does not have to be quantitatively (or even qualitatively) accurate in order to be a local minimum that survives short simulations… There will be countless other local minima, many of which could easily be lower in energy than the native on the force field surface."

Recently another reparametrization was undertaken for oxygen containing compounds including alcohols, ethers, aldehydes, ketones, carboxylic acids, and esters [285]. Again the parameters were optimized against the density and enthalpy of vaporization of the pure liquids and the free energies of solvation in water, ΔG_{wat}, and in cyclohexane, ΔG_{chex}. This time a single force field was derived in an attempt to reproduce the experimental free enthalpies of hydration as well as the properties of the pure liquids and solvation in the apolar solvent cyclohexane. Not surprisingly the resultant force field was a compromise between the results achieved by the two force fields derived previously to fit the aqueous, and apolar systems separately. Rather large deviations from experiment in the free energies of hydration were obtained with several values as high as 35–40% for alcohols, and numerous deviations of 20% occurring in other families. This suggests that the properties of the pure liquids and solvation free energies in cyclohexane may have been over weighted in the optimization. It is however consistent with their previous conclusion (and those of others) that it may not be possible to simultaneously reproduce the properties of the pure liquids, the free enthalpies of solvation in cyclohexane, and the free enthalpies of hydration with a simple quadratic diagonal fixed charge FF.

Yet a further parametrization of the amide backbone parameters was carried out in 2012 [286]. Here only the properties of NMA, including liquid density and enthalpy of vaporization, as well as the Gibbs hydration free energy were used in the parametrization. Additional properties were calculated from the resulting FF including the self-diffusion constant, viscosity, and static relative dielectric permittivity of pure NMA. Results were consistent with previous work. In particular, it was found to be impossible to simultaneously account for hydration free energy and properties of the pure liquid. Thus large errors were observed, for example the free energy of hydration deviated by 30% (−7.0 kcal/mol vs. the experimental value of −10.1 kcal/mol) [286].

11.5.2 Summary: A large set of experimental data for assessing FFs

In summary, a large and powerful set of data involving hydration and solvation has been used in the parametrization and testing of the Gromos force field. The good news and the bad is that the resulting comparison of the calculated results with this experimental data set reveal the inability of the model, including united atoms, and fixed charges to account simultaneously for polar and nonpolar environmental interactions. This, however, is a major contribution in delineating a set of experimental data, which is a comprehensive test of these factors in FFs attempting to account for intermolecular interactions.

11.6 ECEPP-05: Off atom-centered charges and crystal packing prediction

In the early 2000's the Scheraga group undertook a reparametrization of their ECEPP force field [105, 287, 288]. The rigid geometry formulation in which all bonds and valence angles are kept fixed and only torsion angles are variable was retained and thus the FF included only torsional and non-bonded terms. The torsion energy was defined by 3 Fourier terms, while the 6–12 Lennard-Jones potential was replaced in this updated version by the Buckingham (6-exp) potential. The electrostatic energy was defined by the standard coulomb term. They also abandoned the 10–12 H-bond potential and adopted the Hagler–Lifson formulation [86] of the hydrogen bond as being accounted for by the electrostatic and van der Waals interaction with negligible H···O repulsion. The first step involved parametrization of the nonbonded terms—the Buckingham potential and the partial charges for the coulomb term [287, 288]. To this end they introduced a novel methodology, combining quantum mechanical and crystal data, into the refinement of these intermolecular interactions.

The first study involved refinement of the nonbond parameters for aliphatic and aromatic hydrocarbons and alcohols [287] and this was followed by application to amines, imidazoles, amides, and carboxylic acids [288]. In these studies point charges were derived by fitting to the QM electrostatic potential (of monomers), calculated with the 6-31G* basis set using the RESP method [270]. Then an initial set of parameters for the Buckingham nonbond potential were derived from ab initio quantum mechanical interaction energies of dimers calculated at the MP2/6-31G* level. This was followed by a thorough refinement of the latter parameters to reproduce crystal structures (unit cell parameters, positions and orientations of the molecules in

the unit cell) and sublimation enthalpies. In an important addition they required the experimental crystal structures to be the most stable crystal packing mode (global minimum) on the potential surface [287, 289].

The crystal structures of a large number of compounds were employed for the derivation and testing of the potential: For example, 22 amines and imidazoles; 13 amides; and 12 carboxylic acids [288]. Each was separated into training and test sets. For the purpose of derivation of nonbond parameters all molecules were kept rigid. For the most part the experimental crystal structures are well reproduced. Average deviations of the unit cell parameters were less than 4%, and the lattice energies agreed with the experimental sublimation enthalpies within ~2 kcal/mol.

11.6.1 Notable advances introduced in the derivation of the ECEPP nonbond potentials

There are two aspects of this work that are worthy of special note. The first was the finding that, in agreement with the results of Williams [290], electrostatic potentials and crystal structures of the molecules containing nitrogen with sp^2 hybridization (and with no hydrogens bound) cannot be reproduced accurately using only partial charges located on atomic sites. They demonstrated that, as suggested by Williams [290], additional charges located 0.4 Å from the corresponding nitrogen with quasitrigonal symmetry, (intended to model their lone-pair electrons), significantly improved the accuracy of the calculated properties of imidazoles and molecules containing aromatic nitrogen. Although multipole models and other distributed charge models have been developed and used in crystal structure predictions [178, 291] this was the first incorporation of effects such as lone pairs into a "protein force field" (though Krimm's group also introduced atomic dipoles into their polarizable SDFF potential with the aim of developing a general biomolecular FF in the early 2000s) [206, 207]. The importance of better representation of the electron density distribution is further brought home by the comparison of a number of studies of the structure and lattice energy of the acetic acid crystal [288] where the best results were achieved with distributed multipoles (DMA) [292]. This again demonstrates the power of crystal structures for informing potential function derivation.

The second advance in the technology of force field derivations was the introduction of the requirement that the observed crystal structure must be the lowest energy of the crystal packing modes available to the molecule. This was accomplished by exploiting the Conformation-family Monte Carlo method they developed for the prediction of crystal structures of organic molecules [289]. Again, although prediction of packing modes has been an active field of research for some years (see ref [178]), and used, at least implicitly, in the evaluation of force fields, this is the first implementation in the derivation of protein force fields. It provides an additional rigorous test of these potentials, which could be used in future evaluations.

11.6.2 Derivation of ECEPP torsion parameter

The final step in the parametrization was the derivation of the torsion parameters [105]. These were derived by fitting QM (MP2/6-31G**) rotational energy profiles of small molecules representing the side chain functional groups in peptides. In this derivation, the charges were first determined by using the RESP method. The torsional parameters were then obtained by fitting a cosine series to the difference between the QM and MM profiles using these charges and the van der Waals parameters initially transferred from the work described above. The torsion potentials about the backbone angles φ and ψ were determined by fitting blocked Ala, Gly and Pro φ–ψ energy maps (in the case of Pro, φ is fixed in ECEPP due to rigid geometry). Each point on the map was optimized adiabatically at the HF 6-31G** level and then the energy of the optimized structure calculated with MP2 6-31G**. It was found that the backbone nonbond parameters were not transferrable from the previous work, so they were included in the optimization of the φ–ψ maps with *A* and *B*, the 6th power and exponential coefficients, constrained to within 30% of their initial values. The resulting force field was then tested thoroughly by lattice energy minimization, including torsional degrees of freedom, of 29 crystal structures of terminally blocked and unblocked amino acids and di-, tri-, and tetra-peptides.

The authors pointed out that the results indicated some limitations in the rigid geometry model. For example, the constraint of fixed bonds and angles resulted in large deviations in the energies of some conformations of 1, 2-ethanediol, with an overall rms deviation for all ten conformers of 2 kcal/mol. Limitations in transferability were also noted, for example ethanol torsional parameters were found to be inadequate for 2-propanol, and thus separate parameters were derived for the latter, and a similar situation was observed for acetic and propanoic acids. With these caveats, reasonable agreement was found for an extensive set of conformational energies of model compounds. Comparison of the relative energies of blocked alanine minimum energy conformers with the QM values and other force fields yielded mixed results (this is true for all force fields as discussed above). Thus, for example the $C7_{ax}$, and α_L energies deviated by 2–2.5 kcal. However, overall, the performance was comparable to other common diagonal quadratic force fields, with each having problems in the higher energy regions of the φ–ψ map [105]. Most puzzling and inexplicable is the inability of the force field to account for the φ–ψ map of glycine. Not only was it found that the alanine parameters were not transferable to glycine, but even on reparametrization

the rms deviation remained at 5 Kcal/mol [105]. This is a common problem of FFs whose origin remains unexplained.

Finally, a rigorous test of the potential functions was carried out by lattice energy minimization of the 29 peptide crystals described above. The overall crystal structures were reproduced reasonably well with rms deviations in unit cell vectors in the range of 3–5% with the highest deviation, 6.2%, occurring for Val–Gly–Gly. RMS deviations in torsion angles were also reasonable for most crystals though several deviations of ~ 13° were observed. Typical deviations in interatomic distances of ~ 0.5 Å were observed though several instances of deviations of 0.7–0.9 Å were noted. These systems provide an excellent test for protein force fields and should be considered as one of the components of a standard evaluation of these force fields.

12 The ever-increasing trend toward inclusion of quantum mechanical data in force field derivation

As discussed above, as soon as semiempirical molecular orbital methods were introduced, making it possible to treat small molecules researchers including Pullman [114], Hoffman [121], Pople [293], and Scheraga [42] began applying them to small biomolecular model compounds. Exploitation of these methods in FF derivation followed, both in order to determine partial charges [131, 294], and to yield information on torsion potentials [42]. Cox and Williams introduced the use of the electrostatic potentials obtained from HF to derive partial charges [294], based on the work of Scrocco and Tomas [295]. This method, popularized by the Kollman group [296], remains the most widely used approach to date, for determining partial charges. In 1976, we, together with the Leiserowitz lab, combined information from crystal structure determination with Hartree Fock calculations to calculate the rotational potential about φ, and ψ [171], and provide upper bounds for these barriers. Intermolecular forces were also probed early on by HF calculations in a procedure relying solely on this data pursued by Clementi [155, 297].

The recognition of the need to improve force fields toward the beginning of the 1990s spurred an increase in the use of QM in their development [298, 299]. This was facilitated by the exponential increase in computer power following Moore's law [300], and the application of new QM algorithms such as density functional theory (DFT) [301, 302] to "large" organic and biomolecular compounds [303, 304]. The early work included HF calculations to refine H-bond potentials in MM3 [258], calculating φ, and ψ maps to refine these torsions in Amber [267], peptide conformational energies to help optimize Charmm [271], and rotational potentials to optimize these constants in the OPLS-AA force field in 1996 [276].

The early developments marking the utilization of QM in force field derivation cited above, involved the calculation of the energy of the property of interest at several geometries, e.g. intermolecular interaction as a function of distance, or intramolecular energy as a function of torsion angle etc. These energies were then fit to the functional form of the potential to derive the force constants. Subsequent initiatives such as MMFF and CFF discussed below, have continued to build on these early efforts. As MO calculations with increasing large basis sets have been enabled by exponential improvements in hardware QM has become to play a central role in FF development.

12.1 Merck MMFF94

In the mid 1990s Halgren carried out an extensive and thorough parameterization of a force field for a wide range of organic functional groups called MMFF94 [281, 305–308]. Halgren's approach in deriving MMF94 was, in part, patterned after the approach used in developing the CFF93 force field described below. Both, for example, exploited the Quantum Derivative Fitting technique [199, 309–312] using the program PROBE [309, 313, 314] developed in the Biosym potential energy function consortium (PEFC) [315], to derive force constants, from the information contained in a quantum mechanical surface. However, the two approaches also differ significantly in a number of ways. Perhaps the most significant difference is seen in the functional form adopted by Halgren (Eq. 20).

$$E = E_b + E_\theta + E_\phi + E_\chi + E_{Coupling} + E_{Non\text{-}bonded} \tag{20}$$

$$E_b = \sum_b {}^2K_b[(b-b_0)^2 - 2(b-b_0)^3 + 7/3(b-b_0)^4] \tag{20a}$$

$$E_\theta = \sum_\theta {}^2K_\theta[(\theta-\theta_0)^2 - 0.007[(\theta-\theta_0)^3] \tag{20b}$$

$$E_\phi = \sum_\phi {}^1K_\phi(1+\cos\phi) + {}^2K_\phi(1-\cos 2\phi) + {}^3K_\phi(1+\cos 3\phi) \tag{20c}$$

$$E_\chi = \sum_\chi K_\chi \chi^2 \tag{20d}$$

$$E_{Coupling} = E_{b\theta} = \sum_{b\theta'} K_{b\theta}(b-b_0)(\theta-\theta_0) \tag{20e}$$

$$E_{Non\text{-}bonded} = \sum_{i<j} E_{vdW}(r_{ij}) + \sum_{i<j} E_{Coul}(r_{ij}) \tag{20f}$$

$$E_{vdW}(r_{ij}) = \varepsilon_{ij}\left(\frac{1.07r^*_{ij}}{r_{ij}+0.07r^*_{ij}}\right)^7\left(\frac{1.12r^{*7}_{ij}}{r^7{}_{ij}+0.12r^{*7}_{ij}} - 2\right) \tag{20g}$$

$$E_{Coul}(r_{ij}) = \frac{q_i q_j}{D(r_{ij} + 0.05)^n} \tag{20h}$$

$$r^*_{ij} = \frac{r^{*3}_{ii} + r^{*3}_{jj}}{r^{*2}_{ii} + r^{*2}_{jj}}; \quad \varepsilon_{ij} = \frac{181.16 G_i G_j \alpha_i \alpha_j}{(\alpha_i/N_i)^{1/2} + (\alpha_j/N_j)^{1/2}} \frac{1}{r^{*6}_{ij}} \tag{20i}$$

where E_b and E_θ are the strain energies associated with bond (b) and valence angle (θ) distortions from their reference values b_0 and θ_0; (φ) is the dihedral angle and $E\varphi$ is the torsion energy; E_χ is the strain energy associated with deformation of the plane described by the oop angle χ; $E_{coupling}$ is the bond–angle coupling energy; r^*, and ε are the vdW radius and minimum energy (see Fig. 2); q is the partial charge; D is the dielectric; and N and α are the effective number of electrons and polarizability.

Anharmonicity is included through quartic bond stretching and cubic angle bending terms. In contrast to CFF93 (Eq. 21), and MM4 (Eq. 16) that included a broad range of coupling terms, MMFF94 included only the coupling between bond stretching and angle bending. Halgren [307] also pointed out that this results in some molecular behavior that could not be accounted for by MMFF94. An example is the coupling between the torsion and C–N bond length in amides, and other similar delocalized bonds which significantly effects the torsion energy. These require, and can easily be accounted for, by a bond stretching-torsion coupling term as present in both CFF93 and MM4. Another significant difference between the MMFF94 formulation and essentially all current force fields is the functional form used for the nonbond interaction. As seen in Eq. (20g) Halgren introduces a "buffered 14–7" functionality as opposed to the common 12–6 or 9–6 Lennard-Jones potential function. The buffering is invoked to keep the energy finite as the interatomic distance approaches zero [281, 316]. This function was derived by fitting to rare gas non bond potentials taken from the literature which were in turn derived from a variety of experimental data and high level quantum calculations [316]. Following the results of Maple et al. [309] the out of plane deformation was represented by the Wilson definition (Fig. 10) Electrostatics were accounted for by the usual coulomb law, though like the Lennard-Jones it is "buffered" to ensure that it remains finite as the distance approaches zero. Charges are obtained via a bond increment scheme [317] as in CFF93 (Eq. 21p), and coulomb interactions between atoms separated by three bonds (1–4 interactions) are scaled by 0.75.

As in the case of the CFF93, MMFF94 is derived by fitting properties calculated from quantum mechanics. As noted below, quantum derivative fitting [199, 309–312] was used to derive the harmonic bond stretching and angle bending terms. At the time this FF was developed, advances in computational capability allowed more rigorous QM basis sets to be invoked. Thus, Halgren supplemented the HF/6-31G* information by properties obtained from extensive higher level calculations, including − 360 equilibrium conformers derived from full geometry optimization at the MP2/6-31G* level, and energies for ~ 380 MP2/6-31G* optimized geometries carried out at the MP2/TZP level using a triple-zeta plus polarization basis set, and at the MP2 and MP4SDQ levels using modified 6-31G* basis sets. In addition, single-point MP2/TZP calculations carried out at ~ 1450 torsionally incremented geometries from MP2/6-31G* structures were included to sample conformational space. Finally, large basis set calculations of intermolecular interactions in nonpolar systems obtained using highly correlated wave functions were carried out. These cover an impressive range of organic families including: alkanes, alkenes, alcohols, phenols, ethers, aldehydes, ketones, ketals, acetals, hemiketals, hemiacetals, amines, amides, peptide analogs, ureas, imides, carboxylic acids, esters, carboxylate anions, ammonium cations, thiols, mercaptans, disulfides, halides (chlorides and fluorides), imines, iminium cations, amine *N*-oxides, hydroxylamines, hydroxamic acids, amidines, guanidines, amidinium cations, guanidinium cations, imadazolium cations, aromatic hydrocarbons, and heteroaromatic compounds [281].

12.1.1 A novel strategy for FF optimization

Halgren also introduced a novel strategy in his derivation of the force field. Previous force fields had all been derived using a "functional group" approach where, for example, parameters for hydrocarbons are derived by fitting a set of alkanes and alkenes, and these parameters are then used in the derivation of subsequent families such as the amides etc. Halgren instead derived all parameters from the full set of extensive computational data on all the families simultaneously. It was felt that this would reduce possible hidden correlations in the parameters not accounted for in the fit to a single family. However, this straightforward approach was computationally impractical at the time, so a modified protocol was adopted involving a nested optimization of different classes of parameters (e.g. harmonic valence force field constants, nonbonded parameters etc.). An additional caveat with this approach is that if a problem is found in any particular aspect of the force field in a particular functional group the entire force field for all families would need to be rederived.

