J Mol Model (2011) 17:3021–3023
DOI 10.1007/s00894-011-1111-7

SOFTWARE REPORT

SDF2XYZ2SDF: how to exploit TINKER power in cheminformatics projects

Paolo Tosco · Thomas Balle · Fereshteh Shiri

Received: 22 February 2011 / Accepted: 22 March 2011 / Published online: 15 May 2011

Abstract An open-source tool capable of converting SD files (and virtually any other format through OpenBabel) into MMFF-typed XYZ coordinate files to be used with TINKER is described. SDF2XYZ2SDF allows including the power of TINKER molecular mechanics computations in automated cheminformatics workflows, such as conformational searches and virtual screening protocols.

Keywords Cheminformatics · MMFF94 · Molecular mechanics · TINKER

Introduction

The capacity to deal with huge numbers of chemical structures in an automated and unsupervised way is of core importance in cheminformatics. After retrieving a molecule from a chemical fingerprint database, it is often necessary to convert its SMILES representation into a 3D structure. A number of tools serve this purpose; some of them are provided under open source licenses (OpenBabel [1]), others are closed-source, commercial (CORINA [2]) or commercial but "free for academics" (OMEGA [3], SZYBKI [4]). While these tools have to variable extents the capability to carry out additional molecular mechanics calculations on the generated 3D structure, such as conformational searches and geometry optimizations, none of them offers a palette of software modules, force-fields and solvent models as broad as TINKER [5] does. TINKER is a complete, general package for molecular mechanics and dynamics, available with full source code and no license fees. While its field of application has been traditionally focused on biopolymers, recently an implementation of the Merck force field (MMFF) [6] has been included [7], thus potentially making it a very valuable tool in small-molecule cheminformatics. In fact, its collection of lightweight, standalone command-line-driven modules is highly fit for being included in automated scripts and virtual screening workflows. However, a major limitation currently hampers the use of TINKER for such purposes, namely the lack of a tool capable of automatically assigning MMFF atom types, bond types and charges to the input 3D coordinate files. Herein we describe SDF2XYZ2SDF, an open-source software capable of efficiently carrying out this task.

Methods

SDF2XYZ2SDF is written in C and depends on OpenBabel. It consists of two distinct programs, `sdf2tinkerxyz` and `tinkerxyz2sdf`. The first reads a coordinate file in SD format (single or multi-molecule) on

P. Tosco (✉)
Department of Drug Science, University of Torino,
Via Pietro Giuria 9,
10125 Torino, Italy
e-mail: paolo.tosco@unito.it

T. Balle
Department of Medicinal Chemistry, The Faculty of Pharmaceutical Sciences, University of Copenhagen,
2 Universitetsparken,
2100 Copenhagen, Denmark

F. Shiri
Faculty of Chemistry, Razi University,
Baghabrisham,
6714967346 Kermanshah, Iran

standard input and converts it to XYZ format including correct MMFF atom types as required by TINKER:

```
sdf2tinkerxyz < molecule.sdf
```

Piping input through `babel` enables conversion from virtually any format to MMFF-typed XYZ, *e.g.*:

```
babel -imol2 molecule.mol2
-osdf | sdf2tinkerxyz

babel -ipdb protein.pdb
-osdf | sdf2tinkerxyz
```

While our main focus in developing SDF2XYZ2SDF were small molecules, structures with more than 1000 atoms (such as protein or polymer PDB files) may be treated, since both V2000 and V3000 SD files are supported. Besides the XYZ file, a KEY file is also generated with partial charges and bond types needed by TINKER to carry out molecular mechanics calculations.

The second program (`tinkerxyz2sdf`) allows re-converting the XYZ file(s) obtained after performing some computation(s) in TINKER to the original SDF format:

```
sdf2tinkerxyz < molecule.sdf
> molecule_new_crd.sdf
```

Ready-to-use binary distributions of SDF2XYZ2SDF for Windows, Linux, Mac OS X, Solaris and FreeBSD have been prepared which include all necessary OpenBabel dependencies. Source code, binaries and documentation are available under the terms of GNU GPLv3 at the following URL: http://sdf2xyz2sdf.sourceforge.net.

Results and discussion

The implementation of a procedure to correctly assign MMFF atom types, bond types and charges in the format required by TINKER is not trivial from a programmer's point of view. To save a good amount of work, we decided to take advantage of OpenBabel, which is able to assign MMFF atom types, as well as to compute MMFF charges. However, in the TINKER implementation of MMFF the original 99 atom types defined in MMFF are further divided into subtypes, so that SDF2XYZ2SDF needs to correctly assign 212 different atom types. This required complementing the scaffold perception algorithms coded in OpenBabel with new routines capable of discriminating between slightly different atomic neighborhoods. Moreover, a considerable effort was requested to correctly assign MMFF type 1 bonds, namely single bonds connecting sp^2 atoms, which TINKER requires to be explicitly listed in the KEY file via the MMFF-PIBOND keyword. Type 1 bonds occur both in aromatic rings and in conjugate systems; while at a first glance assignment appears to be straightforward, subtle complications arise when condensed ring systems are involved featuring aromatic rings fused to non-aromatic ones. To correctly deal with such systems, a ring perception routine was implemented in SDF2XYZ2SDF using the breadth-first search algorithm described by Figueras [8].

