K306 The new release of 3DMET: Evaluation of 2D-3D converters and structure verification

Miki MAEDA[*1] and Kazumi KONDO[2]

1) *National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, JAPAN*
2) *Quests Research Institute, Otsuka Pharmaceuticals, Co., Ltd., 463-10 Kagasuno Kawauchi-cho, Tokushima, 771-0192, JAPAN*

* mmada@nias.affrc.go.jp

1. Introduction

The release 2.0 of 3DMET (A 3D-structure database of natural METabolites, http://www.3dmet.dna.affrc.go.jp/) [1] was opened to access on Aug. 1, 2008. The main difference between the release of 1 and 2.0 are number of contents and the methods of file conversion and verification for correction between 2D- to 3D-structures.

For collecting correct structures to the database, we tested several programs because we found that not negligible numbers of resulted structures had mistaken structures, especially for chiral atoms/bonds, during this conversion process. We use several converters widely distributed by their program vendors, and some strings describing chemical structures with chiral descriptions. The details are not shown in this abstract, but the results will be also reported in the presentation.

The 3DMET server was also re-constructed for the new release. Thirty descriptors are searchable and a simple substructure search using SMILES [2] strings is provided. Documents are also enriched. We will report the new release of 3DMET and the results of evaluation of 2D-3D converters and their verification.

2. Method

The 3D-structures of the database are converted from the 2D-structures of COMPOUND [3] provided by the GenomeNet. For extraction of natural compounds, the following three limitations were applied to the COMPOUND entries. (1) The entry does not have "R", "n" and "X" in the formula. (2) The entry is included in the KEGG pathway map #1.1 to #1.10, in the "Compounds with biological roles", in the "Lipids", and in the "Phytochemical compounds". (3) The entry does not have some halogens (F, Cl, and Br).

For 2D-3D structure conversion, CONCORD standalone (supplied with SYBYL 7.2) and CORINA (version 3.4) were used. The converted files by CONCORD were minimized by the SYBYL minimizer and the files by CORINA were minimized by using MOE (version 2006.08). The structures of the initial

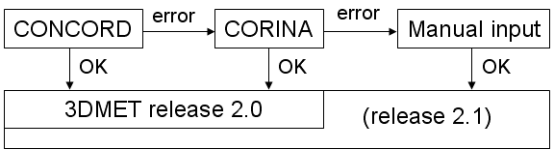

Figure 1. Flow chart of data production

2D-mol file and the 3D-mol2 file after minimization were compared by canonical SMIILES (Daylight Software Version 4.94) and InChI (version 1.00) [4]. The chiral description about phosphate was ignored. The entries of conserved structures between two structures before/after conversion were recorded to 3DMET (Figure 1). If the chiral tags exist in only 3D-structures, the structures were judged as conserved structures. The 30 descriptors of the structures were calculated by MOE (version 2007.09).

3. Results and Discussion

Our results of the evaluation of converters show CONCORD standalone and CORINA are almost the same accuracy to produce 3D-structures. Some structures could not convert by both of the programs. The structures have some common substructures. Thus, the entries will have some substructures not included in the rule database of the programs. Between two CONCORD programs (standalone and sketch), conversion accuracy was much different. Thus, we should choose CONCORD standalone when automatic conversion is needed.

From our tested results, molecular strings by canonical SMILES and InChI were also almost the same accuracy. It is noteworthy that some structures cannot be converted as strings by two programs. The typical error examples for canonical SMILES are symmetric molecules. It would be caused by two or more canonical orders for one molecule.

Figure 2 shows examples of the new release of 3DMET pages. In the data page of the new version, 30 descriptors calculated by MOE and structure verification methods are shown. As a viewer of molecules, Jmol [5] was substituted for MDL Chime because Jmol can be used on many kinds of computers such as MS windows, Macintosh, and Linux, etc.

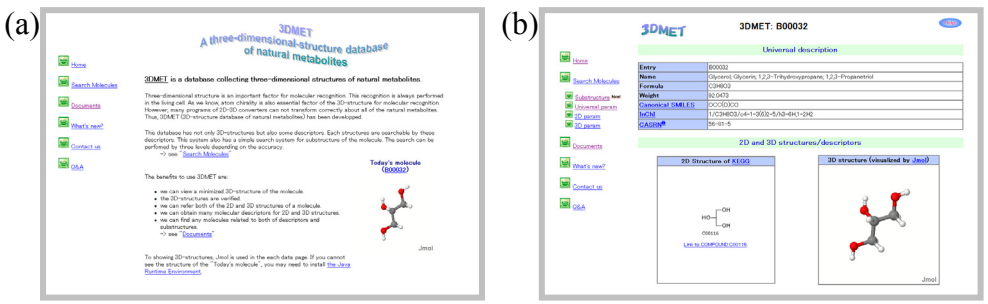

Figure 2. The new pages of the 3DMET release 2.0: (a) The top page and (b) the upper part of view of the data entry page. This data page is following by the 30 descriptors and verification columns.

References

1. Maeda M. *Proceeding of the 33rd QSAR symposium of Japan*, **2005**, K01.
2. Weininger D. *J. Chem. Inf. Comput. Sci.*, **1988**, 28, 31-36.
3. Goto S, Nishioka T, and Kanehisa M. *Bioinformatics*, **1998**, *14*, 591-599.
4. InChI: IUPAC International Chemical Identifier http://www.iupac.org/web/ins/2000-025-1-800
5. Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/