The first step in the procedure was derivation of the nonbond van der Waals radius, r^*, and well depth, ε for the 14–7 Lennard-Jones potential [281, 305]. r^* was derived directly from the atomic polarizability, α_I, while ε was defined by the Slater–Kirkwood equation (Eq. 6c) n terms of the polarizability α_i and the number of effective electrons, N_i. [305, 316]

The nonbond parameters for carbon were taken from previous work [316] and other heavy atom types are derived from compiled values of α_i and N_i [316]. The ε and r^* parameters for aliphatic hydrogens were then determined by fitting to high quality ab initio data on intermolecular interactions for the methane and hydrogen dimers. As noted below Halgren recognized the inadequacies of the commonly used combination rules for derivation of nonbonded interactions between unlike atoms. Exploiting rare gas data, he derived a new set of combining rules based on atomic polarizabilities and number of effective electrons (Eq. 20i) [316]. The bond charge increments [317], were first obtained from fits to HF/6-31G* dipole moments using the PROBE program from the Biosym consortium [309, 313, 315] (HF/6-31G* charges are known to be ~10–20% too high, which was felt to be appropriate for condensed phase calculations). The vdW and bond-charge-increment parameters were then scaled to best reproduce HF/6-31G* interaction energies and hydrogen-bond geometries [281, 305].

The next step in the procedure was optimization of the reference bond lengths and angles, which were determined from HF/6-31G* geometries [281, 306]. These were then used in the determination of the quadratic force constants for bond stretching, angle bending, stretch–bend interaction, and out-of-plane bending [281, 306], determined by quantum derivative fitting of HF/6-31G* first and second derivatives [199, 309]. The coefficients for the cubic and quartic bond stretching term were chosen to fit a Morse Function with an exponent of 2, and the cubic angle bending constant was obtained by considering the effect on the angle-bending force constants from changes in bond angles, in HF/6-31G* calculations on water, ammonia, and ethylene [306]. (We note that in general this will also contain contributions from bond–angle coupling which is included in MMFF). The force constants obtained in this step were then scaled by fitting experimental vibrational frequencies, and the reference values refined by fitting MP2/6-31G* values. These values of the quadratic force constants were then used to update the cubic and quartic force constants for the next PROBE fit. (In general, three to five such iterations produced a converged set of quadratic force constants) [306]. The final step was the derivation of the torsion potentials carried out in the context of the previously determined nonbond and valence potential parameters [281, 307]. These were obtained by fitting relative conformational energies and geometries, determined either from composite "MP4SDQ/TZP" calculations at MP2/6-31G* optimized geometries (380 conformers) or from single-point MP2/TZP calculations carried out at 1450 torsionally incremented geometries derived from MP2/6-31G*-optimized geometries [281, 307]. As noted above, each of these steps was then iterated until a consistent set of parameters was obtained.

In addition to fitting the QM data Halgren carried out a thorough comparison of the results of MMFF94 with a wide range of experimental geometries, vibrational frequencies, and conformational energies [281, 305–307, 318]. Aside from the fact that different parameters are required for small rings an excellent overall fit is achieved to both the quantum and experimental data. Fits to within 0.014 Å, 1.2° and 61 cm^{-1} (RMS) for bond lengths; bond angles and vibrational frequencies were obtained. Dipole moments were reproduced to 0.39D and 5.5°, conformational energies to an RMS of 0.39 kcal/mol, and torsional energies to 0.5K kcal/mol. Given the energy ranges spanned 3.88 and 4.37 kcal/mol respectively, the latter corresponds to deviations of ~10%.

12.1.2 Problem with nonbond parameters?

One of the potential weaknesses of the MMFF94 approach is the omission of any condensed phase data, either in the parameterization or testing. Recall that the nonbond parameters are obtained from atomic polarizabilities and the Slater Kirkwood equation as described above. Bordner et al. [319] report a calculation of sublimation energies of 14 alkanes and 11 oxygen containing compounds using both a nonbond potential they derived and the MMFF94 van der Waals potential. Surprisingly the sublimation energies calculated with MMFF94 were found to be systematically 30–40% too low. This serves to further emphasize Lifson's "consistent force field" approach and the importance of including a wide range of experimental data, sampling both intramolecular properties and intermolecular interactions, including condensed phase, and dimer interactions in the derivation or testing of force fields.

12.2 CFF93 and quantum derivative fitting (QDF), a novel paradigm for force field development

In the 1980s we recognized that a much larger source of information than that contained in the QM energies was readily accessible in the QM Cartesian derivatives of the energy [309, 320]. The latter provides an abundance of quantum data describing all aspects of the energy surface and informs us not only of the value of force constants but also provides information on the deficiencies in the analytical representation of the FF (functional form), as well as transferability of the force constants.

This concept gave rise to a proposal to develop a FF for molecular simulations of biomolecules involving a new approach to the derivation of FFs, and spawned a major program supported by a Consortium [315], composed of ~30 Pharma, chemical, and hardware companies, at Biosym Technologies. The goal of the project was to determine and contrast the functional forms required to faithfully recapitulate the energy of intramolecular deformations and nonbonded interactions; to provide an objective method to rapidly determine the force constants in the functional form

with a protocol that lends itself to automation; and finally to determine the transferability of these force constants [309].

The general form of the energy function to be investigated is given in Eq. (21).

$$E = E_b + E_\theta + E_\phi + E_{Coupling} + E_{Non\text{-}bonded} \tag{21}$$

$$E_b = \sum_b {}^2K_b(b-b_0)^2 + {}^3K_b(b-b_0)^3 + {}^4K_b(b-b_0)^4 \tag{21a}$$

$$E_\theta = \sum_\theta {}^2K_\theta(\theta-\theta_0)^2 + {}^3K_\theta[(\theta-\theta_0)^3 + {}^4K_\theta[(\theta-\theta_0)^4 \tag{21b}$$

$$E_\phi = \sum_\phi {}^1K_\phi(1-\cos\phi) + {}^2K_\phi(1-\cos 2\phi) + {}^3K_\phi(1-\cos 3\phi) \tag{21c}$$

$$E_{Coupling} = E_{bb'} + E_{\theta\theta'} + E_{b\theta} + E_{b\phi} + E_{b'\phi} + E_{\phi\theta} + E_{\theta\phi\theta'} \tag{21d}$$

$$E_{bb'} = \sum_{bb'} K_{bb'}(b-b_0)(b'-b'_0) \tag{21e}$$

$$E_{\theta\theta'} = \sum_{\theta\theta'} K_{\theta\theta'}(\theta-\theta_0)(\theta'-\theta'_0) \tag{21f}$$

$$E_{b\theta} = \sum_{b\theta'} K_{b\theta}(b-b_0)(\theta-\theta_0) \tag{21g}$$

$$E_{b\phi} = \sum_{b\phi'} K_{b\phi}(b-b_0)({}^1K_{b\phi}\cos\phi + {}^2K_{b\phi}\cos 2\phi + {}^3K_{b\phi}\cos 3\phi) \tag{21h}$$

$$E_{b'\phi} = \sum_{b\phi'} K_{b'\phi}(b'-b'_0)({}^1K_{b'\phi}\cos\phi + {}^2K_{b'\phi}\cos 2\phi + {}^3K_{b'\phi}\cos 3\phi) \tag{21i}$$

$$E_{\theta\phi} = \sum_{\theta\phi} K_{\theta\phi}(\theta-\theta_0)({}^1K_{\theta\phi}\cos\phi + {}^2K_{\theta\phi}\cos 2\phi + {}^3K_{\theta\phi}\cos 3\phi) \tag{21j}$$

$$E_{\theta\phi\theta'} = \sum_{\theta\phi\theta'} K_{\theta\phi\theta'}(\theta-\theta_0)(\theta'-\theta'_0)cos\phi \tag{21k}$$

$$E_{Non\text{-}bonded} = \sum_{i<j} E_{vdW}(r_{ij}) + \sum_{i<j} E_{Coul}(r_{ij}) \tag{21l}$$

$$E_{vdW}(r_{ij}) = \varepsilon_{ij}\left[2\left(\frac{r^*_{ij}}{r_{ij}}\right)^9 - 3\left(\frac{r^*_{ij}}{r_{ij}}\right)^6\right] \tag{21m}$$

$$E_{Coul}(r_{ij}) = \frac{q_i q_j}{r_{ij}} \tag{21n}$$

$$r^*_{ij} = [1/2(r^*_{ii} + r^*_{jj})]^{1/6}; \quad \varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}\frac{2(r^*_{ii})^3(r^*_{jj})^3}{(r^*_{ii})^6 + (r^*_{jj})^6} \tag{21o}$$

$$q_i = q_{i0} + \sum_j \delta_{ij} \tag{21p}$$

where E_b and E_θ are the strain energies associated with bond (b) and valence angle (θ) distortions from their reference values b_0 and θ_0; $E\varphi$ is the dihedral angle (φ) energy. $E_{coupling}$ is the coupling energy between valence coordinates b, θ and φ; r^*, ε, and q are the vdW radius and well depth (see Fig. 2); q is the partial atomic charge; and δ is the bond increment [317].

The approach differed from methods existing at the time in two main respects. The first was that most force fields were (and are) determined by fitting a combination of quantum energies and experimental data. Often some of the force constants are derived from experiment while others are obtained from QM *relative energies*, sometimes further adjusted to account for experimental data. The method has been labelled quantum derivative fitting (QDF) [199]. In this approach, rather than fitting the QM energies as a function of configuration, we fit the first and second derivatives of the energy with respect to the coordinates, as well as the energy, of a set of distorted configurations (~15–20) of a variety of molecules (10–20) containing the functional group of interest. For a molecule containing n atoms there are 3n first derivatives and 3n (3n + 1)/2s derivatives, for each configuration, characterizing the slope and curvature of the energy surfaces respectively. Note the first derivatives are essentially the force while the (appropriately mass weighted) second derivatives yield the vibrational frequencies. Thus, from a QM calculation of a single configuration of a molecule we obtain 4.5 ($n^2 + n$) "quantum observables" describing the energy surface, rather than a single observable, energy [309, 312, 320]. The force field derived by fitting this surface is then scaled *in toto* to fit experimental quantities [321], resulting in a "self-consistent" balanced FF, as opposed to using both experimental and QM simultaneously, which can result in a compromised FF which fits neither optimally. In addition, the requirement on the QM energy surface is less stringent—rather than requiring that the QM energy surface is quantitatively accurate as necessary when used directly, we only require that the relative energy over the surface is correct, i.e. the overall shape of the QM surface mirrors the experimental surface. The scaling can then bring them into confluence.

The second advantage of the method is that it readily lends itself to least squares methods, and automation. Thus, the fit to the parameters reduces to minimizing the sum of

square deviations with respect to the parameters as given in Eq. (22);

$$S = \sum_{\alpha} [W_E(E_{ff} - E_{qm})^2 + \sum_{i} W_D(\partial E_{ff}/\partial x_i - \partial E_{qm}/\partial x_i)^2 + \sum_{i,j} W_H(\partial^2 E_{ff}/\partial x_i \partial x_j - \partial^2 E_{qm}/\partial x_i \partial x_j)^2] \quad (22)$$

where S is the sum of squared deviations to be minimized, E_{ff} is the (relative) energy of the configuration calculated with the FF, $\partial E_{ff}/\partial x_i$ is its derivative with respect to coordinate x_i, and $\partial^2 E_{ff}/\partial x_i\ \partial x_j$, its second derivatives. E_{qm}, $\partial E_{qm}/\partial x_i$, and $\partial^2 E_{qm}/\partial x_i\ \partial x_j$ are the corresponding quantities calculated by QM, and appropriate weighting factors of each sum are given by the corresponding weight "W". Finally, each sum goes over all distorted structures, α, of all model compounds of the family or families included in the training set. If W_E, W_H, are set to zero Eq. (22) reduces to a fit of the forces sometimes referred to as "force matching [322]," while If W_E, and W_D are set to zero, the result reduces to the Hessian optimization technique [323].

12.2.1 Gauging the optimal functional form, importance of coupling terms, anharmonicity and out of plane coordinates by QDF

In 1988 Maple, Dinur and Hagler published a study demonstrating the method [309], in which QDF was applied to the formate anion (HCOO–). Optimal potential constants were determined by fitting first and second QM derivatives for various functional forms, and the resulting accuracy of these force fields, such as those containing anharmonicity, coupling terms, and alternative representations of the out-of-plane deformation of a trigonal moiety, was evaluated [309]. As input data the 1st and 2nd energy derivatives of eight distorted structures of the formate ion were calculated at the 4-31G* level. These eight single point calculations yielded 720 Cartesian 1st and 2nd derivatives (and seven relative energies).The potential parameters for six force fields, subsets of the generalized anharmonic, coupled FF given in Eq. (21), were optimized by least squares, and the resulting FFs assessed for their ability to account for the quantum energy surface [309].

The first, FF1, was a simple harmonic diagonal force field. The second, FF2, contained anharmonic cubic terms for all coordinates but the out-of-plane, which has only even terms by symmetry. The third force field, FF3, included quartic anharmonic contributions, while FFs 4–6 introduced coupling between internals. FF4 is a quadratic FF with quadratic cross terms enhancing the usual quadratic diagonal representation, FF1. Likewise, FF5 and FF6 included cubic coupling terms added to the FF2, and FF3 force fields respectively. The consequences of the various level of approximation are given in Table 9.

For the purposes of this study, only the energy derivatives of the eight distorted configurations were used in the optimization of the potential constants. Relative energies, equilibrium geometries, and the frequencies were then used to test how well the optimized force fields could account for properties outside the training set.

12.2.2 Coupling terms contribute significantly to strain energy

Several observations were immediately clear from the results in Table 9. Both anharmonicity and coupling between internals play an important role in the energetics of molecular

Table 9 Formate ion calculations

Force field	Distorted configuration		Equilibrium geometry		
	$\sqrt{(s/s_0)}$	ΔE_{rms}	Δr_{rms}	$\Delta\theta_{rms}$	$\lvert\Delta\nu\rvert$, avg
Diagonal					
1. Quadratic (9)	0.298	2.18	0.17	0.27	155.6
2. Cubic (13)	0.137	70.1	0.010	0.41	186.6
3. Quartic (18)	0.125	2.05	0.009	0.62	172.3
With coupling terms					
4. Second order (14)	0.278	1.00	0.010	0.15	88.3
5. Third order (35)	0.032	0.13	0.002	0.42	31.1
6. Third + quartic diagonals (40)	0.017	0.17	0.001	0.16	7.2
FF6, different definitions of OOP					
Pyramid height	0.017	0.17	0.001	0.16	7.2
Improper torsion	0.032	0.53	0.001	0.45	13.4
Wilson et al.	0.018	0.21	0.000	0.11	11.8

Energy is in kcal/mol, distances are in A, angles are in degrees, and the frequencies, ν, are in cm^{-1}. Numbers of parameters are given in parentheses. OOP is the out-of-plane coordinate.

mechanics and dynamics. In fact, a FF with third order coupling terms and quartic diagonals gives almost indistinguishable structures, energies and dynamics (frequencies) to those obtained with quantum mechanics. It should be noted that this was not a consequence of over parametrization as a total of 720 1st and 2nd derivatives were used in the fit of a maximum of 40 parameters. (In fact, it was observed that this could provide an inexpensive method to carry out QM structural searches—fit several distorted configurations with a coupled anharmonic FF and carry out calculations on this surface. This is equivalent to carrying out the QM calculations, and structures minimized on this surface were virtually indistinguishable from the QM minimum energy structures.) The other important observation was that some interaction terms contribute more to the energy surface than others. For example, the coupled FF4 contained 14 parameters compared to the anharmonic diagonal FF3 which had 18 parameters, yet the former performed significantly better including decreasing strain energy deviations by over 50% (1.0 vs. 2.18 kcal/mol), and vibrational frequencies by almost the same (88 vs. 172 cm^{-1}).

The importance of coupling terms in accounting for structural and energetic properties is not unique to these studies and is now well documented, as reflected in several examples. Similar results were found by the authors for all families addressed, including the amides and peptides, the basic building blocks of proteins [324]. Furthermore, as noted above Allinger found it necessary to include a set of cross terms in MM3 [111, 247], accounting for coupling between bond, angles and torsions to correct deficiencies found in applications of MM2. Another example comes from the work of Palmo et al. [325] who found "angle–torsion cross terms also to be "fairly important", affecting torsion angles by several degrees and carrying a potential energy of up to 0.5 kcal/mol in the case of a distorted hexane structure." To put this in context, the oft-stated ultimate goal for accuracy in drug discovery/ligand binding simulations is 0.5 kcal/mol [326], the energy carried by this single coordinate. Ermer has also addressed these issues with similar conclusions [249, 327].

12.2.3 Out-of-plane representation, a simple but consequential correction

The classic representation for the out of plane deformation of a trigonal center from the plane of its three bonded atoms, such as the peptide nitrogen, out of the plane of its bonded C^{α}, C′, H(N) atoms in protein FFs is the "improper torsion" angle (Fig. 10, left). There is no rationale for this representation beyond the fact that the torsion functionality is already implemented in MM codes. In fact, it is curious that it was ever used, and has persisted, since a more appropriate coordinate, the angle between a bond and the plane formed by the other two bonds (depicted for a carboxyl group in Fig. 10, center) was known since the 1950s [328]. We proposed, what seems perhaps the most logical coordinate for this out of plane motion, simply the distance of displacement of the central atom, e.g. the carboxyl carbon, out of the plane of its bonded atoms e.g. the pyramid height as seen in the rightmost panel of Fig. 10 [309].

The QDF technology was exploited to test the appropriateness of these coordinates and the results are given in the last 3 rows of Table 9 (FF6 incorporates the pyramid height so this line appears twice). Though not huge, the fit to the energy surface is significantly better with either the pyramid height or Wilson Decius and Cross coordinates for the out of plane deformation than with the improper torsion. It would seem evident that the improper torsion coordinate should be discarded in all FFs in favor of either of the alternatives.

12.3 Derivation of a general consistent ff for biopolymers, organic compounds and polymers (CFF93)

The QDF method was then applied to derive the CFF93 series of force fields by fitting compounds containing functional groups from biomolecules, including proteins, and polymers [314, 321, 324, 329]. Distorted conformations of a large, diverse set of compounds were invoked to describe the energy surface of the array of functional groups. The analytical energy surface (FF) was then fit to the QM derivatives and energies of these families of compounds as described above (including highly strained members of the families). The force field included quartic bond stretching and angle bending terms and an extensive set of coupling terms (Eq. 21). It was first applied to alkanes where the properties of distorted configurations of 16 normal, cyclic and strained compounds were fit [310]. The use of QDF on this set of compounds resulted in 128,376 "quantum observables," (relative energies, 1st and 2nd derivatives), which were used to determine 66 force constants (Eq. 21). Both diagonal harmonic and the fully coupled anharmonic functional forms were considered and it was found that the alkane energy surface was anharmonic and significant intramolecular coupling interactions were reflected in its topography.

