

HerbalDB 2.0: Optimization of Construction of Three-Dimensional Chemical Compound Structures to Update Indonesian Medicinal Plant Database

Rezi Riadhi Syahdi¹, Jasmine Tiara Iqbal¹, Abdul Munim¹, Arry Yanuar^{1*}

Rezi Riadhi Syahdi¹, Jasmine Tiara Iqbal¹, Abdul Munim¹, Arry Yanuar^{1*}

¹Faculty of Pharmacy, Universitas Indonesia, Depok, 16424 West Java, INDONESIA.

Correspondence

Arry Yanuar

Faculty of Pharmacy, Universitas Indonesia, Depok 16424, West Java, INDONESIA.

E-mail: arry.yanuar@ui.ac.id

History

- Submission Date: 16-02-2019;
- Review completed: 12-04-2019;
- Accepted Date: 24-05-2019.

DOI : 10.5530/pj.2019.11.184

Article Available online

<http://www.phcogj.com/v11/i6>

Copyright

© 2019 Phcogj.Com. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

ABSTRACT

Objective: Development of novel drugs is an important challenge in the pharmaceutical world and industry. *In-silico* methods are often considered in refinement / correction processes of drug design because they may lower the costs. The *in-silico* drug discovery process requires a three-Dimensional Structure (3DS) of the chemical compounds as input. Computational 3DSs often exhibit structural mismatches thus affecting the validity of the *in-silico* drug design process. In a previous study, a 3DS database with 1405 of Indonesian herbal compounds was developed, named HerbalDB. In this database, various structural mismatches were identified in some of the 3DSs. Our study aimed to identify and correct the structural mismatches in the herbalDB and to determine the best method in creating correct 3DS of chemical compounds. **Methods:** Structural mismatches in the herbal database were identified by molecular visualization. **Results:** The identification process yielded 170 compounds with structural mismatches that were corrected with 10 different parameters using the MarvinSketch and VegaZZ software, evaluated by molecular visualization. **Conclusions:** based on 3DS of chemical compound visualization, *.mol and *.sdf file format created using Dreiding force fields of MarvinSketch are the best method to construct the proper structure of Indonesian medicinal plant's chemical compound database compared with MMFF94, AMBER and CHARMM forcefields.

Key words: Herbal database, MarvinSketch, Three-dimensional structure, VegaZZ.

INTRODUCTION

The process of designing drugs is a major challenge in pharmaceutical industry because it is costly and time-consuming to design, test and register a new drug candidate.¹ There are two commonly used methods of drug designing; high throughput screening and rational drug design. Rational drug design is often performed using virtual screening.² Virtual screening is an inexpensive *in silico* approach to find new potential drug. In virtual screening, hundreds of thousands to millions of molecules can be tested within a short time and this computational screening can reduce the number of *in vitro* assay required. Therefore, this approach also potentially reduces the cost and time of the development process.³

There are two commonly used approaches for *in-silico* drug design; ligand-based drug design and structure-based drug design⁴ Both methods require a three-dimensional structure (3DS) of chemical compound. 3DSs can be obtained either experimentally or computationally.⁵ Computationally generated 3DSs often produce structural mismatch owing to issues with geometry optimization and three-dimensional transformations. These structural mismatches affect the validity of the *in-silico* drug design process. In a previous study, a herbal database with 1405 three-dimensional structure of chemical compound were made.⁶ In this database, we found

a substantial number of mismatches in three-dimensional structure of chemical compound. In the current study, we aimed to identify and correct the structural mismatches and find the best method of constructing 3DSs.

METHODS AND MATERIAL

Hardware

A client computer with this following specification: Dell with Intel Core™ i3-7100U (Intel Core™, USA) Computer Processing Unit (CPU), 8GB RAM with Microsoft Windows 10 Home Single Language 64-bit operating system.

