An Informal AMBER Small Molecule Force Field : parm@Frosst

Credit
Christopher Bayly (1992-2010) initiated, contributed and lead the efforts
Daniel McKay (1997-2010) and Jean-François Truchon (2002-2010) were main contributors

Abstract

This note presents a molecular mechanics force field (FF) extending the AMBER FF to bioorganic small molecules of pharmaceutical interest. The presented parm@Frosst FF enables the simulation of biomolecules (enzymes, DNA, peptides, etc.) in the presence of complex organic molecules such as inhibitor and cofactors. As such it can be used as a small-molecule supplement to the AMBER parm9*x* or ff*xx* biomolecular force fields, as an alternative to e.g. gaff. The development took place at Merck Frosst Canada, a subsidiary of Merck & Co, between 1992 and 2010 in the context of numerous drug-discovery projects. As a result, parm@Frosst, when used to extend one of the "standard" AMBER foce fields such as ff99sb, could successfully parameterize approximately 85% of the Merck corporate collection (of small molecules) in 2009 (personal communication to CIB from V. Hornak). The information you will find is 1) an historical overview of the FF development and technical notes 2) the atom typing definitions 3) the parameters in the Amber format and 4) an implementation validation set. This information is sufficient to reproduce the force field in any software package that implements the AMBER energy function.

Historical Overview

The motivation to develop parm@Frosst was the desire to carry out MD simulations on pharmaceutical-type small-molecule ligands within solvated ligand-protein complexes as part of application modeling in lead optimization projects. At that time (the early 90's) the most current parm9*x* force fields had only a very limited capability to handle small molecules. Even in cases which could be accommodated from a formal standpoint, frequently the parameterizations were unsuitable, for example the single-bond linkage in a biphenyl would hold the two phenyls stiffly planar instead of having a shallow minimum around 40 degrees out-of-plane. As a first step towards extending the force field, new chemotypes and atom types were needed and their application was made possible due to the Merck proprietary chemical perceiver, a descendent of PATTY [Bruce Bush and Robert Sheridan, *J. Chem. Inf. Comput. Sci.,* **33**, (1993), 756-762]. The authors of this group effort were involved in application modeling and extended parm@Frosst as new modeling ideas or new compounds were made. This work spanned a total

of 44 full time employee years. This *ad hoc* approach to extending the FF was initially slow requiring frequent improvement and parameterization of mainly torsion parameters. However, over the 18 years of development, specific parameterization became increasingly infrequent as the FF became more complete. A consequence of this continuous improvement (and the pressures of drug discovery in pharma) is that no global optimization of parameters for small molecules was ever conducted except occasionally for a small number of molecules sharing the same functional group. Different levels of quantum theory were also used over time, a reflection of the advance in quantum theory "best practices" over the years. Furthermore, the molecules used in the parameterization cannot be disclosed and were not accumulated. Our quality insurance was the continuous validation of the FF on multiple projects where explaining and predicting experimental results was the chief of our concern. This occasionally led to a correction of parameters or, more often, to the need to split an atom type into two to separate distinct torsion behaviours. Thus parm@Frosst does not represent a "finished product" but rather an evolved state which responded to the challenges and rigors of many years of use in prospective modeling in pharmaceutical drug discovery. It cannot handle all bioorganic chemistry; it cannot even parameterize all the small molecules in Merck's corporate collection (only ~85% as of 2009). Nevertheless, even by today's standard where a number of parallel FF efforts have been implemented, we claim that parm@Frosst remains a unique asset in terms of its generality and accuracy (based on our own internal studies). The evolution needs to continue: parm@Frosst needs both further parameterization, towards both completeness and improved accuracy, and better validation. This will require time, resources and a diverse set of specialized knowledge. We hope that this public release will motivate a more open and participative effort that should benefit the whole molecular modeling community increasingly as it progresses.

Technical Notes

parm@Frosst elaboration started as an extension of the AMBER parm94 force field [D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules", *J. Am. Chem. Soc.*, **117**, (1995), 5179-5197. Correction: *J. Am. Chem. Soc.*, **118**, (1996), 2309]. As a consequence, one can find many inherited practices. The AMBER energy function required the determination of several parameters: the atomic partial charges, the Fourier parameters related to the torsion terms, the bond angle and associated bending force constant, and the bond length and stretching force constant. No Lennard-Jones radii or well depths were modified or added, keeping the AMBER values and types. With parm@Frosst, most of the parameterization efforts have focused on the torsion terms and the charging scheme. The number of parameters needed is a direct consequence of the addition of new atom types and their occurrences in bioorganic molecules. For this reason, the parsimonious addition of atom types was done with the only objective of fixing general problematic cases. Below is a list of considerations regarding each of the FF terms.

The atomic partial charges were fitted to a RHF/6-31G* electrostatic potential based on the RESP procedure [C.I. Bayly, P. Cieplak, W. D. Cornell, and P.A. Kollman, "A Well-Behaved Electrostatic Potential Based Method using Charge Restraints for Deriving Atomic Charges: The RESP Model", *J. Phys. Chem.,* **97**, (1993), 10269-10280]. This required generally more than one electronic structure calculation per molecule modification examined. The impact of the atomic charges on the parameterization of parm@Frosst is in the torsions since the 1-4 terms couple the electrostatics to the torsions. In principle, if the charge model is modified, the torsion coefficients and phase angle should be revisited. The need for higher throughput while maintaining the accuracy of RESP motivated Christopher Bayly and his Ph.D. student, Araz Jakalian, to develop AM1-BCC [A.Jakalian *et al. J. Comput. Chem.,* ***23,*** (2002), 1623-1641 and A. Jakalian *et al.*, *J. Comput. Chem.*, ***21,*** (2000), 132-146]. As a consequence, all parameterization done after 2001 was done using the AM1-BCC atom charging method. You will find, as part of this release, a molecular dataset typed with the AM1-BCC typer using the original Merck tools. This is to ensure that the AM1-BCC implementations fully benefit from the initial parameterization efforts done by Jakalian et al.