The resulting force field was then tested by calculating structural, energetic and vibrational properties of molecules used in the optimization as well as properties of 12 additional compounds outside the training set [273]. In particular the transferability of the parameters was scrutinized, as well as the previous conclusion of the importance of anharmonicity and coupling interactions. The results, as displayed in Table 10 confirmed the conclusions of the previous study, and demonstrated that transferability of parameters was superior to those used in diagonal harmonic FFs, in large part because additional ("effective") parameters are required

Table 10 RMS deviations of analytical FF properties from quantum mechanics properties

Property	QMFF (anharmonic, coupled)	Harmonic FF
Bonds (Å)	0.003	0.006
Angles (°)	0.4	0.8
Torsions (°)	1.2	2.3
Energies (kcal/mol)	1.0	**3.3**
Frequencies (cm^{-1})	34	100

to account for the missing coupling interactions and anharmonicity [273] as discussed below.

As in the case of the formate ion, the results show that the hydrocarbon energy surface reflects considerable anharmonicity and coupling. Significantly, these results verify the conclusions drawn from the formate ion that contrary to existing dogma, not only does coupling contribute to dynamics, but it makes significant differences in strain energies as well.

12.3.1 Transferability

As noted, the anharmonic, coupled functional form of the energy surface gives rise to a more transferable set of force constants than that obtained typically in harmonic diagonal force fields. An example is the adequacy of a single torsion constant to describe rotation about the sp^3 C–C bond irrespective of whether it's primary, secondary, or tertiary, as opposed to the multiple parameters frequently needed in the diagonal harmonic FFs. This behavior differs from even MM4 [330] and MMFF94 [281] as well as essentially all other "protein" FFs which require different constants for rotation about HCCH, HCCC, and CCCC bonds,. More significantly the same energy surface, without parameter modification, also accounts for the energetics of three and four membered rings. It is the only functional form currently able to account for both these strained rings and normal hydrocarbons simultaneously [199].

The ability to treat small cyclic moieties is of more than just academic interest for several reasons. First of all, the need for different force fields for small rings implies the energy surface is discontinuous, with singularities between unstrained compounds' and four membered ring geometries, and again between 4 and 3 membered ring geometries—intuitively unappealing.

That is, at some angle on the path from the bond angle in an acyclic hydrocarbon to that in cyclobutane K_θ changes discontinuously from its acyclic value to its four-membered ring value in FFs that invoke different parameters. Furthermore, the ability to treat highly strained three and four membered rings gives us confidence that the FF is competent to reproduce strains encountered during dynamic trajectories, and ligand binding.

Finally many biologically interesting compounds and drugs contain both 3 and 4 member heterocyclic and fused rings, including the common antibiotics ciprofloxacin and the β-lactams such as penicillins and cephalosporins [199]. Numerous examples can also be found in the standard Zinc [331] and MDDR [332] databases used in virtual screening for drug discovery. Clearly these small rings would present significant challenges to a FF using different analytical forms for small rings. It is difficult to imagine how fused ring systems would be treated, if different forms were used. For example, would internal which are shared between a fused 4 and 5 or 6—membered ring be treated with the standard force constants or the special ones for the 4-membered ring?

12.3.2 Scaling the QM energy surface to account for experimental properties: amplification of experimental data

Following the derivation of the hydrocarbon CFF from the QM derivatives and energies by QDF, it was shown that an "experimental" force field could be obtained from the "QM" FF by simply scaling it to fit experimental data [321]. Results indicated that this could be achieved by using only a single scale factor for each type of internal. Thus, only five scale factors and two reference values were needed to fit the experimental data (Table 11). The small number of parameters allowed them to be determined from a small subset of experimental observables, for this exercise only 212 experimental frequencies and bond lengths were employed. The resulting force field was then tested by comparing the calculated value of a wide range of structural, conformational and rotational energies and frequencies, including those of small cyclic rings, not included in the fit of the scaling factors. It was found that the calculated values accounted for

Table 11 Seven scale constants and reference values for the hydrocarbon force field

Scaling constants	
Bond stretching	
S_b (C–C)	0.88
S_b (C–H)	0.83
Angle bending—S_θ	0.81
Torsion—Sφ	0.84
Cross-terms—S_{xt}	0.87
Reference values	
b_0 (C–C)	1.535
b_0 (C–H)	1.111

the experimental values, as well or better than existing force fields including MM3 [111, 199, 273].

The scaling described above basically requires that the ab initio energy surface recapitulates the relative topography of the actual energy surface. This is, in fact, a far less stringent requirement of the QM than implicit in its standard use in FF derivations such as those cited above. In those applications, the QM energies are used directly thus requiring that they reproduce absolute values of experimental data, while by scaling an ab initio FF all that is required is that surface is proportional to the actual surface. This is confirmed by the fit achieved [321].

12.3.3 A blueprint for (automated) force field development

The QDF methodology lends itself to automation as well as providing a blueprint for FF derivation. A workflow is straightforwardly constructed in which model compounds are input and distorted configurations generated. These then serve as input to a quantum code, and relative energies, gradients and Hessians generated. These then serve as the input to a least squares procedure where the force constants are optimized to fit the corresponding quantum data. This FF could then be used directly, if high enough level QM data is used, greatly expanding the chemical species that can be treated by MD, or other classical simulation methods, to any functional group that can be treated by QM. It is very likely to be superior to any method of "automatic parameter assignments" used now when parameters are missing in a simulation. Finally, the FF can be scaled, leveraging a small pool of experimental data, to derive an "experimental" FF.

Over the late 1980s into the early 2000s the QDF method was applied to an extensive set of chemical families including amides and peptides [324], carbohydrates [329], heterocycles [314], aromatic sulfones, sulfonamides, thiols, sulfides, and another 25 organic and biological functional groups [314]. In addition it was further extended to a variety of polymer and materials functional groups such as polycarbonates [333], siloxanes [334], silica and zeolite catalysts [335], perchlorates [336] and a variety of others [337]. In the latter studies, where the FF was renamed "COMPASS", liquid data was included to optimize the nonbond parameters, which as discussed elsewhere in this review is a valuable test of the FF, though unfortunately the more powerful information from crystal structures was abandoned [338]. Clearly, the optimal and rigorous protocol is to use both, as per Lifson's paradigm [49, 50].

12.3.4 In conclusion: a widely applicable and transferable FF for calculation of structural, energetic and dynamic properties of biomolecular, organic and synthetic polymers from QDF

The CFF93 force field became close to all encompassing, covering a wider range of chemical species than other existing FFs. Though the intramolecular form has been thoroughly optimized and accounts for conformational energies, rotational barriers, strain energies and dynamics (frequencies) extremely well, in large part due to its ability to account for coupling and its use of a well-founded out of plane functional representation, as with other contemporary FFs the nonbonded interactions need to be further elaborated to include enhanced description of electronic distribution, charge flux, polarizability, nonbond anisotropy, etc [199, 339–342].

12.4 CFF nonbond parameters: 136 crystal structures and 34 sublimation energies

For the reasons described above, crystal structure and energetics provide a powerful, and stringent criterion for assessing force fields. A question that might be asked of force fields proposed for simulating proteins is "before we use these FFs for protein simulations can they account for the structure and energetics of simple crystals such as acetic acid and oxalic acid polymorphs, *N*-methylacetamide, imidazole, or trithiane?" If these simple homogeneous condensed systems present problems one can't expect the FF to be able to yield better results for more complex protein systems containing the same interatomic interactions, among others. Thus in CFF93 the nonbonded force field parameters consisting of atomic partial charges and van der Waals parameters were derived and tested by predominantly fitting crystallographic experimental data as well as dipole moments for a broad set of organic compounds [317]. We exploited the lattice sum method used in the original study of amides invoking the Newton–Raphson approximation to predict deviations in the lattice constants [85, 87], because of its efficiency. Bond increments were introduced in order to account for the dependency of atomic charge on the chemical nature of bonded neighbors [317]. These are analogous to bond dipoles, but have no distance dependence. Thus, the atomic charge on an atom i, q_i, was expressed in terms of bond increments, δ_{ij}, which are the incremental partial charges contributed to atom i by all other atoms to which it is bonded (Eq. 21p). The q_{i0} is a formal charge for atom i. The δ_{ij} are constrained such that $\delta_{ij} = -\delta_{ji}$. Thus, parametrization of the atomic partial charges using bond increments has the additional advantage that $\Sigma q_i = \Sigma q_{i0} = 0$ so that for uncharged molecules the sum of the partial charges derived from bond increments will always be zero [317].

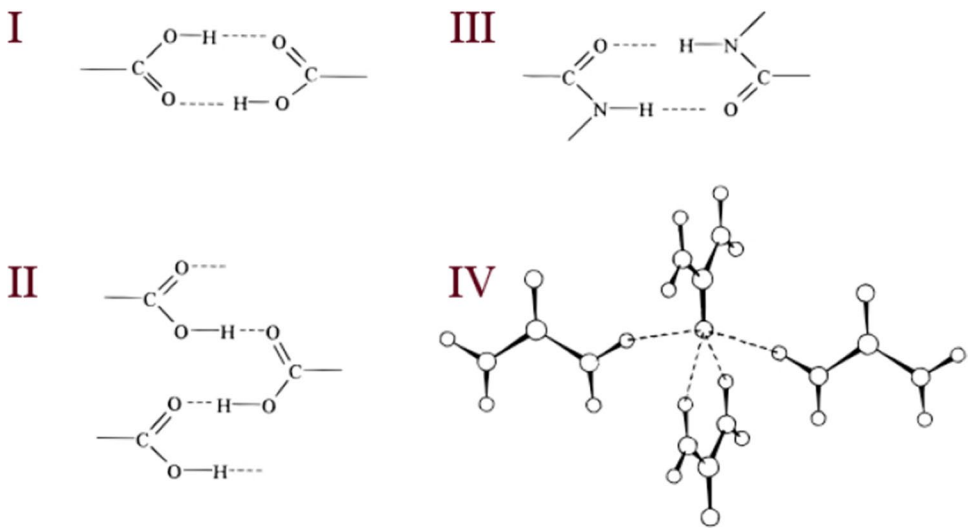


Fig. 11 Hydrogen bond schemes in crystals of acids (I, II) and amides (III, IV)

The compounds in the fit spanned 11 functional groups: alcohols, aldehydes, amides, amines, carboxylic acids, esters, ethers, *N*-heterocycles, hydrocarbons, ketones, and sulfur compounds. The data consisted of the lattice constants of 136 crystal structures, 34 sublimation energies, and 63 gas-phase dipole moments. (Since polarization was not included, this yields a compromise between the gas phase charges and the polarized charge distribution in the crystals, though the latter will dominate due to the preponderance of crystal data). Only 11 atom types and 26 bond dipoles were needed to fit this large data set corresponding to an observable to parameter ratio of 913:48 or almost 20 to 1. As a test of the validity of the derived FF each crystal structure was then minimized with respect to the unit cell parameters and relative position of the molecules in the unit cells.

As crystals display a variety of packing modes, intermolecular orientations, several molecules in the asymmetric unit, and often different polymorphs for the same compound [172–174, 343], many different relative intermolecular orientations are encountered (even within a single crystal). This diverse structural information is exemplified by the different hydrogen bond motifs sampled in the crystals of amides and acids as given in Fig. 11. Of course, the geometry of these motifs is further varied by the packing constraints imposed upon them by the functional groups to which they are bonded.

The quality of the fit to the data described above, as measured by comparing computed and experimental gas-phase dipole moments, lattice constants and energies is summarized in Table 12. A reasonable fit to experiment was obtained

Table 12 Fit to experimental values (RMS deviations)

Family	Crystal structures				Lattice energy (kcal/mol)		Dipole moments (debyes)	
	N[a]	Unit cell length (Å)	Unit cell angel (°)	d[a] < 4.0 Å	N	RMS E_{sublim}	N	RMS
Alcohols	16	0.21	1.63	0.12			8	0.26
Aldehydes/ketones	11	0.43	3.41	0.21	1	1.60	11	0.46
Amides	21	0.50	1.87	0.24	6	1.61	4	0.12
Amines	12	0.16	1.18	0.13	1	0.25	12	0.14
Carboxylic-acids	16	0.34	2.95	0.20	10	1.53[b]	10	0.32
Esters	5	0.44	2.47	0.19			5	0.17
Ethers	5	0.33	2.12	0.25			7	0.33
N-heterocycles	7	0.61	4.00	0.31	2	0.72	3	0.21
Hydrocarbons	25	0.23	2.56	0.15	12	2.18		
Sulphur-compounds	18	0.28	2.27	0.15			2	0.27
Total/average RMS	136	0.35	2.45	0.19	32	1.69	62	0.25

[a]N is the number of molecules; d is the RMS deviation in short interatomic distances

[b]Excluding the two polymorphs of oxalic acid—see text

Table 13 Fit of hydrogen bonds in the crystals (RMS deviations)

Family	CFF93		CFF70s	
	H···A distance[a]	D–H···A angle[a]	H···A distance[a]	D–H···A angle[a]
Alcohols	0.05	2.2		
Amides	0.06	4.0	0.06	3.9
Acids	0.08		0.07	7.7
Esters[b]	0.12			

Distances in Å, angles in degrees

[a]A and D are the acceptor and donor atoms

[b]In molecules with bifunctional groups

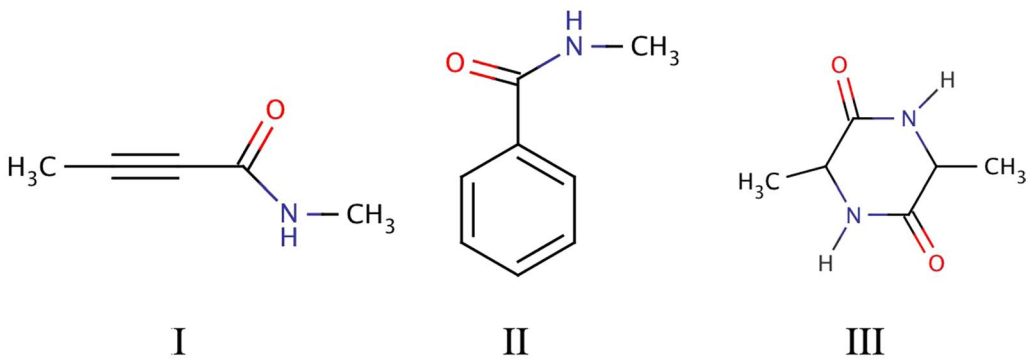
overall though one would like to see better than a 1 kcal error in lattice energies. Lattice energies deviated by 1.7 kcal (or ~5%) and dipoles by 0.25D. The rms deviation for unit cell lengths and angles was 0.35 Å, and 2.45°. Perhaps a better measure is the deviation of close interatomic distances (<4 Å), and the rmsd for those was 0.19 Å. An even better fit was obtained for hydrogen bonding properties (Table 13)—the rmsd of H···A distances and D–H···A angles ranged from 0.05 to 0.12 Å, and 2.2–4.0°, respectively (D = donor atom and A, acceptor). The accuracies of the predicted dipole moments, crystal lattice energies, and crystal lattice vectors are remarkably similar among the functional groups and packing motifs considered, demonstrating that the same method of deriving nonbonded parameters is appropriate to all cases.

The results are similar to those obtained in earlier studies of amides and carboxylic acids discussed above [85, 101, 344]. The increase in the number of molecules in the database affected the fit to experiment to a degree, but it remained reasonable, (compare Tables 1, 12), demonstrating the transferability of the parameters in the FF. For example, for amides, 21 crystals were included compared to 12 used in the original fit to amides [85]. The rmsd of the unit cell lengths and angles changed from 0.24 Å and 1.26° to 0.50 Å and 1.87° respectively, perhaps because such multifunctional compounds as *N*-methyltetrolamide(I), *N*-methylbenzamide(II), and 3,6-dimethyl-2,5-piperazinedione(III) were included.

As noted, such compounds give rise to a wide range of interatomic interaction environments and geometries sampled in the crystal data set. In contrast to the lattice constants, the rms of lattice energies improved—from 1.93 to 1.61 kcal/mol.

12.4.1 Systems with large deviations highlight deficiencies in FF and point way to improvements in functional form and assumptions

We note that there were some outliers which display larger deviations than might be expected from the overall fit. As pointed out above, these are more important than the systems fit well, as they provide valuable data for detecting deficiencies in the FF and lay the basis for future improvements. An example of this are the two polymorphs of the oxalic acid crystal [317]. The calculated sublimation energy for both crystal polymorphs deviates by ~7 kcal from the corresponding experimental values. It appears that this deviation is due to the lack of transferability of the electrostatic parameters to this system. Thus, when atomic partial charges for oxalic acid are determined from ab initio calculations, it is found that the charge on the carbonyl oxygen is appreciably less negative in oxalic acid than in the other dicarboxylic acids containing a methylene buffer (or even a hydrogen as in formic acid) adjacent to the carboxylic groups [100]. Thus the deviation in the calculated lattice energies suggests that there is a deficiency in the representation of the electrostatics,



I II III

which dominates the lattice energy of the oxalic acid crystal [101], perhaps the lack of inclusion of a better description of the spatial electron density, polarization, or deficiency in the bond increments used to calculate the atomic charges [317]. We observed, however, that the relative energies of the two polymorphs agree with experiment, so the error affects both equally (which would be expected if the error is due to the partial charges).

It should be noted that, with the exception of the cholesterol crystals and aqueous ion solutions discussed below, the properties of the systems discussed above were calculated by minimization. To be more rigorous these systems should be recalculated with MD as in the latter studies, which is very tractable with current computational resources. These systems provide a challenge to current force fields, especially as to whether inclusion of polarization effects and atomic multipoles, coupled with MD or MC could account for the outliers.