Software

The software used in this research are MarvinSketch (ChemAxon, Budapest),⁷ VegaZZ 3.1.1.42 (Universita Degli Milano, Italy),⁸ and PyMOL (DeLano Scientific LLC, Italy).⁹

3D structures

Three-dimensional structures of chemical compound were obtained from Herbal Database (Biomedical Computation and Drug Design Laboratory, Universitas Indonesia).⁶

2D Reference structures

Two-dimensional reference structures were obtained from Kanaya KNApSAcK (<http://kanaya.naist.jp/>)

Cite this article: Syahdi RR, Iqbal JT, Munim A, Yanuar A. HerbalDB 2.0: Optimization of Construction of Three-Dimensional Chemical Compound Structures to Update Indonesian Medicinal Plant Database. Pharmacogn J. 2019;11(6):1189-1194.

KNAPSAcK¹⁰ and Pubchem compound (<http://PubChem.ncbi.nlm.gov>).¹¹

Examination of three-dimensional structure of compounds in HerbalDB

3DSs in HerbalDB were examined to identify compounds having structural mismatch with their two-dimensional reference structures. The 1405 compounds in HerbalDB were examined using visualization generated by the PyMOL software. Structural mismatches were identified by a visual comparison between 3DSs in HerbalDB and two-dimensional reference structures obtained from Kanaya KNApSAcK database and Pubchem compounds. Compounds that were identified as having structural mismatches with their reference structures were listed using Microsoft Excel.

Repair of two-dimensional structures

Compounds with mismatched structures were reconstructed into two-dimensional structures with *.mol file format using the MarvinSketch software. The reconstruction was manually performed by drawing structures based on the two-dimensional reference structures from Kanaya KNApSAcK and Pubchem compounds databases. The two-dimensional structures were in *.mol format and were named according to the ID number.

Repair of three-dimensional structures

The reconstructed two-dimensional structures were converted into 3DSs using 10 different parameters / file formats, as listed in Figure 1, in the MarvinSketch and VegaZZ software. The two-dimensional

structures were converted into three-dimensional structures using two different force fields in MarvinSketch; Dreiding and MMFF94. The three-dimensional conformer with the lowest energy of each force field were then saved in three different file formats; *.mol, *.sdf and *.pdb. Total six 3DSs files were obtained for each chemical compound using the MarvinSketch software.

In VegaZZ, the two-dimensional structures were converted into 3DSs using two different force fields; AMBER and CHARMM. The three-dimensional conformers for each force field were then saved in two different file formats: *.mol and *.pdb. Four 3DSs files were obtained for each chemical compound using the VegaZZ software. Ten 3DSs files were obtained for each repaired chemical compound.

Analysis and visualization of the repaired structures

Next, the validity of the reconstructed 3DS structures was then examined. The examination was performed by manual visualization using the PyMOL software. Visualization results were then recorded using Microsoft Excel. Based on the visualization results, the structures that had corrected with reference to the two-dimensional structures could be identified. The best methods for constructing 3DSs for chemical compounds in herbal database could also be determined.

RESULTS

Visualization was performed on 1405 3DSs of chemical compound in HerbalDB using the PyMOL software to identify compounds with structural mismatches. The visualization result is summarized in Table 1. Next, the structures in priority I then were reconstructed using a few parameters in the MarvinSketch and VegaZZ software, the final