SDF2XYZ2SDF is able to correctly assign MMFF atom types, bond types and charges to both the MMFF94 [9] and MMFF94s [10] validation suites available on CCL.NET. These tasks can be carried out with two simple commands:

```
babel -imol2 MMFF94_dative.mol2
-osdf | sdf2tinkerxyz

babel -imol2 MMFF94s_dative.mol2
-osdf | sdf2tinkerxyz
```

A validation script is included in the SDF2XYZ2SDF distribution which, after executing these commands, computes energies with TINKER's `analyze` module and compares them to the ones reported by Kearsley in the MMFF validation suites. All tests are passed for both MMFF94 and MMFF94s, the only exception being the CYGUAN01 molecule; exhaustive explanations for this discrepancy are given by Staelens [11], who first implemented MMFF in TINKER.

Two enhancements are introduced by SDF2XYZ2SDF to complement the native TINKER MMFF implementation:

1. MMFF94s parameters become available to TINKER via the SDF2XYZ2SDF distribution;
2. SDF2XYZ2SDF allows overcoming a limitation concerning ionic heteroaromatic rings [7], so that it becomes possible to deal with these systems in TINKER by passing explicit charges through the KEY file.

Summary

We have realized a tool capable of converting SD files (and virtually any other format through OpenBabel) into MMFF-typed XYZ files to be used with TINKER. This opens up the possibility to include the power of TINKER molecular mechanics computations in automated, script-driven chem-informatics workflows such as conformational searches and virtual screening protocols; the SDF2XYZ2SDF API is currently used for this purpose by our Open3DALIGN project [12]. The modular and open-source nature of this project will enable us, as well as other developers willing to contribute, to add other desirable features in the future, such as atom-typing according to other force-fields supported by TINKER and integration in OpenBabel as a plugin. Further information and examples are available on the SDF2XYZ2SDF website (http://sdf2xyz2sdf.sourceforge.net).

Acknowledgments The first acknowledgment goes to OpenBabel developers, particularly to Geoffrey Hutchison, who merged in real time our changes and additions to the OpenBabel development trunk. Without OpenBabel, SDF2XYZ2SDF would have probably never seen the light of day. Next we wish to thank Prof. Jay Ponder for his kind feedback and for inspiring the development and the public release of SDF2XYZ2SDF; thanks also to Nicolas Staelens for providing the first implementation of MMFF in TINKER. T.B. acknowledges the Lundbeck Foundation for financial support. F.S. acknowledges the Iranian Ministry of Education for funding her stay at the Department of Drug Science in Torino as a visiting scientist.

References

1. OpenBabel development version (2011) http://openbabel.org/. Accessed 9 March 2011
2. Gasteiger J, Rudolph C, Sadowski J (1990) Automatic generation of 3D-atomic coordinates for organic molecules. Tetrahedron Comput Methodol 3:537–547. doi:10.1016/0898-5529(90)90156-3
3. Boström J, Greenwood JR, Gottfries J (2003) Assessing the performance of OMEGA with respect to retrieving bioactive conformations. J Mol Graph Model 21:449–462. doi:10.1016/S1093-3263(02)00204-8
4. Wlodek S, Skillman AG, Nicholls A (2010) Ligand entropy in gas-phase, upon solvation and protein complexation. Fast estimation with quasi-Newton hessian. J Chem Theor Comput 6:2140–2152. doi:10.1021/ct100095p
5. TINKER - Software tools for molecular design, version 5.1. http://dasher.wustl.edu/tinker/. Accessed 9 March 2011
6. Halgren TA (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem 17:490–519. doi:10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P
7. Staelens N, Ponder J (2009) TINKER implementation of the MMFF94 Force Field. http://dasher.wustl.edu/tinker/distribution/mmff/0README. Accessed 9 March 2011
8. Figueras J (1996) Ring perception using breadth-first search. J Chem Inf Comput Sci 36:986–991. doi:10.1021/ci960013p
9. Kearsley S (1999) MMFF94 Validation Suite. http://ccl.net/cca/data/MMFF94/index.shtml. Accessed 9 March 2011
10. Kearsley S (1999) MMFF94s Validation Suite. http://ccl.net/cca/data/MMFF94s/index.shtml. Accessed 9 March 2011
11. Staelens N (2007) Validation Annex 1. http://dasher.wustl.edu/tinker/distribution/mmff/Info.pdf. Accessed 9 March 2011
12. Tosco P, Balle T (2011) Open3DALIGN: an open-source software aimed at unsupervised molecular alignment. http://open3dalign.org. Accessed 9 March 2011