12.4.2 Extension to lipids and related molecules

The QDF method was extended to lipids and related molecules in the early 1990s in a series of studies [345–347]. To this end extensive quantum calculations were carried out on a set of model compounds followed by derivation and testing of a CFF as described above. In the first study the energy surfaces of dimethyl phosphate and methyl propyl phosphate were studied at the 6-31G* and 6-31+G* levels [345]. First and 2nd derivatives were calculated analytically, at various points on the energy surface. In addition, all stationary points were found as well. The structures, relative energies and vibrational frequencies were characterized in detail. Among other observations it was found that most of the critical structural features of phospholipids observed in crystals, in the gel and liquid crystal phases, and in solution are reproduced by the QM calculations, indicating that they are not due to induced structural changes by the medium [345]. These calculations were then extended to the characterization of the eight cis–trans isomers of inositol, as well as two phosphorylinositols [346]. DFT calculations were also applied to compare with the HF methods. Ring conformations and structures as well as relative energies, and free energies (calculated from vibrational spectra), were studied.

Following these investigations on lipid model compounds a study was undertaken to derive and test parameters for cholesterol, one of the major components of cell membranes [347]. At this point the only required parameters missing from the CFF93 force field were the C–C–C–O torsion of the A ring and its cross terms. These parameters were determined as described above, from QM calculations of the energy surface of 4-hydroxyl-1-butene. Energies, and the first and second derivatives of the energy with respect to coordinates, for the four energy minima and two transition states of this compound were characterized. Reasonably good agreement between the force field and quantum results were obtained for the properties of these six states. For example, the largest deviations were found in one of the transition state structures, with a difference in torsion angle of 7° (ab initio 128.1° vs. 121.3° in the FF) and a maximum relative energy difference of 0.6 kcal/mol (6.1 and 5.5 kcal/mol for QM and FF respectively).

12.4.3 Simultaneous testing of intramolecular and intermolecular force field by MD

The FF was then tested by extensive simulation of both the low-temperature neutron-diffraction structure of cholesteryl acetate crystals and the X-ray diffraction crystal structure of the cholesterol crystal. It is important to note that no new nonbond parameters were required for this test. These parameters were transferred from the corresponding functional groups derived previously [317], and all valence parameters with the exception of the C–C–C–O torsion were also transferred. Thus, these studies constituted an additional, stringent test of both nonbond and intramolecular parameters. Both energy minimization and molecular dynamics simulations of these crystals were carried out. The cell vectors of the low temperature cholesterol acetate structure were reproduced to within 2.4% maximum error (in the unit cell vector a), while the molecular structures in the crystal environment were also well accounted for with a maximum rms deviation in positions and torsions of heavy atoms in the two molecules of 0.097 Å and 2.35° respectively (the internals containing hydrogen were not fit as well with the corresponding quantities being 1.36 Å and 3.32°). Interestingly, the rms deviation of all non-hydrogen atoms in the unit cell of the time averaged MD structure from experiment was only 0.06 Å, even better than that obtained with minimization.

The cholesterol crystal is more complex, with eight molecules exhibiting considerable flexibility in the asymmetric unit, resulting in some molecular regions being ill-defined with large B-factors [347]. The observed fit to the crystal cell vectors obtained by crystal minimization and by MD simulations was reasonable with a maximum deviation of ~1% in the b vector. Several large deviations are observed including deviations in torsion angles up to 18° in molecule E of the asymmetric unit. However, it was pointed out that all the largest deviations are associated with the C_{18} atom. The experimentally determined position of this atom is highly questionable as reflected in its unusually long C–C bond length (1.705 Å) as well as the unusual C–C–C angles (99.2° and 127.1°) it participates in. Excluding this atom, the maximum deviations between experimental and calculated internal coordinates are 0.07 Å for bond lengths, 6° for angles, and 10° for torsions for atoms with B-factors < 10.

Though better, these deviations are still large. Interestingly, once again MD yields a better fit to the overall structure. For atoms with B-factors < 15 Å^2, the corresponding rms deviations were 0.152 and 0.189 Å for the MD averaged and energy-minimized structures, respectively. The fit to the more well determined atoms, B-factors < 10 Å^{2,} maintained this trend with even better agreement to experiment (rms of 0.14 Å for the time averaged structure and 0.175 Å for the minimized structure).

12.5 Combination rules: as important as parameters themselves

The nonbond interactions for unlike atoms are generally approximated from the interactions for like atoms by the use of combining rules involving the nonbond parameters of the like atom–atom pair potentials. This reduces the number of parameters dramatically. For the most part, standard assumptions for these rules have been adopted in biomolecular force fields without investigation. As discussed above the common rules routinely used were the so-called Lorentz–Berthelot rules derived in the late 1800's [348, 349]. These consist of (1) a geometric mean rule for the well depth, ε, of the nonbond potential; and (2) a geometric or arithmetic mean rule for the position van der Waals radius, r^*, of the nonbond potential (Eq. 4e, Fig. 2). Not much attention had been paid to the nature or validity of these combining rules in the early force fields. (In retrospect this is somewhat incongruous as with the exception perhaps of hydrogen there are more unlike interactions than like interactions for most intermolecular systems). In the early 1980s with more accurate experimental data, it became clear that these simple combining rules lead to large errors in predicting the potential parameters of mixed rare gas atom pairs. An example of the magnitude of the errors introduced by these simple combining rules is given in Table 14 [27]. This led physicists to propose more elaborate combination rules that significantly improved the fit of the rare gas data, as well as other simple molecular systems [350, 351].

The problem escaped the notice of the molecular mechanics and dynamics community until the early 1990s when Halgren derived a new set of combining rules based on high quality quantum calculations of rare gas interactions [316] (Eqs. 23a, b).

$$r_{ij}^* = \frac{r_{ii}^{*3} + r_{jj}^{*3}}{r_{ii}^{*2} + r_{jj}^{*2}}; \quad \varepsilon_{ij} = \frac{181.16 G_i G_j \alpha_i \alpha_j}{(\alpha_i/N_i)^{1/2} + (\alpha_j/N_j)^{1/2}} \frac{1}{r_{ij}^{*6}} \tag{23a}$$

$$\varepsilon_{ii} = 1/2kG_i^2 C_{6ii} \frac{1}{r_{ii}^{*6}} \tag{23b}$$

where as above, r^* is the van der Waals radius, ε is the well depth, and N and α are the effective number of electrons and atomic polarizability respectively. G is a scaling parameter determined by requiring that the relationship in Eq. (23b) holds for each of the like-pair interactions of the rare gases (in this Eq. 23b, k is a unit conversion factor).

12.5.1 Waldman–Hagler combining rules

At roughly the same time we took a graphical approach to the combination rule problem [27]. This was in part stimulated, by the fact that the newer combination rules generally involved additional parameters such as polarizabilities, ionization potentials, or dispersion force coefficients [316, 350, 351] (see e.g. Eqs. 23a, b). Therefore, these combination rules are less easily incorporated into molecular force fields because they lead, to an excess of parameters that are not easily determined for systems other than rare gases.

We showed that the combining rule formulas could be reformulated to a single-parameter problem. This allowed graphical analysis of the combining rules against experimental data [352]. This analysis technique showed that the rare gas potentials do not obey a single combining rule for ε, but do follow a single combining rule for r^*. However, it was shown that a combining rule using both ε and r^* could be used to predict the ε parameters for the mixed rare gas pairs. The resulting combining rules derived [27] and then used in the CFF93 [317] and elsewhere (see eg [353–356] and references therein) are defined in Eq. (21o).

The resulting fit to the rare gas properties using the standard geometric combining rule is given in Table 14. It is seen by comparing the results given in Table 15 with those in Table 14 that the new formulation of the combining rules results in a dramatic improvement in accounting for the nonbond interactions of the rare gases.

One of the important advantages of this formulation lies in the simplicity of the equations and most importantly that no additional experimental information beyond the values of r^* and ε is required.

Table 14 Errors introduced in intermolecular energies by use of geometric mean combining rule for ε

	He	Ne	Ar	Kr	Xe
He	*10.44*	*19.44*	*30.01*	*31.05*	*29.77*
Ne	7.7%	*42.00*	*64.17*	*67.32*	*67.25*
Ar	28.1%	20.2%	*141.5*	*165.8*	*182.6*
Kr	46.4%	35.4%	0.9%	*197.8*	*225.5*
Xe	79.7%	59.5%	7.8%	3.3%	*274.0*

Upper right triangle (italics) is experimental data. Lower left triangle is % error

Table 15 Test of geometric mean combining rule for $\varepsilon(r^*)^6$ using sixth power arithmetic mean combining rule for r^*

	He	Ne	Ar	Kr	Xe
He	*10.44*	*19.44*	*30.01*	*31.05*	*29.77*
Ne	6.3%	*42.00*	*64.17*	*67.32*	*67.25*
Ar	−1.0%	2.1%	*141.5*	*165.8*	*182.6*
Kr	−0.8%	2.7%	−0.9%	*197.8*	*225.4*
Xe	−0.2%	0.9%	−2.0%	0.1%	*274.0*

Upper right triangle (italics) is experimental data. Lower left triangle is % error

12.5.2 Consequences of inaccurate combining rules

Perhaps the more important question is whether the impact of the different combining rules on the nonbond parameters derived for biomolecular systems is significant. The answer is a resounding yes! They indeed have a major impact on both the nonbond parameters themselves as well as the fit to the model systems used to derive these parameters, as might be expected from the results on the rare gases. The effect on crystal properties of model compounds as reflected in structural properties can be seen in Table 16.

The AG rules give about the same or marginally more accurate results only for the amides and amines. For alcohols, the deviations with the AG rules are roughly twice those obtained with the WH (Waldman–Hagler) combining rules, and overall the deviations in unit cell vector lengths, angles, and interatomic distances in the crystal are ~50% greater with the AG combining rules as compared to the WH combining rules. The largest errors introduced by the commonly used AG rules occur for the sulfur compounds. This is consistent with the results for the rare gases where as seen from Table 14, the error in the intermolecular energy is largest 78% for the He–Xe interaction, i.e. the pair with the greatest disparity in atomic radii.

Desgranges and Delhommelle [354] found that the use of the Lorentz–Berthelot (AG) rules result in the Gibbs free energy of the solution being 7% higher than the value predicted by the Waldman–Hagler [27] rule. They conclude that this "… emphasizes the importance of the combining rule for the determination of hydration free energies using molecular simulations."

In a subsequent study, Al-Matar and Rockstraw [353] derived a generating equation for combination rules and in a comparison of various rules found the WH rule to be generally superior to others, including the MMFF (Halgren) rule [316], which led to large errors in some cases. They recommended using either WH or their generating equation [353].

13 Summary and conclusions

In this perspective, we have reviewed the history of the origins of biomolecular FFs and the advancements made in the twentieth century. The evolution of the rigor of the functional form describing the physics, approximations made and their effects, have been discussed. A synopsis of some of the salient points of this history are summarized below. In the following paper, we discuss the advances made since the initial FFs and the applications reviewed above.

Table 16 Effect of combining rules on crystal structure: RMS deviation from experimental unit cell lengths, angles and distances < 4 Å with Wildman–Hagler combination rules and with arithmetic–geometric mean rules (AG)

Molecular family	RMS deviation							
	No. of crystals	Unit cell length (Å)		Unit cell angle (°)			Distances < 4 (Å)	
		WH	AG	WH	AG		WH	AG
Alcohols	12	0.13	0.33	1.91	3.45		0.09	0.17
Amides	12	0.26	0.24	1.78	1.74		0.15	0.16
Amines	9	0.16	0.15	1.09	0.80		0.14	0.13
Carboxylic acids	14	0.35	0.62	3.20	4.34		0.20	0.36
N-heterocycles	9	0.41	0.39	1.49	2.46		0.17	0.21
Hydrocarbons	22	0.11	0.16	1.64	2.73		0.21	0.27
Sulfur compounds	15	0.35	0.91	2.50	2.83		0.18	0.47
Total/average RMS	93	0.25	0.40	1.94	2.62		0.16	0.25

13.1 Polarizability, multipoles, nonbond anisotropy, and coupling between internals recognized and accounted for in calculations in the 1930s through the 1950s

We saw that the first "force fields" were inspired by the desire to account for vibrational spectra and describe the normal modes of vibration in the early decades of the twentieth century [9, 17, 18]. Concurrently there was an interest in accounting for the properties of water. The first energetic calculations of water (and ice) were carried out by Bernal and Fowler in 1933 [8], who included vdW and electrostatic interactions, and recognized, even in 1933, that off-atom charges were necessary. Rowlinson introduced a 4-point model for water in 1950 [201]. Anticipating the refinements of this century, he included higher order multipoles and polarizability in his calculation of the structure and energy of the ice lattice and the second virial coefficient of water.

These were soon followed by the application of the FF concept to study conformations and energies of molecules. The first calculations were carried out by hand, as were those described above, and addressed steric effects, especially with respect to reaction mechanisms [28–31]. The modern era of classical computational chemistry was ushered in by Hendrickson in 1961 [13], who, for the first time, exploited computers to carry out MM calculations (of the conformational energetics of medium ring hydrocarbons), and Wiberg who introduced Cartesian coordinates, derivatives, and minimization in studying strain energies of cycloalkanes [33].

13.2 The forebearers of modern force fields

The potential of applying computers to energetics of organic and bio molecules was soon recognized by numerous researchers. Scheraga [357], Liquori [38], Ramachandran [37] and Flory [39] among others pioneered the application of these techniques to peptides. They produced the first Ramachandran (φ–ψ) maps, the earliest energetic studies of the elements of protein structure. Concurrently, several labs addressed the structure and thermodynamics of small organic compounds and began to focus on the energy surfaces or force fields underlying these properties. Seminal advances were introduced by Hendrickson, who carried out the first molecular mechanics calculations on a computer [13] and Wiberg, who introduced the use of Cartesian coordinates in these calculations [33]. Notable among the pioneering works which ushered in the modern era are the works of the Lifson [26, 51] and Allinger [69] laboratories. Warshel and Lifson's consistent force field (CFF) paradigm, where the principle that a FF should "give the best possible agreement with a large amount and variety of observed data" is a cornerstone of the field. Note, Lifson emphasized both size of the compound array; i.e. a large diverse set of compounds (~ 10–20 acyclic, cyclic, strained), and variety; e.g. structure, conformational energies and barriers, vibrational frequencies, and lattice energy and structure. The accompanying CFF software they developed evolved into many of the current biomolecular simulation packages used today. Allingers initial work evolved to a decades long exploration onto the improvement of the representation of molecular energy surfaces. He systematically accumulated reported deviations, added these systems to an already large set of observables, improved the functional form accounting for the physics and reparametrized the FF. This led to the important MMn series of force fields and algorthims [358]. These were truly pioneering works and still some of the best papers on FF derivation. Unfortunately, in more recent years, many of the FF reparametrization efforts have strayed from these principles, using smaller, and different data sets in each subsequent reparametrization.

13.3 The hydrogen bond

Prior to 1974 explicit functional forms representing an orientaional covalent interaction were used to account for the Hydrogen bond [46]. Through analysis of crystals packing and sublimation energies, Hagler and Lifson (H–L) demonstrated that in fact electrons about the polar hydrogen were shifted along the bond into the more electronegative nitrogen atom. This allowed the donor to come closer to the acceptor increasing the attraction and accounting for a major portion of the H-bond energy [86]. Thus, the H–L representation including vdW and electrostatic interactions with a modified donor in which exchange repulsion of the donor hydrogen was significantly diminished or removed, was able to account for the properties of these amide crystal systems as well or better than the more complex functions with fewer parameters [85, 86]. This was subsequently found to hold true for the carboxylic acids as well [101]. It has since been adopted in essentially all current simulation packages, though as pointed out in the text, it is still far from perfect (Fig. 4).

13.4 The power of the Lifson paradigm and Allinger methodology

It was seen that, though the H–L form is now essentially universally adopted, new complex explicit H-bond potentials, containing additional parameters, continued to be introduced and used in applications for some two decades after the H–L lattice studies, only to be ultimately discarded [102]. This clearly represents an inefficiency of efforts that could have been devoted to building upon, and improving the representation. It was noted that, had the Lifson paradigm, or Allinger's methodology of systematic incremental improvements of the functional form representing the physics based on collected problematic systems, been retained,

this inefficiency could have been avoided. Both of these protocols call for a large number of compounds and diverse properties be included in the training and validation tests, and all compounds and properties employed in the derivation and testing of the previous version of the FF tested with the new functional form or parameters. The unproductive detours in H-bond derivations involved relatively small sets of new different observables, sometimes only qualitative. Had the new proposed functions been tested against the crystal properties treated previously, it would have been immediately obvious that these more complex functions with additional parameters not only didn't improve the fit but often introduced additional artifacts! This could have saved 10–20 years of effort that has been deemed flawed.

Though this is history, unfortunately it is important to learn from it as similar practices persist. As discussed in part II of this perspective [113]. Thus following the Lifson–Allinger protocols, and at the very least when deriving a new and "more accurate" FF, addressing *all* observables used in deriving and testing the previous FFs which are being improved upon, as well as augmenting them, is an initial investment which will avoid much wasted effort in application of flawed FFs and subsequent reparametrizations.

13.5 The pivotal role of QM in FF derivation

It was seen that from the earliest days of FF research QM has played a key role in the determination of both force constants and functional forms. Thus, in the early 1970s Scheraga exploited semiempirical QM calculations (CNDO/2) to determine charges and rotational barriers in peptides [42, 131]. They also developed the 12–10 H-bond potential by fitting CNDO/2 energy surfaces [103]. Clementi was the first to exploit rigorous SCF-MO calculations to the derivation FFs, first deriving a water potential in 1973 [143] and water-peptide intermolecular FFs later in the decade [359]. He continued these studies for decades producing an impressive body of work in these areas [159]. In 1976 Hagler et al. [171] exploited HF calculations in combination with crystallographic structure determinations and MM to demonstrate that the rotational barriers around φ and ψ were small with the former being of the order of 0.7 kcal/mol, while it is likely the intrinsic rotational barrier about ψ is negligible.

Since these early applications, QM has played a role in virtually every effort to derive FFs, most often in the determination of charges and rotational barriers, but it has also been applied to all inter and intramolecular terms.

