The Consistent Force Field. 1. Methods and Strategies for Optimization of Empirical Potential Energy Functions

Søren Balling Engelsen, Jesper Fabricius and Kjeld Rasmussen[†]

Chemistry Department A, The Technical University of Denmark, DK-2800 Lyngby, Denmark

Engelsen, S. B., Fabricius, J. and Rasmussen, Kj., 1994. The Consistent Force Field. 1. Methods and Strategies for Optimization of Empirical Potential Energy Function. – Acta Chem. Scand. 48: 548–552 © Acta Chemica Scandinavica 1994.

A systematic approach to the development of empirical potential energy functions using exclusively high precision experimental data is outlined. This paper, the first in a series, describes our qualitative choice of potential energy functions intended for use in glycolipid research. We describe a conservative empirical approach to optimize the chosen potential energy functions, avoiding conflicts with molecular structure definitions and avoiding redundant experimental data. A method for converting molecular structures obtained from electron diffraction (r_a and r_g) into the ground vibrational state structure at absolute zero, r_z, is reported.

We have developed a methodology[1,2] for quick and interactive evaluation of the performance of a set of potential energy functions, which are the kernel of molecular mechanics.[3] This paper describes our methods and strategies of parametrization; the succeeding paper[4] represents our first parametrization effort using the described approach to obtain an optimized force field for the functional groups methine, methylene and methyl. We use the Consistent Force Field, originally conceived by Lifson and Warshel,[5] and documented and used extensively by this group.[6,7] The uniqueness of this method is that it allows for systematic least squares refinement of force field parameters on experimental data.

The final result of any effort to develop a new set of empirical potential energy functions (a molecular mechanics force field) depends on the pathway or strategy used in the fitting process. Apparently the parameter 'hypersurface' is very complex and contains many minima each of which is a region of partially optimal solutions, and for this reason it is important to collect and fit as many different data as possible in order to restrict parameter hyperspace. The two most important elements of any optimization strategy are (1) data collection and selection, and (2) data reproduction. We describe an approach utilizing only non-redundant high precision experimental data.

The methods and strategies described here were devised with the ultimate purpose of developing a molecular mechanics force field for simulation of glycolipids using molecular statics as well as molecular dynamics. It is considered of major importance to create a force field which is able to simulate lipids in aqueous environment and in

gel phases, as well as in the gaseous phase, where applicable. In the parametrization strategy we have avoided implicit environment variables during the optimization process. Previous parametrizations carried out in this group[6–9] were fitted or optimized on data measured in crystal phase or in solution, with no explicit account of intermolecular interactions.

Methods

Program. Development of new parameter sets is to be made with the Lyngby version[1,6,7] of the Consistent Force Field (CFF) program.[5] CFF is a molecular mechanics program designed for development of molecular potential energy functions, due to its integrated least squares optimization algorithm. The algorithm used to optimize the potential energy function parameters and the recently implemented interactive graphical application to guide the progress of the optimization, are described elsewhere.[1,2,6]

In CFF, energy minimization is mostly performed using the variable metric method published by Dennis and Moré.[10] This method implements BFGS updating of the Hessian matrix which is recommended by Stewart[11] and employs soft line search due to Powell.[12] Minimization is continued until the 2-norm of the gradient vector is less than 0.001 J mol^{-1} pm^{-1}.

Crystal simulation[13] is performed using Williams' variant[14–16] of the Ewald lattice summation method with a spherical and abrupt cutoff limit of 0.8 nm.

Potential energy functions. Since the final performance of any force field will depend on the functional form of the potential energy functions as well as their parametriza-

---

[†] To whom correspondence should be addressed.

tion, the choice of the functional form must be considered carefully. As a pragmatic way of thinking, the potential energy functions are divided into three categories: bonded energies, non-bonded energies, and correction terms, respectively. Our choice of functional form in each of the three categories is described below.

An important consideration concerns the number of parameters. While the addition of more terms and therefore more parameters improves performance, this approach may lead to an excessive number of parameters as the number of atom types is increased. Our work consistently aims at producing small parameter sets with good performance by a judicious choice of functions. An earlier example of a simple PEF applied to the overcrowded tris(tert-butyl)methane[17] illustrates this principle.

Bonded energies[2,7] The Morse potential [eqn. (1)] is chosen to simulate bonded energies, giving a theoretical possibility to approximate the true bonded potential in a larger region around the equilibrium than by using the harmonic potential.

$$V_b = \sum_{all\ bonds} D_e (e^{-2\alpha(b - b_0)} - 2e^{-\alpha(b - b_0)}) \qquad (1)$$

The explicit inclusion of anharmonicity by the Morse potential is expected to improve calculation of normal modes in the harmonic limit by the added functional anharmonicity. Lifson and Stern[18] and more recently Dillen[19] have successfully developed molecular mechanics force fields for alkanes using the Morse potential.