The torsion terms were obtained with a diverse set of electronic structure methods including RHF, MP2 and DFT with a diversity of basis sets. Although details in relative minima height and exact position may vary, these methods usually show quantitative agreement within the seeked accuracy of a FF. Great care was taken to obtain the torsional energy function minimum angle(s) accurately and less care to get the height of the transition peaks correctly. Whenever possible, simultaneous fit was done on few analogs sharing the same torsion type. However, it was common practice to steal parameters from an electronically similar torsion type.

While a great many bond angle bending terms appear in parm@Frosst, very few were actually fitted. The great majority of these parameters were “grandfathered in” from parm96 by analogy with chemically similar atom types; the need for many of these was a side-effect of creating new atom types for the purpose of improving the torsion terms. The AMBER standard was followed. In the few cases where bond angle parameters were fitted, again the level of electronic structure calculation varied.

Finally, very few adjustments were needed for the improper bending angle term, mainly following the AMBER generic values.

Content of the Supporting Data

Because atom typing rules are difficult to implement without ambiguity, two datasets with typed atoms and/or bonds are provided. One is the MMFF94 training set and the other is a subset from the ZINC collection (http://zinc.docking.org).

1) The type definition file

The type definition file, named parm@Frosst.pcp, assigns atom type labels to each atom in a molecule. It uses the atomic element and the connection table as input and assigns the atom type that will be matching force field parameters. The definition language used here is based on PATTY [Bruce Bush and Robert Sheridan, *J. Chem. Inf. Comput. Sci.,* **33**, (1993), 756-762]. Each statement matches a chemical pattern and assigns a label to some (or all) of the atoms in the order of the match. As the typer moves down through the file, types are refined with further definitions and labeling. A defined label can be used in a subsequent matching rule. To understand the details of this file, please refer to the cited reference and to the comments in the file. In the data repository, you will find many molecules with an assign type to each of the atoms. The typing depends on the connectivity given in the structure file, therefore to reproduce the atom types, care should be taken to keep the connection table unchanged.

2) The fitted parameters

The parameter file, parm@Frosst.frcmod, follows the actual AMBER format. Please refer to the Table 14 of reference [W.D. Cornell *et al*., *J. Am. Chem. Soc.*,**117**, (1995), 5179-5197] for the precise definition of each of the terms. Here is a summary of the terms you will find in the file with the symbol used in the paper.

MASS				
A	mass (in atomic mass units)			
BOND				
A-B	K_r	r_{eq}		
ANGLE				
A-B-C	K_θ	θ_{eq}		
DIHEDRAL				
A-B-C-D	M	$V_N/2$	γ_N	*-N*
A-B-C-D	M	…	…	…
A-B-C-D	M	$V_m/2$	γ_m	m
IMPROPER				
A-B-C-D	$V_N/2$	γ_N	*n*	
NONBON				
A	R^*	ε		

Where M corresponds to the number of paths or the number of occurrences of a particular dihedral in a rotatable bond. The negative sign in front of *N* is announcing more entries for a given dihedral. The equation for the dihedral and improper energies is given below. In the case of the improper energy term the number of path *M* is equal to 1.

It is important to note that typically, with the AMBER tools, the protein parameter file needs to be loaded first and then the parm@Frosst file is loaded ensuring that the initial AMBER parameters are first assigned. None of them should be overwritten by parm@Frosst which provides only the missing parameters.

3) The typed molecules

The molecules are obtained from the publicly available ZINC collection as of May 25th 2011. The subset #16 called all-clean which does not have property filtering but has been 'clean' by the Zinc authors. This subset contains 8 513 583 molecules and the authors of the Zinc collection provide a diverse subset of 7562 molecules obtained from a Tanimoto similarity cutoff of 0.6 is applied. This subset is provided by http://zinc.docking.org. The presented dataset contains 7535 molecules, where 27 molecules were identified as duplicates and discarded.

For the sake of completeness, the MMFF94 training set is also included. This set exercises many rare atom types and has a broad chemical coverage.

4) A summary of the provided files

All the structure files are provided in the MDL SD format. You will find 10 files.

File name	Content
parm@Frosst.pcp	The typing rules in PATTY format.
parm@Frosst.frcmod	The AMBER FF parameters specific to small bioorganic molecules
zinc.sdf	Three dimensional structures in the MDL SD format.
zinc_p@f_types.txt	The parm@Frosst atom types. One type per line. This exactly matches the atom ordering found in the zinc.sdf file.
zinc_am1bcc_atypes.txt	The AM1-BCC atom types. One type per line. This exactly matches the atom ordering found in the zinc.sdf file.
zinc_am1bcc_btypes.txt	The AM1-BCC bond types. One type per line. This exactly matches the bond list found in the zinc.sdf file.
mmff94.sdf	Three dimensional structures in the MDL SD format.
mmff94_p@f_types.txt	The parm@Frosst atom types. One type per line. This exactly matches the atom ordering found in the mmff94.sdf file.

mmff94_am1bcc_atypes.txt	The AM1-BCC atom types. One type per line. This exactly matches the atom ordering found in the mmff94.sdf file.
mmff94_am1bcc_btypes.txt	The AM1-BCC bond types. One type per line. This exactly matches the bond list found in the mmff94.sdf file.