13.6 Adapting the pioneering MM approaches to biomolecular systems

The 1980s saw the adaptation of the pioneering work of Hendrickson and Wiberg as developed by Allinger and especially Lifson's CFF software, to the development of the first biomolecular FFs and simulation software, including AMBER [210], the peptide valence FF (VFF) version of the CFF [225], CHARMM [109], OPLS [218] and Gromos [214]. With the exception of the VFF all these FFs invoked the united atom approximation where nonpolar hydrogens were omitted in favor of "united" methyl or methylene "atoms," in an attempt to treat large protein systems for longer simulation times. In general, again with the exception of VFF, these FFs were quadratic diagonal FFs and the parameters were chosen on a very limited set of data, and in some cases taken from the literature, resulting in an inconsistent set of force constants.

In the 1990s and early 2000's the crude FFs incorporating the united atom approximation were discarded and more rigorous parametrizations were carried out. Thus, as we described, new versions of CHARMM [271], AMBER [268], OPLS [277] and Gromos [280], were all introduced in this period. However. unlike the others Gromos has continued to invoke the united atom approximation. In contrast to the early incarnations, these FFs were fit to a large set of experimental data and tested against additioanl observables, including crystal data, outside the training set. As described, the CHARMM22 parametrization was particularly rigorous [271], following the Lifson paradigm of including a large number of diverse observables, and as noted above, documenting problematic systems indicating deficiencies in the simple diagonal quadratic fixed charge functional form characterizing all these FFs [271].

These extensive data sets, and especcially the documentation of discrepancies provided the framework for systematic improvement of the FFs following Allingers example. Unfortunately, as described in the second half of this perspective [113], subsequent "refinements" of these FFs discarded many of these observables, fitting entirely different data, rather than adding the additional data to the comprehensive sets used previously. This led, as in the case of the H-bond potential described above, to flawed FFs and the need for many reparametrizations, a situation continuing into the present [113].

13.7 Extending the Lifson paradigm: incorporating quantum derivative fitting

As we saw, energies, charges, and vibrational properties derived from quantum mechanics have played a central role in FF development from the beginning. In 1988 and the years following, this concept was extended by the Hagler lab, who noted that an energy surface (or any surface for that matter) can be described by sampling the value, slope and curvature (energy, first and second derivatives respectively) of points on it [309]. As pointed out by Lifson and Warshel [26] these are the first three terms in the Taylor expansion of

the energy and correspond to the system's energy, structure and vibrational frequencies respectively. These quantities are readily available from QM and can provide orders of magnitude more information, for the same computational cost, than, for example simply tracing out the adiabatic energy of rotation about a torsion as a function of torsion angle by single point calculations, or calculating dimer energies at different geometries.

As discussed above, Hagler et al. developed an algorithm to fit FF parameters to these quantities by least squares and showed that it was a powerful methodology, capable of rapidly assessing functional forms in terms of their ability to represent the physics of molecular energetics. In addition to deriving the corresponding potential parameters, a complete FF can be determined by fitting these quantities [309–311, 313]. For example fitting of the Quantum energy surface revealed significant deficiencies in the representation of out-of-plane distortions by the standard improper torsion angle which could be addressed simply by using either the distance of the apex atom above the plane of the three atoms it's bonded to, or the Wilson definition [309]. They also demonstrated the significance of cross terms in accounting for relative energies. Another attractive feature of this algorithm is it can be easily automated to return an entire QM FF. It was then shown how this rich source of information on the nature of the energy surface could be incorporated into Lifson's methodology by simply scaling the quantum energy surface or FF by fitting a small number of scale factors to the available experimental data [321]. Unfortunately this scaling was not extended beyond the hydrocarbons.

13.8 Partitioning QM energy into atom–atom contributions

Dinur and Hagler went on to develop the concept of derivatives noting that the second derivatives extracted specific atom–atom interactions from for example the monomer–monomer interaction in dimers [96, 199, 339]. Similar reasoning showed that the intramolecular energy could be partitioned into individual bond, angle, and torsion strain by analogous 2nd derivatives of the quantum energy [312, 360]. Derivatives also allowed for the derivation of natural charges, atomic multipoles, and charge flux, or the important geometry dependence of charges [341, 361]. These are powerful tools for interrogating the Quantum energy surface and teasing out both the functional form of these interactions as well as their associated parameters. They also allow us to account for such physical phenomena such as charge flux, and anisotropy of nonbond interactions, to name two. These developments are discussed in more detail in the following paper [113]. Unfortunately, with only few exceptions, they have not been incorporated into ongoing FF research.

13.9 Combination rules: more important than they've been given credit for

As outlined, there has been a great deal of attention and discussion devoted to the derivation of nonbond parameters. However, this effort has focused almost exclusively on the like atom–atom interactions. The mixed interactions have simply been assumed to follow simple combining rules, such as a geometric mean rule for the well depth parameter, ε, of the nonbond potential; and a geometric or arithmetic mean rule for the well depth position, r^*. Further examination revealed however, that these rules lead to large errors in the parameters for the mixed atom–atom pairs [27]. This was noticed and addressed by Halgren [316] and Waldman and Hagler [27]. The latter proposed simple combination rules based on graphical analysis, though it was shown that the rule for ε was a function of r^*. These improved combining rules have been slow to be adopted in biomolecular simulations, despite their profound effect on structure and energetics [27, 317], though this may be changing.

13.10 In conclusion

The second half of the twentieth century saw the birth of the field of conformational analysis and computer simulations of organic and biomolecular systems. The early calculations of molecular properties by hand gave way in 1961 to Hendrickson's first relative conformational energy calculations (of hydrocarbon rings) on recently introduced computers. This was followed by rapid adoption by early adopters in two schools, the molecular mechanics school who focused on structural and energetic properties of small organic compounds, and the peptide conformation school, who applied FFs to study the conformational properties of dipeptides and peptide helices. The early MM pioneers included such names as, Lifson, Allinger, Bartell, while the conformational group included Scheraga, Liquori, Ramachandran, and Flory among other. The critical importance of the FFs underlying these calculations was immediately recognized by these pioneers, starting from the first paper by Hendrickson.

The next wave of development, occurring roughly in the 1980s saw the application of these techniques to proteins and the development of the early protein FFs by labs including Karplus, Kollman, Jorgensen, Van Gunsteren, Berendsen and Hagler among others, while Lifson, and Scheraga also continued to expand their efforts in this field. It also saw the introduction of software packages from these labs to carry out these simulations (CHARMm, AMBER, GROMOS, CFF/Discover, ECEPP) The 1970s and 1980s also saw the incorporation of QM into FF derivations, with the most extensive application by Clementi, the introduction of rigorous methodologies to derive FFs from diverse properties of large numbers of small model compounds (e.g. Lifson

Table 17 Seminal developments in the development of FFs for organic and biomolecular systems

Year(s)	Contribution/significance	Citations
1930s	First FFs described by spectroscopists to calculate vibrational spectra	Urey and Bradley [15], Wilson [9]
1933	First water potential—employed off atom charges, realized atom-centered charges inadequate	Bernal and Fowler [8]
1946	Calculation of steric effects on reaction. Van der Waals interactions are calculated to account for steric effects in reactions. Anisotropy of the van der Waals interaction and the differences between intra and intermolecular vdW interactions accounted for, aspects still not included in biomolecular simulations	Dostrovsky et al. [30]
1951	Rowlinson water potential. Incorporated multipoles as well as polarizability, anticipating these contributions in biomolecular FFs by half a century	Rowlinson [201]
1961	First molecular mechanics calculation by computer	Hendrickson [13]
1963	Derivation of transferable (spectroscopic) FF—demonstration of need for coupling terms and superiority of valence FF over Urey–Bradley (and diagonal) FFs	Snyder and Schachtschneider [19, 20]
1965	Introduction of Cartesian coordinates as a basis for calculations and use of minimization algorithm	Wiberg [33]
1965–	Application of energy methods to peptides—first φ, ψ maps	Liquori [38], Ramachandran [36], Flory [39], Scheraga [362]
1965–	First exploitation of crystal structure and sublimation energies to derive intermolecular FFs	Kitaigorodskii and Mirskaya [82, 84], Williams [79], Scheraga [61]
1967	MM1 FF published. First in Allingers series of MM	Allinger et al. [69]
1969	Introduction of Consistent FF paradigm. Objective optimization of parameters (LSQ) against numerous observables of different physical observables in different states	Lifson and Warshel [26, 54]
1973	Introduction of coupling terms (VFF) to molecular mechanics energy function—demonstrating need for these terms and superiority of VFF to Urey–Bradley or diagonal FF	Ermer and Lifson [55]
1973	Clementi water potential derived exclusively from Hartree–Fock calculations of numerous water dimer configurations. First potential derived from HF calculations	Popkie et al. [143]
1974	Hagler–Lifson H-bond. Derived the now commonly accepted H-bond form as a vdW-coulomb representation with negligible donor radius, and showed it accounted for H-bonded systems as any explicit formulation to date (w/I the contest of a n-6-or exp-6-1 FF)	Hagler-Lifson et al. [85, 86]
1975	Development of the ECEPP rigid geometry FF for proteins	Momany et al. [67]
1976	Intrinsic φ–ψ torsion barriers determined from first HF application to peptide conformation, along with X-ray structure determination and lattice energy calculations	Hagler et al. [171]
1977	MM2 published: Improves representation of physics to correct discrepancies uncovered by extensive application of MM1 by Allinger and others	Allinger et al. [72]
1977	First MD simulation of a protein (PTI, in vacuo, no hydrogens)	Karplus et al. [211]
1981	AMBER FF and program published—simplified FF for proteins, no coupling terms, anharmonicity or hydrogens (united atom) and a 10–12 H-bond potential	Weiner et al. [210]
1981	TIP3 model for water derived	Jorgensen [216]
1982	First simulation of a drug receptor complex and, the first all atom valence FF for proteins developed (CVFF)	Dauber et al. [220, 221]
1983	CHARMM FF and Program published—simplified FF for proteins, no coupling terms, anharmonicity or hydrogens and a complex angle dependent H-bond potential	Brooks et al. [109]
1985	OPLS FF united atom FF derived from MC calculations of liquid amides published	Jorgensen and Swenson [218]
1986	Restoration of nonpolar hydrogens to AMBER resulting in all-atom AMBER FF. Charges derived by fitting ESP, and explicit H-bond potential discarded in favor of Hagler–Lifson formulation. Coupling and anharmonicity not included (Diagonal quadratic form)	Weiner et al. [213]
1987	GROMOS FF and Program published—simplified FF for proteins, no coupling terms, anharmonicity or hydrogens, Hagler–Lifson H-bond	van Gunsteren and Berendsen [214]
1988	Quantum Derivative Fitting method for derivation of FF analytical form and parameters presented, comprised of fitting QM energies, Hessians, 1st and second derivatives and dipole moments of numerous distorted geometries of multiple diverse compounds in a family. Shown to be extremely efficient and easily incorporated into a workflow	Maple et al. [309]

Table 17 (continued)

Year(s)	Contribution/significance	Citations
1988	The errors introduced by fatally flawed improper torsion OOP representation were characterized and replaced with forms invoking Wilson coordinate, or distance of central atom from plane	Maple et al. [309]
1989–1996	MM3-MM4 FFs derived. Improved representation of physics including terms to represent coupling and anharmonicity	Allinger et al. [111, 265]
1992–1993	Combination rules—large errors in common geometric–arithmetic van der Waals combination rules documented and improved rules derived	Halgren [316], Waldman and Hagler [27]
1993	NCC-vib flexible and polarizable potential for water published. One of most rigorously derived and tested water potentials	Corongiu and Clementi [151]
1994–2001	CFF93, a fully coupled anharmonic FF for proteins, employing valid out of plane representation and combination rules	Maple et al. [310], Ewig et al. [314]
1996	OPLS-AA all atom harmonic, diagonal OPLS FF published	Jorgensen et al. [276]
1996	MMFF94 published. Diagonal anharmonic FF, with 14–7 nonbond, Wilson OOP, and Halgren combination rules. Employed PROBE to derive parameters from high level QM Hessians, first derivatives and energies	Halgren [281]
1998	All atom CHARMm FF published. Diagonal quadratic FF with Hagler–Lifson H-bond. Parametrized and tested against a large set of diverse properties including crystals	MacKerell et al. [271]

and Allinger), reformulation of the H-bond potential to the current simple vdW electrostatic model (Hagler–Lifson), and the introduction of the first simplified united atom FFs to allow for the calculation of protein (in vacuo). These decades also saw the adoption of molecular dynamics and Monte Carlo from the physics community, and the introduction of aqueous solvent into simulations of proteins (Hagler, Clementi). Research continuing into the 1990s saw the abandonment of the crude United atom potentials and a rigorous reparametrization of the CHARMm, AMBER and OPLS FFs, by the Karplus, Kollman and Jorgensen groups respectively, the introduction of a new methodology to derive FFs from QM energy surfaces characterized by the (QM) energies, first and second derivatives sampled over the surface, along with a complete protein valence FF, CFF, by the Hagler group, and more rigorous representations of out of plane functions and combination rules (Halgren, Hagler). It also saw the first applications of these techniques to drug discovery.

For convenience, a summary of the seminal developments are given in Table 17.

This was clearly an exciting period, which saw the field emerge and develop, side by side with parallel developments in hardware capabilities and X-ray crystallography, from calculations on small hydrocarbons to large biomolecular systems and applications to drug design. In the second half of this perspective we will review the continued evolution of the field in the twenty-first century.

Acknowledgements We would like to thank Dr. Mike Gilson, for reading parts of the manuscript and helpful discussions and Dr. Ruth Sharon for reading, discussing and valuable help with editing. We also thank Eitan Hagler for help with the figures. Special thanks to the editor, Dr. Terry Stouch for his invitation to write this perspective, encouragement, and endless patience.

Abbreviations

AG: Arithmetic–geometric; Ala: Alanine; AMBER: Assisted model building with energy refinement; AMOEBA: Atomic multipole optimized energetics for biomolecular applications; BNS: Ben Naim–Stillinger; CFF: Consistent force field; CHARMM: Chemistry at HARvard Macromolecular Mechanics; CNDO: Complete neglect of differential overlap; COMPASS: Condensed-phase optimized molecular potentials for atomistic simulation studies; CVFF: Consistent valence force field; DFT: Density functional theory; ECEPP: Empirical conformational energy program for peptides; EHT: Extended Huckel theory; FF: Force field; FQ: Fluctuating charges; Gly: Glycine; GROMOS: *GRO*ningen *MOL*ecular *S*imulation; Hyp: Hydroxyproline; LCAO: Linear combination of atomic orbitals; LJ: Lennard-Jones; LSQ: Least squares; MC: Monte Carlo; MCMS FF: Momany, Carruthers, McGuire, and Scheraga Force Field; MCY: Matsuoka–Clementi–Yoshimine; MD: Molecular dynamics; MDDR: MDL drug data report; MDL: Molecular design limited; MM: Molecular mechanics; MMFF: Merck molecular force field; NMA: *N*-methylacetamide; OPLS: Optimized potential for liquid simulations; OPLS-AA: OPLS-AA/L OPLS all atom FF (L for LMP2); PCILO: Perturbative configuration interaction using localized orbitals; PDB: Protein data base; PEFC: Potential energy function consortium (Biosym); QCPE: Quantum chemistry program exchange; QDF: Quantum derivative fitting; QDP: Charge dependent polarizability; QM: Quantum mechanics; RESP: Restrained electrostatic potential;

RMS: Root mean square; RMSD: Root mean square deviation; SCF-LCAO-MO: Self-consistent field-linear combination of atomic-molecular orbital (wave function); SDFF: Spectroscopically determined force fields (for macromolecules); SPC: Simple point charge (water model); ST2: Four point water model replacing Ben-Naim Stillinger (BNS) model; STO: Slater-type atomic orbitals; TIP3P: Transferable intermolecular potential (functions for water, alcohols and ethers); TTBM: Tri-*tert*-butylmethane; UB: Urey–Bradley; VDW: van der Waals; VFF: Valence force field; WH: Waldman–Hagler