Non-bonded energies. The non-bonded potential consists of a van der Waals and a Coulomb contribution. The van der Waals contribution is a combination of a repulsive term accounting for interatomic repulsions and the Pauli exclusion principle and an attractive term accounting for the dispersion attraction (the London potential). The Coulomb contribution represents electrostatic interactions.

The non-bonded potentials are in principle multipole potentials which are assumed to work on both inter- and intramolecular interactions although derived mainly from intermolecular experiments. Multipole potentials are unsuitable for molecular mechanics calculations; as an efficient compromise we use effective pair potentials which implicitly include average multibody effects. The drawback is that the effective pair potential needed to reproduce experimental data may turn out to depend on temperature, density, etc. while the true potential does not. For the van der Waals potential we employ the computationally efficient Lennard-Jones 12–6 term [eqn. (2)].

$$V_{nb} = \sum_{i<j} [A_{ij} r^{-12} - B_{ij} r^{-6} + e_i e_j (Dr)^{-1}] \qquad (2)$$

The Lennard-Jones plus Coulomb 12–6–1 potential energy function is parametrized using one-atom parameters,[13,16] which cuts the number of parameters significantly, especially as more types of atoms are introduced.

For electrostatic interactions the simple Coulomb term is used. It models the interaction between electron clouds by a superposition of monopole electrostatic interactions.[8] This crude approximation placing the partial atomic charges at the center of the nuclei and using a global dielectric constant for the intervening medium probably represents the strongest deviation from Nature in molecular mechanics; however, more refined potential energy functions in this respect, such as charge-flux or topology-dependent Coulomb potentials, are very cumbersome to code and very expensive to evaluate. It should be mentioned that the bond-dipole approach to the electrostatic interactions has been proven to work equally well,[20] although not superior to the fractional charge approach.

In the CFF program the fractional charges are assigned from their parameter values by a charge redistribution algorithm;[1] therefore the assigned fractional charges are not identical to the values of the charge parameters. The advantages of this approach are that (1) from one charge parameter per atom, we may obtain fractional charges on the atoms in different contexts such that reasonable *ab initio* charge distributions are mimicked, and (2) the molecules are neutralized unless otherwise specified.

The use of a global dielectric constant D is worth a comment. Atoms interacting through free space would experience a dielectric constant of 1. When they interact across a molecule, the dielectric constant of a 'normal' medium would apply, such as 80 for water. A pragmatic solution is to use a value in the range 2–3.5. Too small a value causes the electrostatic interactions to dominate and too high a value renders them superfluous. A theoretical derivation[21] of the effective dielectric constant for the interaction between two charged particles in a spherical cavity lends support to this pragmatic method. The use of an arbitrarily chosen value of D will be reflected implicitly in the numerical values of the partial charges when optimized on crystal properties.

In order not to introduce further complications, intramolecular scaling factors are not included, and non-bonded interactions therefore contribute fully for 1–4 and higher intramolecular interactions (more than two bonds away). Since we see no theoretical reason for the exclusion of geminal interactions, we have conducted some calculations which included geminal interactions in full, but they all turned out to give results of poorer quality than for the potential which excluded geminal interactions. Test molecules 'blew up', and the size could not be reduced by reasonable values of bond energy parameters. The reason for this is not fully investigated, but has to do with an overestimated 1–3 repulsion originating from the combined valence angle potential and the non-bonded potential.

Correction energies. The correction potentials are indispensable in the present state of molecular mechanics calculations. They may turn out to be superfluous if better bonded and especially non-bonded potentials are

developed. We find it necessary to include a harmonic valence angle potential [eqn. (3)] and a Pitzer torsional (dihedral) angle potential [eqn. (4)].

$$V_\theta = \sum_{\text{all angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \tag{3}$$

$$V_\phi = \sum_{\text{all torsions}} \frac{1}{2} K_\phi (1 + \cos k\phi) \tag{4}$$

For reasons of simplicity and efficiency the correction potentials have all been designed to express harmonic behavior around a minimum energy conformation. The valence angle term is strongly needed in order to maintain the structural rigidity, presumably mainly because of the lack of proper representation of geminal non-bonded intramolecular interactions. The torsional potential is needed, since intramolecular 1–4 van der Waals interactions strong enough to reproduce hindered rotation around sp^3–sp^3 bonds create nonphysical behavior in condensed phase simulations.