References

1. Hofmann AW (1865) On the combining power of atoms. Proc R Inst 4:401–430
2. Pasteur L (1848) Recherches sur les relations qui peuvent exister entre la forme cristalline, la composition chimique, et le sens de la polarisation rotatoire. Anal Chim Phys 24:442–459
3. Le Bel JA (1874). Bull Soc Chim Fr 22:335–347
4. van't Hoff JH (1874) Sur les formules de Structure dans l'Espace. Arch Neerl Sci Exactes Nat 9:1–10
5. Dreiding VAS (1959) Einfache molekularmodelle. Helv Chim Acta 48:1339–1344
6. Koltun WL (1965) Precision space-filling atomic models. Biopolymers 3:665–679
7. Larson GO (1964) Atomic and molecular models made from vinyl covered wire. J Chem Educ 41:219
8. Bernal JD, Fowler RH (1933) A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. J Chem Phys 1:515–548
9. Wilson EBJ (1939) A method of obtaining the expanded secular equation for the vibration frequencies of a molecule. J Chem Phys 7:1047–1052
10. Barton DHR (1946) Interactions between non-bonded atoms, and the structure of cis-decalin. J Chem Soc 174:340–342
11. Mason EA, Kreevoy MM (1955) A simple model for barriers to internal rotation. J Am Chem Soc 77:5808–5814
12. Pitzer KS, Donath WE (1959) Conformations and strain energy of cyclopentane and its derivatives. J Am Chem Soc 81:3213–3218
13. Hendrickson James B (1961) Molecular geometry I. Machine computation of the common rings. J Am Chem Soc 83:4537–4547
14. Wilson EBJ, Decius JC, Cross PC (1955) Molecular vibrations: the theory of infrared and Raman vibrational spectra. McGraw-Bill, New York
15. Urey HC, Bradley CA (1931) The vibrations of pentatonic tetrahedral molecules. Phys Rev 38:1969
16. Simanouti T (1949) The normal vibrations of polyatomic molecules as treated by Urey–Bradley field. J Chem Phys 17:245–248
17. Dennison DM (1931) The infrared spectra of polyatomic molecules part I. Rev Mod Phys 3:280–345
18. Cross PC, Van Vleck JH (1933) Molecular vibrations of three particle systems with special applications to the ethyl halides and ethyl alcohol. J Chem Phys 1:350–356
19. Snyder RG, Schachtschneider JH (1963) Vibrational analysis of the n-paraffins-I. Assignments of infrared bands in the spectra C_3H_8 through n-$C_{19}H_{40}$ Spectrochim Acta 19:85–116
20. Schachtschneider JH, Snyder RG (1963) Vibrational analysis of the n-paraffins-II. Normal co-ordinate calculations. Spectrochim Acta 19:117–168
21. Williams JE, Stang PJ, Schleyer PVR (1968) Physical organic chemistry: quantitative conformational analysis; calculation methods. Annu Rev Phys Chem 19:531–558
22. Scott RA, Scheraga HA (1966) Conformational analysis of macromolecules. I. Ethane, propane, n-butane, and n-pentane. Biopolymers 4:237–238
23. Scott RA, Scheraga HA (1966) Conformational analysis of macromolecules. II. The rotational isomeric states of the normal hydrocarbons. J Chem Phys 44:3054–3069
24. Jones JE (1924) On the determination of molecular fields-II. From the equation of state of a gas. Proc R Soc London Ser A 106:463–477
25. Buckingham RAA (1938) The classical equation of state of gaseous helium, neon and argon. Proc R Soc London Ser A 168:264–283
26. Lifson S, Warshel A (1968) Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and n-alkane molecules. J Chem Phys 49:5116
27. Waldman M, Hagler ATT (1993) New combining rules for raregas van der Waals parameters. J Comput Chem 14:1077–1084
28. Hill TL (1946) On steric effects a reactive potential for hydrocarbons with intermolecular interactions on steric effects. J Chem Phys 14:465
29. Westheimer FH, Mayer JE (1946) The theory of the racemization of optically active derivatives of diphenyl. J Chem Phys 14:733–738
30. Dostrovsky BI, Hughes ED, Ingold CK (1946) Mechanism of substitution at a saturated carbon atom.I. The role of steric hindrance. J Chem Soc 173–194
31. Westheimer FH (1956) In: Newman MS (ed) Steric effects in organic chemistry, Chap. 12. Wiley, Hoboken
32. Allinger NL (1959) Conformational analysis. III Application to some medium ring compounds. J Am Chem Soc 81:5727–5733
33. Wiberg KB (1965) A scheme for strain energy minimization. Application to the cycloalkanes. J Am Chem Soc 87:1070–1078
34. Edsall JT et al (1966) A proposal of standard conventions and nomenclature for description of polypeptide conformation. J Biol Chem 241:1004–1008
35. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7:95–99
36. Ramakrishnan C, Ramachandran GN (1965) Stereochemical criteria for polypeptide and protein chain conformation II. Allowed conformations for a pair of peptide units. Biophys J 5:909–933
37. Ramachandran GN, Venkatachalam CM, Krimm S (1966) Stereochemical criteria for polypeptide and protein chain conformations III. Helical and hydrogen-bonded polypepide chains. Biophys J 6:849–872
38. De Santis P, Giglio E, Liquori AM, Ripamonti A (1965) van der Waals interaction and the stability of helical polypeptide chains. Nature 206:456–458
39. Brant DA, Flory PJ (1965) The configuration of random polypeptide chains. II. Theory. J Am Chem Soc 87:2791–2800
40. Scott RA, Scheraga HA (1965) Method for calculating internal rotation barriers. J Chem Phys 42:2209
41. Scott RA, Scheraga HA (1966) Conformational analysis of macromolecules. III. Helical structures of polyglycine and poly-L-alanine. J Chem Phys 45:2091–2101
42. Momany FA, McGuire RF, Yan JF, Scheraga HA (1970) Energy parameters in polypeptides. 3. Semiempirical molecular orbital calculations for hydrogen-bonded model peptides. J Phys Chem 74:2424–2438

43. Lippincott ER, Schroeder R (1955) One-dimensional model of the hydrogen bond. J Chem Phys 23:1099
44. Schroeder R, Lippincott ER (1957) Potential function model of hydrogen bonds. II. J Phys Chem 61:921–928
45. Moulton WG, Kromhout RA (1956) Nuclear magnetic resonance: structure of the amino group. II. J Chem Phys 25:34–37
46. Chidambaram R, Balasubramanian R, Ramachandran GN (1970) Potential functions for hydrogen bond interactions I. A modified Lippincott–Schroeder potential function for NH–O interaction between peptide groups. Biochim Biophys Acta 221:182–195
47. Perutz MF et al (1960) Structure of haemoglobin—3-dimensional fourier synthesis At 5.5-Å resolution, obtained by X-ray analysis. Nature 185:416–422
48. Kendrew JC (1961) Three-dimensional structure of a protein molecule—way in which chain of amino acid units in a protein molecule is coiled and folded in space has been worked out for first time—protein is myoglobin, molecule of which contains 2,600 atoms. Sci Am 205:96
49. Lifson S (1972) In: Jaenicke R, Helmreich E (eds) Protein–proptein interactions. Springer, New York, pp 3–16
50. Lifson S (1973) Recent developments in consistent force field calulations. Dyn Asp Conform Chang Biol Macromol 421–430
51. Bixon M, Lifson S (1967) Potential functions and conformations. Cycloalkanes Tetrahedr 23:769–784
52. Warshel A, Lifson S (1969) An empirical function for second neighbor interactions and its effect on vibrational modes and other properties of cyclo- and n-alkanes. Chem Phys Lett 4:255–256
53. Warshel A, Levitt M, Lifson S (1970) Consistent force field for calculation of vibrational spectra and conformations of some amides and lactam rings. J Mol Spectrosc 33:84–99
54. Warshel A, Lifson S (1970) Consistent force field calculations. II. Crystal structures, sublimation energies, molecular and lattice vibrations, molecular conformations, and enthalpies of alkanes. J Chem Phys 53:582–594
55. Ermer O, Lifson S (1973) Consistent force field calculations. III. Vibrations, conformations, and heats of hydrogenation of nonconjugated olefins. J Am Chem Soc 95:4121–4132
56. Ypma TJ (1995) Historical development of the Newton–Raphson method. SIAM Rev 37:531–551
57. Hirschfelder JO, Curtiss CF, Bird RB (1954) Molecular theory of gases and liquids. Wiley, Hoboken
58. Hagler AT, Stern PS, Lifson S, Ariel S (1979) Urey–Bradley force field, valence force field, and ab initio study of intramolecular forces in tri-tert-butylmethane and isobutane. J Am Chem Soc 101:813–819
59. Morse PM (1929) Diatomic molecules according to the wave mechanics. II. Vibrational levels. Phys Rev 34:57–64
60. Levitt M, Lifson S (1969) Refinement of protein conformations using a macromolecular energy minimization procedure. J Mol Biol 46:269–279
61. Momany FA, Vanderkooi G, Scheraga HA (1968) Determination of intermolecular potentials from crystal data. I. General theory and application to crystallne benzene at several temperatures. Proc Natl Acad Sci USA 61:429–436
62. McGuire RF et al (1971) Determination of intermolecular potentials from crystal data. II. Crystal packing with application to poly(amino acids). Macromolecules 4:112–124
63. Momany FA, Carruthers LM, McGuire RF, Scheraga HA (1974) Intermolecular potentials from crystal data. III. Determination of empirical potentials and applications to the packing configurations and lattice energies in crystals of hydrocarbons, carboxylic acids, amines, and amides. J Phys Chem 78:1595–1620
64. Momany FA, Carruthers LM, Scheraga HA (1974) Intermolecular potentials from crystal data. IV. Application of empirical potentials to the packing configurations and lattice energies in crystals of amino acids. J Phys Chem 78:1621–1630
65. Pople J, Beveridge D (1970) Approximate molecular orbital theory. McGraw Hill, New York
66. Momany FA, McGuire RF, Burgess AW, Scheraga HA (1975) Energy parameters in polypeptides.7. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen-bond interactions, and intrinsic torsional potentials for naturally occurring amino-acids. J Phys Chem 79:2361–2381
67. Nemethy G, Pottle MS, Scheraga HA (1983) Energy parameters in peptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino acids. J Phys Chem 87:1883–1887
68. Nemethy G et al (1992) Energy parameters in polypeptldes. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with appllcatlon to proline-containing peptides. J Phys Chem 96:6472–6484
69. Allinger NL, Miller MA, Vancatledge FA, Hirsch JA (1967) Conformational analysis. LVII. The calculation of the conformational structures of hydrocarbons by the Westheimer–Hendrickson–Wiberg method. J Am Chem Soc 89:4345–4357
70. Allinger NL, Tribble MT, Miller MA, Wertz DH (1971) Conformational analysis. LXIX. An improved calculations of the structures and energies of hydrocarbons. J Am Chem Soc 93:1637–1648
71. Allinger NL (1976) Calculation of molecular structure and energy by force-field methods. Adv Phys Org Chem 13:1–82
72. Allinger NL (1977) Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. J Am Chem Soc 99:8127–8134
73. Allinger NL, Hickey MJ (1975) Conformational analysis CVIII. The calculation of the structures and energies of alkanethiols and thiaalkanes by the molecular mechanics method. J Am Chem Soc 97:5167–5177
74. Allinger NL, Hickey MJ, Kao J (1976) Conformational analysis CXI. The calculation of the structures and energies of disulfides by the molecular mechanics method. J Am Chem Soc 98:2741–2745
75. Allinger NL, Kao J (1976) Conformational analysis XCIV. Molecular mechanics studies of sulfoxides. Tetrahedron 32:529–536
76. Allinger NL, Chung DY (1976) Conformational analysis. 118. Application of the molecular-mechanics method to alcohols and ethers. J Am Chem Soc 98:6798–6803
77. Allinger NL, Chang SHM (1977) Conformational analysis—CXXIII. Tetrahedron 33:1561–1567
78. Williams DE (1966) Nonbonded potential parameters derived from crystalline aromatic hydrocarbons. J Chem Phys 45:3770–3778
79. Williams DE (1967) Nonbonded potential parameters derived from crystalline hydrocarbons. J Chem Phys 47:4680–4684
80. Kitaigorodskii AI (1965) The principle of close packing and the condition of thermodynamic stability of organic crystals. Acta Crystallogr 18:585–590
81. Kitaigorodskii AI (1968) Studies on organic chemical crystallography in USSR. Sov Phys Crystallogr USSR 12:692
82. Kitaigorodskii AI, Mirskaya KV, Tovbias AB (1968) Lattice energy of crystalline benzene in atom-atom approximation. Sov Phys Crystallogr USSR 13:176
83. Kitaigorodskii AI, Mirskaya KV (1969) Atom-atom potential method in physics of molecular crystal. Acta Cryst A 25:S91
84. Kitaigorodskii AI, Mirskaya KV (1965) Quadrupole interaction in a molecular crystal. Sov Phys Crystallogr 10:121
85. Hagler AT, Huler E, Lifson S (1974) Energy functions for peptides and proteins. I. Derivation of a consistent force field

including the hydrogen bond from amide crystals. JAm Chem Soc 96:5319–5327
86. Hagler AT, Lifson S (1974) Energy functions for peptides and proteins. II. The amide hydrogen bond and calculation of amide crystal properties. J Am Chem Soc 96:5327–5335
87. Hagler A, Lifson SA (1974) Procedure for obtaining energy parameters from crystal packing. Acta Cryst B30:1336–1341
88. Hagler ATT, Dauber P, Lifson S (1979) Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 3. The C=O···H–O hydrogen bond and the analysis of the energetics and packing of carboxylic acids. J Am Chem Soc 101:5131–5141
89. Murthy ASN, Rao CNR (1970) Recent theoretical treatments of the hydrogen bond. J Mol Struc 6:253–282
90. Pearlman Da et al (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput Phys Commun 91:1–41
91. Stockmayer WH (1941) Second virial coefficients of polar gases second virial coefficients of polar gases. J Chem Phys 9:398–402
92. Liquori AM (1960) The stereochemical code and the logic of a protein molecule. Q Rev Biophys 2:65–92
93. Berkovitch-yellin Z, Leiserowitz L (1980) The role of Coulomb forces in the crystal packing of amides. A study based on experimental electron densities. J Am Chem Soc 102:7677–7690
94. Spackman MA (1987) A Simple quantitative model of hydrogen bonding. Application to more complex systems. J Phys Chem 91:3179–3186
95. Spackman MA, Weber HP, Craven BM (1988) Energies of molecular interactions from Bragg diffraction data. J Am Chem Soc 110:775–782
96. Dinur U, Hagler AT (1992) The role of nonbond and charge flux in hydrogen bond interactions. The effect on structural changes and spectral shifts in water dimer. J Chem Phys 97:9161–9172
97. Shaik MS, Liem SY, Popelier PLA (2010) Properties of liquid water from a systematic refinement of a high-rank multipolar electrostatic potential. J Chem Phys 132:174504
98. Ren P, Wu C, Ponder JW (2011) Polarizable atomic multipole-based molecular mechanics for organic molecules. J Chem Theory Comput 7:3143–3161
99. Bakó I et al (2010) Hydrogen bonded network properties in liquid formamide. J Chem Phys 132:14506
100. Lifson S, Hagler AT, Dauber P (1979) Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 1. Carboxylic acids, amides, and the C=O···H-hydrogen bonds. J Am Chem Soc 101:5111
101. Hagler AT, Lifson S, Dauber P (1979) Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 2. A benchmark for the objective comparison of alternative force fields. J Am Chem Soc 101:5122–5130
102. Morozov AV, Kortemme T (2006) Potential functions for hydrogen bonds in protein structure prediction and design. Adv Protein Chem 72:1–38
103. McGuire RF, Momany FA, Scheraga HA (1972) Energy parameters in polypeptides. V. An empirical hydrogen bond potential function based on molecular orbital calculations. J Phys Chem 76:375–393
104. Sippl MJ, Ncmethy G, Scheraga HA (1984) Intermolecular potentials from crystal data. 6. Determination of empirical potentials for O–H···O=C hydrogen bonds from packing conflgurat-lons. J Phys Chem 7:6231–6233
105. Arnautova YA, Jagielska A, Scheraga HA (2006) A new force field (ECEPP-05) for peptides, proteins, and organic molecules. J Phys Chem B 110:5025–5044
106. Weiner SJ et al (1984) A new force-field for molecular mechanical simulation of nucleic-acids and proteins. J Am Chem Soc 106:765–784
107. Cornell WD, Kollman, PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117:5179–5197
108. Mccammon JA, Wolynes PG, Karplus M (1979) Picosecond dynamics of tyrosine side chains in proteins. Biochemistry 18:927–942
109. Brooks BR et al (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4:187–217
110. Neria E, Fischer S, Karplus M (1996) Simulation of activation free energies in molecular systems. J Chem Phys 105:1902–1921
111. Allinger NL, Yuh YH, Jenn-Huei L (1989) Molecular mechanics. The MM3 force field for hydrocarbons. J Am Chem Soc 11:8551–8566
112. Allinger NL, Chen K, Lii J (1996) An improved force field (MM4) for saturated hydrocarbons. J Comp Chem 17:642–668
113. Hagler AT (2018) The second phase of FF research introduces a fork in the road: relaxation of physics-based criteria through the introduction of correction factors… or Inclusion of more rigorous physics into the representation of molecular energetics. J Comput Aided Mol Des. https://doi.org/10.1007/s10822-018-0134-x
114. Pullman B, Pullman A (1959) The electronic structure of the purine-pyrimidine pairs of DNA. Biochim Biophys Acta 36:343–350
115. Pullman B (1969) Opening remarks: quantum-mechanical calculations of biological structures and mechanisms. Ann N Y Acad Sci 158:1–19
116. Hückel VE (1931) Quantentheoretische beitrgge zum benzolproblem. I. Die elekfronenkonfigurafion des benzols und verwandfer verbindungen. Zeitsch Phys 70:204–286
117. DelRe G, Pullman B, Yonezawa T (1963) Electronic structure of the alpha-amino acids of proteins. I. Charge distributions and proton chemical shifts. Biochim Biophys Acta 75:153–182
118. DelRe G (1958) Charge distributions in saturated organic molecules. A simple MO-LCAO method. J Chem Soc 4031–4040
119. Hoffmann R (1963) An extended hückel theory. I. Hydrocarbons. J Chem Phys 39:1397
120. Pople JA, Santry DP, Segal GA (1965) Approximate self-consistent molecular orbital theory. I. Invariant procedures. J Chem Phys 43:S129–S129
121. Hoffmann R, Imamurs A (1969) Quantum mechanical approach to the conformational analysis of macromolecules in ground and excited states. Biopolymers 7:207–213
122. Rossi A, David CW, Schor R (1969) Extended Hückel calculations on polypeptide chains. The alpha helix. Theor Chim Acta 14:429–431
123. Kier LB, George JM (1969) Extended Hiickel MO calculations of the conformation of several amino acids. Theor Chim Acta 260:258–260
124. Govil G (1970) Molecular orbital calculations on polypeptides and proteins. Part 1. Extended Huckel theory calculations on trans-dipeptides. J Chem Soc A 2464–2469
125. Govil G (1971) Molecular orbital calculations on polypeptides and proteins. Part 2 stability and conformational structure of cis-dipeptides. J Chem Soc A 386–388
126. Govil G (1971) Molecular orbital calculations on polypeptides and proteins. 3 conformation of side chains. J Indian Chem Soc 48:731
127. Govil G, Saran A (1971) Molecular orbital calculations on polypeptides and proteins. 4. Studies by EHT and CNDO on the influence of hydrogen bond and chain length on certain structures. J Chem Soc A 3624–3627

128. Govil G, Saran A (1972) Molecular orbital calculations on polypeptides and proteins 5. Stable conformations of oxygen-containing peptides. J Chem Soc Faraday Trans 68:1176–1180
129. Yan JF, Momany FA, Hoffmann R, Scheraga HA, McGuire RF (1970) Energy parameters in polypeptides 2. Semiempirical molecular orbital calculations for model peptides. J Phys Chem 74:420
130. Yan JF, Momany FA, Scheraga HA (1970) Conformational analysis of macromolecules. VI. Helical structures of 0-, rn-, and p-chlorobenzyl esters of poly-L-aspartic acid. J Am Chem Soc 92:1109–1115
131. Momany FA, McGuire RF, Yan JF, Scheraga H (1971) Energy parameters in polypeptides. IV. Semiempirical molecular orbital calculations of conformational dependence of energy and partial charge in di- and tripeptides. J Phys Chem 75:2286–2297
132. Diner S, Malrieu JP, Claverie P (1969) Localized bond orbitals and the correlation problem I. The perturbation calculation of the ground state energy. Theor Chim Acta 13:1–17
133. Maigret B, Pullman B, Dreyfus M (1970) Molecular orbital calculations on the conformation of polypeptides and proteins. 1. Preliminary investigations and simple dipeptides. J Theor Biol 26:321–333
134. Maigret M, Pullman B, Caillet J (1970) The conformational energy of an alanyl residue preceding proline: a QM approach. Biochem Biophys Res Comm 40:808–813
135. Pullman B, Pullman A (1974) Molecular orbital calculations on the conformation of amino acid residues of proteins. Adv Prot Chem 28:347–526
136. Boyd DB (2013) Pioneers of quantum chemistry: ACS symposium series, vol. 1122. American Chemical Society, Washington, DC, pp 221–273
137. Clementi E (1962) SCF-MO wave functions for the hydrogen fluoride molecule. J Chem Phys 36:33
138. Veillard A, Clementi E (1967) Complete multi. configuration self-consistent field theory. Theor Chim Acta 7:133–143
139. Clementi E (1972) Computation of large molecules with the Hartree–Fock model. Proc Natl Acad Sci USA 69:2942–2944
140. Clementi E, Davis DR (1967) Electronic structure of large molecular systems. J Comp Phys 2:223–244
141. Clementi E, Andre JM, Andre MC, Klint D, Hahn D (1969) Study of the electronic structure of molecules. X ground state for adenine, cytosine, guanine and thyamine. Acta Phys Acad Sci Hungaricae 27:439–521
142. Clementi E, Mehl J, von Niessen W (1971) Study of the electronic structure of molecules. XII. Hydrogen bridges in the guanine–cytosine pair and in the dimeric form of formic acid. J Chem Phys 54:508
143. Popkie H, Kistenmacher H, Clementi E (1973) Study of the structure of molecular complexes. IV. The Hartree–Fock potential for the water dimer and its application to the liquid state. J Chem Phys 59:1325
144. Kistenmacher H (1974) Study of the structure of molecular complexes. VI. Dimers and small clusters of water molecules in the Hartree–Fock approximation. J Chem Phys 61:546
145. Matsuoka O, Clementi E, Yoshimine M (1976) CI study of the water dimer potential surface. J Chem Phys 64:1351
146. Barker JA, Watts RO (1969) Structure of water: a Monte Carlo calculation. Chem Phys Lett 3:144–145
147. Narten AH (1972) Liquid water: atom pair correlation functions from neutron and X-ray diffraction. J Chem Phys 56:5681
148. Lie GC, Clementi E (1975) Study of the structure of molecular complexes. XII. Structure of liquid water obtained by Monte Carlo simulation with the Hartree–Fock potential corrected by inclusion of dispersion forces. J Chem Phys 62:2195
149. Bartlett RJ, Shavitt I, Purvis GD (1979) The quartic force field of H_2O determined by many-body methods that include quadruple excitation effects. J Chem Phys 71:281
150. Lie GC, Clementi E (1986) Molecular-dynamics simulation of liquid water with an ab initio fiexible water-water interaction potential. Phys Rev A 33:2679–2693
151. Corongiu G, Clementi E (1993) Molecular dynamics simulations with a flexible and polarizable potential: density of states for liquid water at different temperatures. J Chem Phys 98:4984–4990
152. Corongiu G, Clementi E (1992) Liquid water with an ab initio potential: X-ray and neutron scattering from 238 to 368 K. J Chem Phys 97:2030
153. Corongiu G, Clementi E (1993) Solvated water molecules and hydrogen-bridged networks in liquid water. J Chem Phys 98:2241–2249
154. Sciortino F, Corongiu G (1993) Structure and dynamics in hexagonal ice: a molecular dynamics simulation with an ab initio polarizable and flexible potential. J Chem Phys 98:5694
155. Clementi E, Cavallone F, Scordamaglia R (1977) Analytical potentials from 'ab Initio' computations for the Interaction between biomolecules. 1. Water with amino acids. J Am Chem Soc 99:5531–5545
156. Scordamaglia R, Cavallone F, Clementi E (1977) Analytical potentials from 'ab initio' computations for the interaction between biomolecules. 2. Water with the four bases of DNA. J Am Chem Soc 99:5545–5550
157. Clementi E, Corongiu G, Lelj F (1979) Analytical potentials from ab initio computations for the interaction between biomolecules. V. The phosphate group in nucleic acids. J Chem Phys 70:3726
158. Clementi E, Corongiu G (1979) Interaction of water with DNA single and double helix in the B conformation. Int J Quantum 16:897–915
159. Aida M, Corongiu G, Clementi E (1992) Ab initio force field for simulations of proteins and nucleic acids. Int J Quantum Chem 42:1353–1381
160. Corongiu G (1992) Molecular dynamics simulation for liquid water using a polarizable and flexible potential. Int J Quantum Chem 42:1209–1235
161. Clementi E, Corongiu G (1985) Computer simulations of complex chemical systems. Adv Biophys 20:75–107
162. Clementi E et al (1991) Selected topics in ab initio computational chemistry in both very small and very large chemical systems. Chem Rev 91:699–879
163. Pople JA (1970) Molecular orbital methods in organic chemistry. Acct Chem Res 3:217–223
164. Hehre WJ, Radom L, Schleyer PVR, Pople JA (1986) Ab initio molecular orbital theory. Wiley, Hoboken
165. Pople J (1999) Nobel lecture: quantum chemical models. Rev Mod Phys 71:1267–1274
166. Hehre WJ, Stewart RF, Pople JA (1969) Self-consistent molecular-orbital methods. I. Use of Gaussian expansions of slater-type atomic orbitals. J Chem Phys 51:2657
167. Hehre WJ, Lathan WA, Ditchfield R, Newton MD, Pople JA (1970) Gaussian 70. Quantum Chem Exch Progr 237
168. Del Bene J, Pople JA (1969) Intermolecular energies of small water polymers. Chem Phys Lett 4:426–428
169. Del Bene J, Pople JA (1970) Theory of molecular interactions. I. Molecular orbital studies of water polymers using a minimal slater-type basis. J Chem Phys 52:4858
170. Del Bene JE (1973) Theory of molecular interactions. III. A comparison of studies of H_2O polymers using different molecular-orbital basis sets. J Chem Phys 58:3605
171. Hagler AT, Leiserowitz L, Tuval M (1976) Experimental and theoretical studies of barrier to rotation about N–C-alpha and C-alpha-C′ bonds (phi and psi) in amides and peptides. J Am Chem Soc 98:4600–4612

172. Bernstein J, Hagler AT (1978) Conformational polymorphism the influence of crystal structure on molecular conformation. J Am Chem Soc 100:673–681
173. Hagler AT, Bernstein J (1978) Conformational polymorphism 2. Crystal energetics by computational substitution—further evidence for sensitivity of method. J Am Chem Soc 100:673–681
174. Bernstein J, Hagler AT (1979) Polymorphism and isomorphism as tools to study the relationship between crystal forces and molecular-conformation. Mol Cryst Liq Cryst 50:223–233
175. Lommerse JPM et al (2000) A test of crystal structure prediction of small organic molecules research papers. Acta Cryst B 56:697–714
176. Van Eijck BP (2002) Crystal structure predictions using five space groups with two independent molecules. The case of small organic acids. J Comput Chem 23:456–462
177. Dunitz JD, Gavezzotti A (2009) How molecules stick together in organic crystals: weak intermolecular interactions. Chem Soc Rev 38:2622–2633
178. Bardwell D et al (2011) Towards crystal structure prediction of complex organic compounds—a report on the fifth blind test. Acta Crystallogr B 67:535–551
179. Hornak V et al (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins-Struct Funct Bioinforms 65:712–725
180. Feig M, MacKerell AD, Brooks CL (2003) Force field influence on the observation of π-helical protein structures in molecular dynamics simulations. J Phys Chem B 107:2831–2836
181. Bochevarov AD et al (2013) Jaguar: a high-performance quantum chemistry software program with strengths in life and materials sciences. Int J Quantum Chem 113:2110–2142
182. Kitano M, Fukuyama T, Kuchitsu K (1973) Molecular structure of *N*-methylacetamide as studied by gas electron diffraction. Bull Chem Soc Jpn 46:384–387
183. Kitano M, Kuchitsu K (1974) Molecular structure of *N*-methylformamide as studied by gas electron diffraction. J Bull Chem Soc Jpn 47:631–634
184. Kitano M, Kuchitsu K (1973) Molecular structure of acetamide as studied by gas electron diffraction. J Bull Chem Soc Jpn 46:3048–3051
185. Brock CP, Ibers JA (1975) The role of crystal packing forces in the structure of pentaphenylantimony coordinate molecules. In such molecules the trigonal. Acta Cryst A 31:38–42
186. Hagler AT, Leiserowitz L (1978) Amide hydrogen-bond and anomalous packing of adipamide. J Am Chem Soc 100:5879–5887
187. Guo H, Karplus M (1992) Ab initio studies of hydrogen bonding of *N*-methyiacetamide: structure, cooperativlty, and internal rotational barriers. J Phys Chem 96:7273–7287
188. Mennucci B, Martínez JM (2005) How to model solvation of peptides? Insights from a quantum-mechanical and molecular dynamics study of *N*-methylacetamide. 1. Geometries, infrared, and ultraviolet spectra in water. J Phys Chem B 109:9818–9829
189. Villani V, Alagona G, Ghio C (1999) Ab initio studies on *N*-methylacetamide. Molec Eng B 8:135–153
190. Katz L, Post B (1960) The crystal structure and polymorphism of *N*-methylacetamide. Acta Crystallogr 13:624–628
191. Rauscher S et al (2015) Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment. J Chem Theory Comput 11:5513–5524
192. Hagler AT (1977) Relation between spatial electron-density and conformational properties of molecular systems. Isr J Chem 16:202–212
193. Hagler AT, Lapiccirella A (1976) Spatial electron distribution and population analysis of amides, carboxylic acid, and peptides and their relation to empirical poteintial functions. Biopolymers 15:1167–1200
194. Hagler AT, Lapiccirella A (1978) Basis set dependence of spatial electron-distribution—implications for calculated conformational equilibria. J Am Chem Soc 100:4026–4029
195. Nyburg SC, Faerman CH (1985) A revision of van der Waals atomic radii for molecular crystals: N, O, F, S, CI, Se, Br and I bonded to carbon. Acta Cryst B 41:274–279
196. Stone AJ, Price SL (1988) Some new ideas in the theory of intermolecular forces: anisotropic atom-atom. J Phys Chem 92:3325–3335
197. Price SL, Leslie M, Welch GWA, Habgood M, Price LS, Karamertzanis PG, Day GM (2010) Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. Phys Chem Chem Phys 12:8478–8490
198. Eramian H, Tian Y-H, Fox Z, Beneberu HZ, Kertesz M, Se S (2013) On the anisotropy of van der Waals atomic radii of O. F, Cl, Br, and I. J Phys Chem A 117:14184–14190
199. Hagler AT (2015) Quantum derivative fitting and biomolecular force fields—functional form, coupling terms, charge flux, nonbond anharmonicity and individual dihedral potentials. J Chem Theory Comput 11:5555–5572
200. Hagler AT (1977) On the relation between the spatial electron density and the conformational properties of molecular systems. Isr J Chem 16:202–212
201. Rowlinson JS (1951) The lattice energy of ice and the second virial coefficient of water vapour. Trans Faraday Soc 47:120–129
202. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) In: Pullman B (ed) Intermolecular forces. Reidel Publishing Company, Dordrecht, pp 331–342
203. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935
204. Berkovitch-Yellin Z, Leiserowitz L (1982) Atom–atom potential analysis of the packing characteristics of carboxylic acids. A study based on experimental electron density distributions. J Am Chem Soc 104:4052–4064
205. Spackman MA (2012) Charge densities and crystal engineering in modern charge-density analysis. Springer, Dordrecht
206. Mannfors B, Palmo K, Krimm S (2000) A new electrostatic model for molecular mechanics force fields. J Mol Struct 556:1–21
207. Palmo K, Mannfors B, Mirkin NG, Krimm S (2003) Potential energy functions: from consistent force fields to spectroscopically determined polarizable force fields. Biopolymers 68:383–394
208. Qian W, Krimm S (2005) Limitations of the molecular multipole expansion treatment of electrostatic interactions for C–H···O and O–H···O hydrogen bonds and application of a general charge density approach. J Phys Chem A 109:5608–5618
209. Shi Y et al (2013) Polarizable atomic multipole-based AMOEBA force field for proteins. J Chem Theory Comput 9:4046–4063
210. Weiner PK, Kollman PA, AMBER (1981) Assisted model building with energy refinement. A general program for modeling molecules and their interactions. J Comput Chem 2:287–303
211. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267:585–590
212. Engler EM, Andose JD, Schleyer PVR (1973) Critical evaluation of molecular mechanics. J Am Chem Soc 95:8005
213. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An all atom force field for simulations of proteins and nucleic acids. J Comput Chem 7:230–252
214. van Gunsteren WF, Berendsen HJC (1987) GROningen MOlecular Simulation (GROMOS) Library Manual. Biomos, Groningen
215. Hermans J, Berendsen HJC, van Gunsteren WF, Postma JPM (1984) A consistent empirical potential for water–protein interactions. Biopolymers 23:1513–1518

216. Jorgensen WL (1981) Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. J Am Chem Soc 103:335–340
217. Jorgensen WL (1982) Revised TIPS for simulations of liquid water and aqueous solutions. J Chem Phys 77:4156
218. Jorgensen WL, Swenson CJ (1985) Optimized intermolecular potential functions for amides and peptides. Structure and properties of liquid amides. J Am Chem Soc 107:569–578
219. Jorgensen WL, Tirado-Rives J (1988) The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 110:1657–1666
220. Dauber P, Osguthorpe DJ, Hagler AT, Structure (1982) Energetics and dynamics of ligand binding to dihydrofolate-reductase. Biochem 10:312–318
221. Dauber-Osguthorpe P et al (1988) Structure and energetics of ligand binding to proteins: *Escherichia coli* dihydrofolate reductase-trimethoprim, a drug-receptor system. Proteins Struct Funct Bioinform 4:31–47
222. Hagler AT, Osguthorpe DJ, Dauber-Osguthorpe P, Hempel JC (1985) Dynamics and conformational energetics of a peptide hormone: vasopressin. Science 227:1309–1315
223. Struthers RS, Rivier J, Hagler AT (1985) Molecular dynamics and minimum energy conformations of GnRH and analogs. Ann N Y Acad Sci 439:81–96
224. Struthers RS et al (1990) Design of biologically-active, conformationally constrained Gnrh antagonists. Proteins-Struct Funct Genet 8:295–304
225. Dauber-Osguthorpe P et al (1988) Structure and energetics of ligand binding to proteins: *Escherichia coli* dihydrofolate reductase-trimethoprim, a drug-receptor system. Proteins-Struct Funct Genet 4:31–47
226. Rick SW, Stuart SJ (2002) Potentials and algorithms for incorporating polarizability in computer simulations. Rev Comput Chem 18:89–146
227. Vega C, Abascal JLF (2011) Simulating water with rigid nonpolarizable models: a general perspective. Phys Chem Chem Phys 13:19663–19688
228. Wang L-P et al (2013) Systematic improvement of a classical molecular model of water. J Phys Chem B 117:9956–9972
229. Cardamone S, Hughes TJ, Popelier PLA (2014) Multipolar electrostatics. PhysChemChemPhys 16:10367–10387
230. Clark GNI, Cappa CD, Smith JD, Saykally RJ, Head-Gordon T (2010) The structure of ambient water. Mol Phys 108:1415–1433
231. Demerdash O, Yap E-H, Head-Gordon T (2014) Advanced potential energy surfaces for condensed phase simulation. Annu Rev Phys Chem 65:149–174
232. Ben-Naim A, Stillinger FH (1972) Aspects of the statistical-mechanical theory of water. Wiley, Hoboken
233. Rahman A, Stillinger FH (1971) Molecular dynamics study of liquid water. J Chem Phys 55:3336–3359
234. Alder BJ, Wainwright TE (1959) Studies in molecular dynamics. I. General method. J Chem Phys 31:459
235. Alder BJ, Wainwright TE (1960) Studies in molecular dynamics. II. Behavior of a small number of elastic spheres. J Chem Phys 33:1439
236. Stillinger FH, Rahman A (1974) Improved simulation of liquid water by molecular dynamics. J Chem Phys 60:1545–1557
237. Mahoney MW, Jorgensen WL (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. J Chem Phys 112:8910
238. Rick SW, Stuart SJ, Berne BJ (1994) Dynamical fluctuating charge force fields: application to liquid water. J Chem Phys 101:6141–6156
239. Chen B, Xing J, Siepmann JI (2000) Development of polarizable water force fields for phase equilibrium calculations. J Phys Chem B 104:2391–2401
240. Horn HW et al (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys 120:9665–9678
241. Abascal JLF, Vega C (2005) A general purpose model for the condensed phases of water: TIP4P/2005. J Chem Phys 123:234505
242. Bauer BA, Warren GL, Patel S (2009) Incorporating phase-dependent polarizability in nonadditive electrostatic models for molecular dynamics simulations of the aqueous liquid—vapor interface. J Chem Theory Comput 359–373
243. Bauer BA, Patel S (2009) Properties of water along the liquid-vapor coexistence curve via molecular dynamics simulations using the polarizable TIP4P-QDP-LJ water model. J Chem Phys 131:84709
244. Cisneros GA et al (2016) Modeling molecular interactions in water: from pairwise to many-body potential energy functions. Chem Rev 116:7501 – 7528
245. Lipkowitz KB, Allinger NL, Lipkowitz KB, Allinger NL (1987) QCPE Bull 7
246. Allinger NL, Lii J-H (1987) Benzene, aromatic rings, van der Waals molecules, and crystals of aromatic molecules in molecular mechanics (MM3). J Comput Chem 8:1146–1153
247. Lii JH, Allinger NL (1989) Molecular mechanics. The MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics. J Am Chem Soc 111:8566–8575
248. Lii JH, Allinger NL (1989) Molecular mechanics. The MM3 force field for hydrocarbons. 3. The van der Waals' potentials and crystal data for aliphatic and aromatic hydrocarbons. J Am Chem Soc 111:8576–8582
249. Ermer O (1976) Calculation of molecular properties using force fields. Applications in organic chemistry. Bond Forces Struct Bond 27:161–211
250. Lifson S, Stern PS (1982) Born–Oppenheimer energy surfaces of similar molecules: interrelations between bond lengths, bond angles, and frequencies of normal vibrations in alkanes. J Chem Phys 77:4542
251. Allinger NL, Rahman M, Lii JA (1990) Molecular mechanics force field (MM3) for alcohols and ethers. J Am Chem Soc 112:8293–8307
252. Schmitz LR, Allinger NL (1990) Molecular mechanics calculations (MM3) on aliphatic amines. J Am Chem Soc 112:8307–8315
253. Allinger NL, Quinn M, Rahman M, Chen K (1991) Molecular mechanics (MM3) calculations on sulfides. J Phys Org Chem 4:647–658
254. Chen K, Allinger NL (1991) Molecular mechanics (MM3) calculations on disulfides. J Phys Org Chem 4:659–666
255. Allinger NL, Zhu ZQS, Chen K (1992) Molecular mechanics (MM3) studies of carboxylic acids and esters. J Am Chem Soc 114:6120–6133
256. Tai JC, Yang L, Allinger NL (1993) Molecular mechanics (MM3). Calculations on nitrogen-containing aromatic heterocycles. J Am Chem Soc 115:11906–11917
257. Lii J, Allinger NL (1991) The MM3 force field for amides, polypeptides and proteins. J Comput Chem 12:186–199
258. Lii J, Allinger NL (1994) Directional hydrogen bonding in the MM3 force field. I. J Phys Org Chem 7:591–609
259. Dauber P, Hagler AT (1980) Crystal packing hydrogen bonding, and the effect of crystal forces on molecular conformation. Acct Chem Res 13:105–112
260. Nevins N, Lii JH, Allinger NL (1996) Molecular mechanics (MM4) calculations on conjugated hydrocarbons. J Comput Chem 17:695–729
261. Nevins N, Allinger NL (1996) Molecular mechanics (MM4) vibrational frequency calculations for alkenes and conjugated hydrocarbons. J Comput Chem 17:730–746