Two different systematic definitions of torsional angles can be used: group torsion and bond torsion.[7,22] The group torsion approach defines only one torsional angle per bond, whereas the bond torsion approach defines e.g. nine torsions about a $C(sp^3)$–$C(sp^3)$ bond. The group torsion model requires much fewer parameters than the bond torsion model, but we find that it presents problems of definitions which introduce numerical problems when the torsional energy has a non-negligible value. On the other hand, the bond torsional approach causes a highly correlated behavior of individual torsional terms which may create problems in the optimization process.

Cross-term potentials. Since our intention with the potential is to create a 'simple' force field for structural research on glycolipids and not to make a vibrational force field intended for use in spectroscopy, no cross terms are implemented. This means that it is accepted, at the outset, that vibrational frequencies, in particular in the interval 1600–500 cm^{-1}, may be erroneously reproduced on an individual basis.

Optimization strategy

Principles. Once the functional forms of the different energy contributions are selected, it is important to decide the strategy of parametrization bearing in mind the intended future applications. As previously mentioned the ambition is to obtain a 'vacuum like' force field applicable to both gas phase and condensed phase simulations. The surrounding environment in a molecular mechanics calculation can be either implicitly incorporated on an average basis in the parameter set, which usually is called potential of mean force, or it can be calculated explicitly as a 'physically' present environment. Either choice will affect the procedure of parametrization. If a potential of mean force is wanted, the potential energy functions are

optimized to reproduce experimental data for substances in the environment of choice. Similarly, if an explicit environment potential is wanted, the potential energy functions are made to reproduce experimental data for substances in the explicit environment. In the latter case the result is a 'vacuum like' potential, to which symmetry operations must be applied, in order to simulate condensed phases. The environment of choice usually represents a crystalline environment or a water model which is developed separately and represents a self-consistent model, able to reproduce many thermodynamic data. The complexity of the development of a new water model has resulted in a strong dependency on the few successful water potentials like Stillinger and Rahmans ST2-potential[23] and Jorgensens TIPS-potential.[24,25] With the purpose of developing a force field for glycolipids it is desired to approximate an explicit environment potential as closely as possible by adjusting non-bonded parameters to condensed phase observables, i.e. X-ray diffraction and neutron diffraction.

Initial parameter values. Using the selected potential energy functions, the parameter vector to be optimized can be symbolically written as shown in eqn. (5).

$$p = (D_e^1, \alpha^1, b_0^1, D_e^2, \alpha^2, b_0^2, \ldots, D_e^n, \alpha^n, b_0^n, \tag{5}$$

$$K_\theta^1, \theta_0^1, K_\theta^2, \theta_0^2, \ldots, K_\theta^m, \theta_0^m,$$

$$K_\phi^1, K_\phi^2, \ldots, K_\phi^k,$$

$$A^1, B^1, A^2, B^2, \ldots, A^s, B^s,$$

$$e^1, e^2, \ldots, e^s)$$

Here, n runs over all types of X–X 'pair' bonds, m runs over all types X–X–X of valence angles, k runs over all types of X–X–X–X torsional angles and s runs over all atoms involved in the simulation. Note that $s \leq n < m < < k$ owing to the large number of permutations of the torsional arrangement X1–X2–X3–X4 in the bond torsional approach.

To initiate an optimization process one has to provide the optimization algorithm with a realistic initial parameter vector. D_e can be interpreted as the bond dissociation energy, K_θ and K_ϕ are interpreted as force constants and b_0, θ_0 and ϕ_0 as equilibrium values for bond lengths, bond angles and torsional angles. The A_i and A_j, and B_i and B_j, of the van der Waals potential can be calculated from Lennard-Jones diameters or non-bonded equilibrium distances and the dispersion well depth. The e_i of the Coulomb potential can be obtained from *ab initio* calculations. Finally a critical examination of previously published molecular mechanics force fields can serve as a guide in the selection of initial parameters.

Choice of experimental data. Selection of experimental data is not straightforward if consistency in force field performance is desired. The geometry of a molecule in a hypothetical, motionless state at the bottom of the potential well (as optimized using *ab initio* methods) is

different from the geometry obtained by diffraction and spectroscopic methods. There are even significant differences between observables (not only hydrogen positions) measured by neutron diffraction and by X-ray diffraction at the same temperature.