262. Langley CH, Allinger NL (2002) Molecular mechanics (MM4) calculations on amides. J Phys Chem A 106:5638–5652
263. Chen K, Lii J, Fan Y, Allinger NL (2007) Molecular mechanics (MM4) study of amines. J Comput Chem 28:2391–2412
264. Allinger NL, Chen K, Lii J, Durkin KA (2003) Alcohols, ethers, carbohydrates, and related compounds. I. The MM4 force field for simple compounds. J Comput Chem 24:1447–1472
265. Nevins N, Chen K, Allinger NL (1996) Molecular mechanics (MM4) calculations on alkenes. J Comput Chem 17:669–694
266. Langley CH, Allinger NL (2003) Molecular mechanics (MM4) and ab initio study of amide-amide and amide-water dimers. J Chem Phys A 107:5208–5216
267. Cornell WD et al (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117:5179–5197
268. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 21:1049–1074
269. Brüesch P (1966) X-ray and infrared studies of bicyclo(2.2.2) octane, triethylenediamine and quinuclidine—II. Normal coordinate calculations of bicyclo(2.2.2.)octane, triethylenediamine and quinuclidine. Spectrochim Acta 22:867–875
270. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem 97:10269–10280
271. MacKerell AD et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616
272. Gelin BR, Karplus M (1975) Sidechain torsional potentials and motion of amino acids in proteins: bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci USA 72:2002–2006
273. Maple JR, Hwang MJ, Stockfisch TP, Hagler AT (1994) Derivation of class II force fields.3. Characterization of a quantum force field for alkanes. Isr J Chem 34:195–231
274. Debiec KT et al (2016) Further along the road less traveled: AMBER ff15ipq, an original protein force field built on a self-consistent physical model. J Chem Theory Comput 12:3926 – 3947
275. Huang J et al (2017) Charmm36M: an improved force field for folded and intrinsically disordered proteins. Nat Methods 14:71–73
276. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118:11225–11236
277. Kaminski GA, Friesner RA, Tirado-rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105:6474–6487
278. Beachy MD, Chasman D, Murphy RB, Halgren TA, Friesner RA (1997) Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. J Am Chem Soc 119:5908–5920
279. Daura X, Mark AE, van Gunsteren WF (1998) Parametrization of aliphatic CH United Atoms of GROMOS96 force field. J Comp Chem 19:535–547
280. Schuler LD, Daura X, van Gunsteren WF (2001) An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. J Comp Chem 22:1205–1218
281. Halgren TA (1996) Merck molecular force field.I. Basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem 17:490–519
282. Schuler LD, van Gunsteren WF (2000) On the choice of dihedral angle potential energy functions for n-alkanes. Mol Simul 25:301–319
283. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25:1656–1676
284. Schmid N et al (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur Biophys J 40:843–856
285. Horta BAC, Fuchs PFJ, van Gunsteren WF, Hunenberger Philippe H (2011) New interaction parameters for oxygen compounds in the GROMOS force field: improved pure-liquid and solvation properties for alcohols, ethers, aldehydes, ketones, carboxylic acids and esters. J Chem Theory Comput 7:1016–1031
286. Horta BAC et al (2012) Reoptimized interaction parameters for the peptide-backbone model compound *N*-methylacetamide in the GROMOS force field: influence on the folding properties of two beta-peptides in methanol. J Comput Chem 33:1907–1917
287. Arnautova Ya, Jagielska A, Pillardy J, Scheraga H (2003) Derivation of a new force field for crystal-structure prediction using global optimization: nonbonded potential parameters for hydrocarbons and alcohols. J Phys Chem B 107:7143–7154
288. Jagielska A, Arnautova YA, Scheraga HA (2004) Derivation of a new force field for crystal-structure prediction using global optimization: nonbonded potential parameters for amines, imidazoles, amides, and carboxylic acids. J Phys Chem B 108:12181–12196
289. Pillardy J, Arnautova Y, Czaplewski C, Gibson KD, Scheraga HA (2001) Conformation-family Monte Carlo: a new method for crystal structure prediction. Proc Natl Acad Sci USA 98:12351–12356
290. Williams DE, Cox SR (1984) Nonbonded potentials for azahydrocarbons: the importance of the coulombic interaction. Acta Cryst B40:404–417
291. Berkovitch-Yellin Z (1985) Toward an ab initio derivation of crystal morphology. J Am Chem Soc 107:8239–8253
292. Mooij WTM, van Eijck BP, Price SL, Verwer P, Kroon J (1998) Crystal structure predictions for acetic acid. J Comput Chem 19:459–474
293. Hehre WJ, Pople JA (1968) Atomic electron populations for some simple molecules. Chem Phys Lett 2:379–380
294. Cox SR, Williams DE (1981) Representation of the molecular electrostatic potential by a net atomic charge model. J Comput Chem 2:304–323
295. Scrocco E, Tomasi T (1978) Electronic molecular structure, reactivity and intermolecular forces: an euristic interpretation by means of electrostatic molecular potentials. Adv Quant Chem I1:116–193
296. Besler BH, Merz KM, Kollman PA (1990) Atomic charges derived from semiempirical methods. J Comput Chem 11:431–439
297. Clementi E (1985) Ab initio computational chemistry. J Phys Chem 89:4426–4436
298. Zhou T, Huang D, Caflisch A (2010) Quantum mechanical methods for drug design. Curr Top Med Chem 10:33–45
299. Lopes PEM, Guvench O, MacKerell AD (2015) Molecular modeling of proteins. In: Kukol A (ed) Methods in molecular biology, vol. 1215. Springer, New York, pp 47–71
300. Moore GE, Fellow L (1965) Cramming more components onto integrated circuits. Electronics 114–117
301. Becke AD, Perspective (2014) Fifty years of density-functional theory in chemical physics. J Chem Phys 140:18A301
302. St-Amant A, Salahub DR (1990) New algorithm for the optimization of geometries in local density functional theory. Chem Phys Lett 169:387–392

303. Bajorath J, Kraut J, Li ZQ, Kitson DH, Hagler AT (1991) Theoretical-studies on the dihydrofolate-reductase mechanism—electronic polarization of bound substrates. Proc Natl Acad Sci USA 88:6423–6426
304. Andzelm JW, Nguyen DT, Eggenberger R, Salahub DR, Hagler AT (1995) Applications of the adiabatic connection method to conformational equilibria and reactions involving formic-acid. Comput Chem 19:145–154
305. Halgren TA (1996) Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. J Comput Chem 17:520–552
306. Halgren TA (1996) Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. J Comput Chem 17:553–586
307. Halgren TA, Nachbar RB (1996) Merck molecular force field. IV.Conformational energies and geometries for MMFF94. J Comput Chem 17:587–615
308. Halgren TA (1996) Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. J Comput Chem 17:616–641
309. Maple JR, Dinur U, Hagler AT (1988) Derivation of force fields for molecular mechanics and dynamics from ab initio energy surfaces. Proc Natl Acad Sci USA 85:5350–5354
310. Maple JR et al (1994) Derivation of class-Ii force-fields.1. Methodology and quantum force-field for the alkyl functional-group and alkane molecules. J Comput Chem 15:162–182
311. Hagler AT, Ewig CS (1994) On the use of quantum energy surfaces in the derivation of molecular-force fields. Comput Phys Commun 84:131–155
312. Dinur U, Hagler AT (1991) Lipkowitz KB, Boyd DB (eds) Reviews in computational chemistry, vol. 15. VCH, New York, pp 99–164
313. *Probe Manual* (1989) Biosym Technologies, Inc
314. Ewig CS et al (2001) Derivation of class II force fields. VIII. Derivation of a general quantum mechanical force field for organic compounds. J Comput Chem 22:1782–1800
315. Palca J (1986) Computer models: cooperation on new molecules. Nature 322:586
316. Halgren TA (1992) The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. J Am Chem Soc 114:7827–7843
317. Ewig CS, Thacher TS, Hagler AT (1999) Derivation of class II force fields. 7. Nonbonded force field parameters for organic compounds. J Phys Chem B 103:6998–7014
318. Halgren TA, MMFF VII (1999) Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular interaction energies and geometries. J Comput Chem 20:730–748
319. Bordner AJ, Cavasotto CN, Abagyan RA (2003) Direct derivation of van der Waals force field parameters from quantum mechanical interaction energies. J Phys Chem B 107:9601–9609
320. Hagler AT, Maple JR, Thacher TS, Fitzgerald GB, Dinur U Weiner PK (1989) Potential energy functions for organic and biomolecular systems. Comput Simul Biomol Syst ESCOM Leiden 1:149–167
321. Hwang MJ, Stockfisch TP, Hagler AT (1994) Derivation of class II force fields. 2. Derivation and characterization of a class-II force field, CFF93, for the alkyl functional group and alkane molecules. J Am Chem Soc 116:2515–2525
322. Ercolessi F, Adams JB (1994) Interatomic potentials from first-principles calculations: the force-matching method. Eur Lett 26:583–588
323. Dasgupta S, Goddard WA (1989) Hessian-biased force fields from combining theory and experiment. J Chem Phys 90:7207–7215
324. Maple JR, Hwang MJ, Jalkanen KJ, Stockfisch TP, Hagler AT (1998) Derivation of class II force fields: V. Quantum force field for amides, peptides, and related compounds. J Comput Chem 19:430–458
325. Palmo K, Mirkin NG, Pietila: L, Krimm S (1993) Spectroscopically determined force fields for macromolecules. 1 n-alkane chains. Macromolecules 26:6831–6840
326. Stouch TR (2012) The errors of our ways: taking account of error in computer-aided drug design to build confidence intervals for our next 25 years. J Comput Aided Mol Des 26:125–134
327. Ermer O, Lex J (1987) Shortened C–C, bondsand antiplanar O=C–0–H torsion angles in 1,4-cubanedicarboxyiic acid. Anyen Chem Int Ed Engl 41:447–449
328. Wilson E, Decius J, Cross PC (1980) Molecular vibrations. Dover, New York
329. Hwang MJ, Ni X, Waldman M, Ewig CS, Hagler AT (1998) Derivation of class II force fields. VI. Carbohydrate compounds and anomeric effects. Biopolymers 45:435–468
330. Chen K-H, Allinger NL (2002) Molecular mechanics (MM4) study of saturated four-membered ring hydrocarbons. J Mol Struct Theochem 581:215–237
331. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG (2012) ZINC: a free tool to discover chemistry for biology. J Chem Inf Model 52:1757–1768
332. MDDR Accelrys. http://accelrys.com/products/collaborative-science/databases/bioactivity-databases/mddr.html
333. Sun H, Mumby SJ, Maple JR, Hagler AT (1994) An ab initio Cff93 all-atom force field for polycarbonates. J Am Chem Soc 116:2978–2987
334. Sun H, Rigby D (1997) Polysiloxanes: ab initio force field and structural, conformational and thermophysical properties title. Spectrochim Acta A 53:1301–1323
335. Hill JR, Sauer J (1994) Molecular mechanics potential for silica and zeolite catalysts based on ab initio calculations. 1. Dense and microporous silica. J Phys Chem 98:1238–1244
336. Zhu W et al (2009) Molecular dynamics simulations of AP/HMX composite with a modified force field. J Hazard Mater 167:810–816
337. McQuaid MJ, Sun H, Rigby D (2004) Development and validation of COMPASS force field parameters for molecules with aliphatic azide chains. J Comput Chem 25:61–71
338. Sun HCOMPASS (1998) An ab initio force-field optimized for condensed-phase applications—overview with details on alkane and benzene compounds. J Phys Chem B 102:7338–7364
339. Dinur U, Hagler AT (1989) Direct evaluation of nonbonding interactions from ab initio calculations. J Am Chem Soc 111:5149–5151
340. Dinur U, Hagler AT (1989) Determination of atomic point charges and point dipoles from the Cartesian derivatives of the molecular dipole moment and second moments, and from energy second derivatives of planar dimers II. Applications to model systems. J Chem Phys 91:2959–2970
341. Dinur U, Hagler AT (1995) Geometry-dependent atomic charges—methodology and application to alkanes, aldehydes, ketones, and amides. J Comput Chem 16:154–170
342. Galimberti D, Milani A, Castiglioni C (2013) Charge mobility in molecules: charge fluxes from second derivatives of the molecular dipole. J Chem Phys 138:164115
343. Cruz-Cabeza AJ, Bernstein J (2014) Conformational polymorphism. Chem Rev 114:2170–2191
344. Dauber P, Hagler AT, Crystal, Packing (1980) Hydrogen-bonding, and the effect of crystal forces on molecular-conformation. Acc Chem Res 13:105–112
345. Liang CX, Ewig CS, Stouch TR, Hagler AT (1993) Abinitio studies of lipid model species.1. Dimethyl-phosphate and methyl propyl phosphate anions. J Am Chem Soc 115:1537–1545

346. Liang CX, Ewig CS, Stouch TR, Hagler AT (1994) Ab initio studies of lipid model species. 2. Conformational-analysis of inositols. J Am Chem Soc 116:3904–3911
347. Liang C et al (1995) Force field studies of cholesterol and cholesteryl acetate crystals and cholesterol–cholesterol intermolecular interactions. J Comp Chem 16:883–897
348. Lorentz HA (1881) Ueber die anwendung des satzes vom virial in der kinetischen theorie der gase. Ann Phys 248:127–136
349. Berthelot D (1898) Sur le mélange des gaz. Comptes Rendus Hebd Séances l'Acad Sci 126:1703–1855
350. Diaz Peña M, Pando C, Renuncio JA (1982) R. Combination rules for intermolecular potential parameters. I. Rules based on approximations for the long-range dispersion energy. J Chem Phys 76:325
351. Tang KT, Toennies JP (1986) New combining rules for well parameters and shapes of the van der Waals potential of mixed rare gas systems. Z Phys D 1:91–101
352. Kestin J et al (1984) Equilibrium and transport properties of the noble gases and their mixtures at low density. J Phys Chem Ref Data 13:229–303
353. Al-Matar AK, Rockstraw DW (2004) A generating equation for mixing rules and two new mixing rules for interatomic potential energy parameters. J Comp Chem 25:660–668
354. Desgranges C, Delhommelle J (2014) Evaluation of the grand-canonical partition function using expanded Wang-Landau simulations. III. Impact of combining rules on mixtures properties. J Chem Phys 140:104109
355. Song W, Rossky PJ, Maroncelli M (2003) Modeling alkane + perfluoroalkane interactions using all-atom potentials: failure of the usual combining rules. J Chem Phys 119:9145
356. Reinhold J et al (2014) Molecular dynamics simulations on scattering of single Ar, N_2, and CO_2 molecules on realistic surfaces. Comput Fluids 97:31–39
357. Nemethy G, Scheraga HA (1965) Theoretical determination of sterically allowed conformations of a polypeptide chain by a computer method. Biopolymers 3:155–184
358. Allinger NL (2011) Understanding molecular structure from molecular mechanics. J Comput Aided Mol Des 25:295–316
359. Clementi E, Ranghino G, Scordamaglia R (1977) Intermolecular pontentials: interaction of water with lysozyme. Chem Phys Lett 49:218–224
360. Dinur U, Hagler AT (1990) A novel decomposition of torsional potentials into pairwise interactions—a study of energy 2nd derivatives. J Comput Chem 11:1234–1246
361. Dinur U, Hagler AT (1989) Determination of atomic point charges and point dipoles from the cartesian derivatives of the molecular dipole-moment and 2nd moments, and from energy 2nd derivatives of planar dimers. I. Theory. J Chem Phys 91:2949–2958
362. Ooi T, Scott RA, Vanderkooi G, Scheraga HA (1967) Conformational analysis of macromolecules. IV. Helical structures of poly-L-alanine, poly-L-valine, poly-β-methyl-L-aspartate, poly-γ-methyl-L-glutamate, and poly-L-tyrosine. J Chem Phys 46:4410–4426

Affiliations

Pnina Dauber-Osguthorpe[2] · A. T. Hagler[1,3]

[1] Department of Chemistry, University of Massachusetts, Amherst, MA 01003, USA

[2] Art Pearl Studios, Newlyn, Cornwall, UK

[3] Present Address: Valis Pharma, San Diego, CA, USA