Dipole moments and normal modes are readily available from microwave and from infrared and Raman experiments, respectively, but for molecular structure the experimental discipline of choice is not obvious. Electron diffraction, X-ray diffraction and neutron diffraction all give high precision measurements of molecular geometry. They do not give the same results because they use different definitions of molecular geometry. The definition of molecular geometry and interatomic distances is not simple, since it depends on temperature and structural model. The structures and their relationships have been thoroughly examined by Kuchitsu and Cyvin.[26,27]

The most desirable structure for use in deriving a potential energy function is r_e, defined as the distances between the atomic nuclei in the hypothetical vibrationless state. As this structure is very difficult to derive, and because only a few are available, the r_z structure, defined as the average nuclear distances in the vibrational ground state at absolute zero, is preferable. Publications of electron diffraction experiments usually report either the r_a or the r_g structure, and only in rare cases the r_z structure, which therefore in most cases must be calculated from the reported data.

Systematic inconsistencies in reproducing molecular geometry, as seen with previously published force fields, including some developed by this group,[8,28] were related to the use of data from different molecular models. For this reason we now convert all structural data found by electron diffraction into r_z structures before we use them in an optimization process. It is done using two approximations published by Kuchitsu and Cyvin.[29]

$$r_z \approx r_a + \frac{\langle \Delta z^2 \rangle_T}{r_e} - \frac{3}{2} a_3 (\langle \Delta z^2 \rangle_T - \langle \Delta z^2 \rangle_0)$$

$$- \frac{\langle \Delta x^2 \rangle_0 + \langle \Delta y^2 \rangle_0}{2 r_e} \qquad (6)$$

$$r_z \approx r_g - \frac{3}{2} a_3 (\langle \Delta z^2 \rangle_T - \langle \Delta z^2 \rangle_0)$$

$$- \frac{\langle \Delta x^2 \rangle_0 + \langle \Delta y^2 \rangle_0}{2 r_e} \qquad (7)$$

where $\langle \Delta z^2 \rangle_T$ and $\langle \Delta z^2 \rangle_0$ are the mean-square amplitudes parallel to an imaginary axis determined by the equilibrium positions of two arbitrary atoms. $\langle \Delta x^2 \rangle$ and $\langle \Delta y^2 \rangle$ are the mean-square amplitudes perpendicular to this axis. The scalar a_3 is a constant, approximately 20 nm^{-1}, analogous to the Morse parameter for diatomic molecules.[7] In theory, r_e must be used in the denominators, but r_a or r_g can be used in practice. The amplitudes are also designated generalized mean-square amplitudes.[29]

Actual procedure. In the course of exploratory work the following procedure has proved satisfactory. Initially, sets of non-bonded Lennard-Jones parameters over a broad range are subjected to optimization, while all other parameters are fixed. Only very small improvements occurred in the force field performance during these initial optimizations. It was apparent that the magnitudes of the non-bonded parameters are sensitive only to optimization using crystal data, and an iterative scheme was adopted which is shown in Fig. 1. This optimization scheme, when applied in its full extent, gives a multi-level iterative optimization process.

Conclusion

In this work we have discussed a qualitative choice of analytical forms for empirical potential energy functions with the explicit aim of describing glycolipids. We further described a systematic approach for optimizing a set of parameters for the functions using exclusively high precision structural and vibrational data to restrict parameter space. With the described approach we have developed a set of parameters for a new combination of potential energy functions for modelling the alkanes; this work is described in the succeeding paper.[4] It represents a first and probably most essential step towards the de-
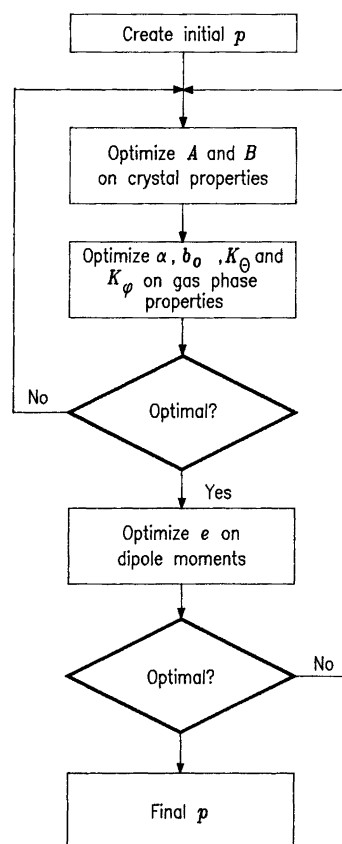

Fig. 1. Optimization strategy.

velopment of a more general molecular mechanics force field for modelling the multifunctional structures of glycolipids. Extension to other classes of compounds, such as ethers, alcohols, saccharides, ketones, aldehydes, esters, amines, amides and phosphates is under way.[30-32]

References

1. Rasmussen, Kj., Engelsen, S. B., Fabricius, J. and Rasmussen, B. In: Fausto, R., Ed., *Recent Experimental and Computational Advances in Molecular Spectroscopy* (Proceedings of the NATO Advanced Study Institute on Molecular Spectroscopy, Ponta Delgada, São Miguel, Açores, Portugal 1992), NATO ASI Ser. C: Mathematical and Physical Sciences, Kluwer Academic Publishers, Dordrecht 1993, Vol. 406, pp. 381–419.
2. Engelsen, S. B., Fabricius, J. and Rasmussen, Kj. *Computers and Chemistry. In press.*
3. Burkert, U. and Allinger, N. L. *Molecular Mechanics*, ACS Monograph, Vol. 177, 1982.
4. Engelsen, S. B., Fabricius, J. and Rasmussen, Kj. *Acta Chem. Scand. 48* (1994) 553.
5. Lifson, S. and Warshel, A. *J. Chem. Phys. 49* (1968) 5116.
6. Rasmussen, Kj. *Potential Energy Functions in Conformational Analysis*, Lecture Notes in Chemistry, Springer-Verlag, Berlin 1985, Vol. 37.
7. Niketić, S. R. and Rasmussen, Kj. *The Consistent Force Field: A Documentation*, Lecture Notes in Chemistry, Springer-Verlag, Berlin 1977, Vol. 3.
8. Melberg, S. and Rasmussen, Kj. *J. Mol. Struct. 57* (1979) 215.
9. Rasmussen, Kj. and Fabricius, J. In: French, A. D. and Brady, J. W., Eds., *Computer Modelling of Carbohydrate Molecules.* ACS Symposium Series No. 430, The American Chemical Society, Washington, DC 1990.
10. Dennis, J. E. and Moré, J. J. *SIAM Rev. 19* (1977) 46.
11. (a) Stewart, J. J. P. *Comput. Chem. 13* (1989) 157; (b) Engelsen, S. B. *Unpublished.*
12. Powell, M. J. D. *Report C.S.S.*, AERE Harwell, Oxfordshire, UK 1975, Vol. 15.
13. Pietilä, L.-O. and Rasmussen, Kj. *J. Comput. Chem. 5* (1984) 252.
14. Williams, D. E. *Acta Crystallogr., Sect. A 27* (1971) 452.
15. Williams, D. E. *Acta Crystallogr., Sect. A 28* (1972) 629.
16. Williams, D. E. In: Metzger, R. M., Ed., *Crystal Cohesion and Conformational Energies*, Springer-Verlag, Berlin 1981, p. 3.
17. Rasmussen, Kj. *J. Mol. Struct. 72* (1981) 171.
18. Lifson, S. and Stern, P. S. *J. Chem. Phys. 77* (1982) 4542.
19. Dillen, J. L. M. *J. Comput. Chem. 11* (1990) 1125.
20. Allinger, N. L. *J. Am. Chem. Soc. 99* (1977) 8127.
21. Ramachandran, G. N. and Srinivasan, R. *Ind. J. Biochem. 7* (1970) 95.
22. Gleicher, G. J. and Schleyer, P. v. R. *J. Am. Chem. Soc. 89* (1967) 582.
23. Stillinger, F. H. and Rahman, A. *J. Chem. Phys. 60* (1974) 1545.
24. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. and Klein, M. L. *J. Chem. Phys. 79* (1983) 926.
25. Jorgensen, W. L. *J. Am. Chem. Soc. 103* (1981) 335.
26. Kuchitsu, K. In: Lide, D. R. and Paul, M. A., Eds., *Critical Evaluation of Chemical and Physical Structural Information*, Natl. Acad. Sci., Washington, DC 1974, pp. 132–139.
27. Kuchitsu, K. and Cyvin, S. J. In: Cyvin, S. J., Ed., *Molecular Structures and Vibrations*, Elsevier, Amsterdam 1972, pp. 183–211.
28. Rasmussen, Kj. *Acta Chem. Scand., Ser. A 36* (1982) 323.
29. Cyvin, S. J. *Molecular Vibrations and Mean Square Amplitudes*, Universitetsforlaget, Oslo, and Elsevier, Amsterdam 1968.
30. Engelsen, S. B. *Ph.D. Thesis*, Technical University of Denmark, Lyngby 1991.
31. Fabricius, J. *Ph.D. Thesis*, Technical University of Denmark, Lyngby 1991.
32. Jónsdottir, S. Ó. and Rasmussen, Kj. *To be published.*