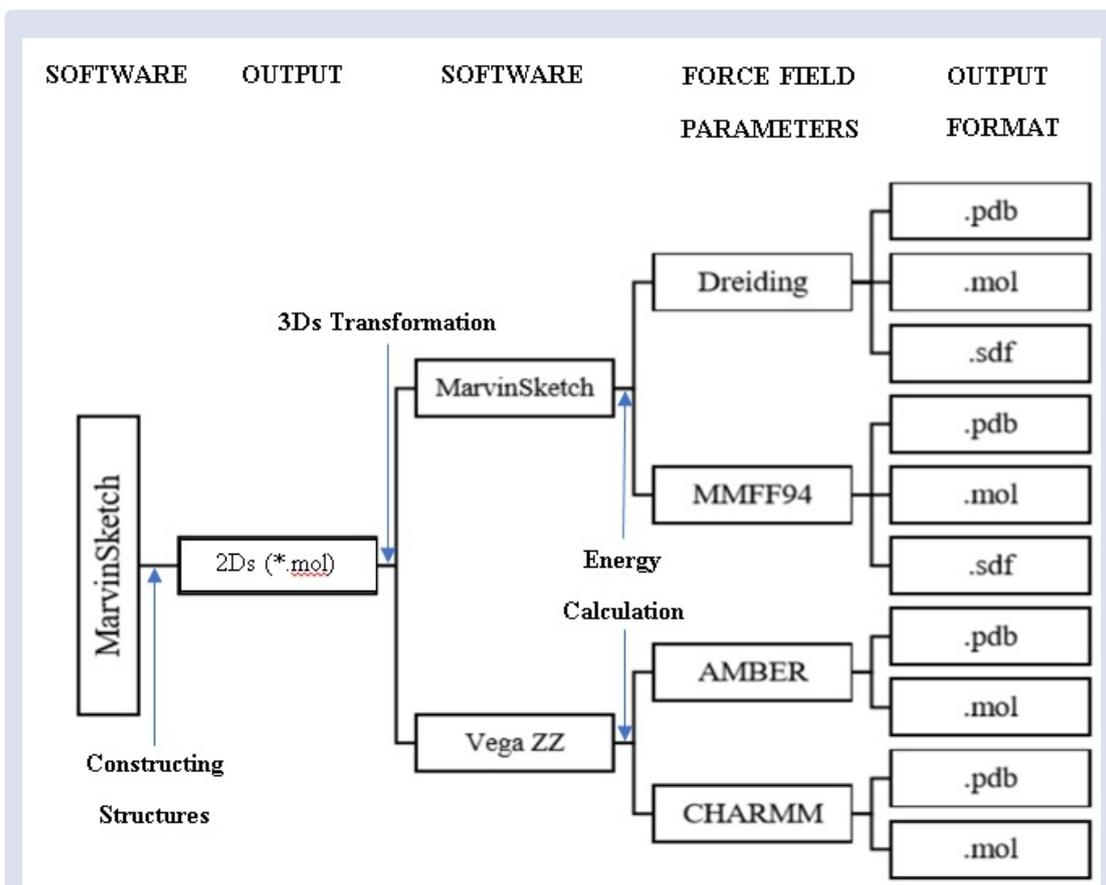
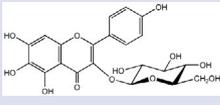
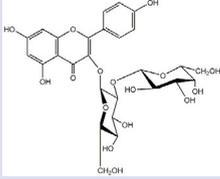
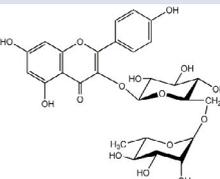
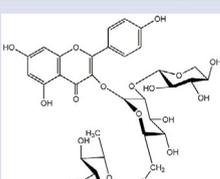
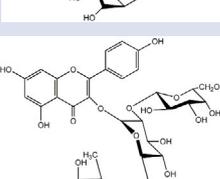
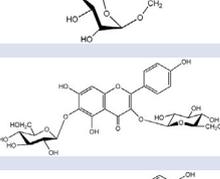
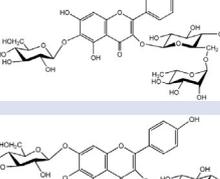
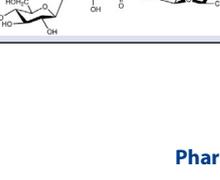


Figure 1: Parameters varied for creating three-dimensional structure.

Table 1: HerbalDB visualization results.

No	Categories	Description	Amount of Structures
1	Priority I	Different structure Empty structure Disappearance of double and triple bond(s) Appearance of double and triple bond(s) Substitution changes Detaches structure	178
2	Priority II	Stereochemical changes	134
3	Priority III	Resonance	711
4	Accepted	The structure matches reference structure	294
Total			1405

Table 2: Example of energy calculation comparison between similar structures constructed using Marvin Sketch.

No	Compound Name	Structure	Energy (kcal/mol)	
			Dreiding	MMFF94
1.	Hydroxykaempferol 3-glucoside M00005306		70.86	171.37
2.	Camelliaside C M00005164		86.68	268.74
3.	Nicotiflorin M00005169		87.87	256.11
4.	Camelliaside B M00005214		103.69	330.49
5.	Camelliaside A M00005216		105.48	345.29
6.	6-Hydroxykaempferol 3,6-diglucoside M00005308		106.88	290.82
7.	6-Hydroxykaempferol 3-rutinoside-6-glucoside M00005310		109.59	388.41
8.	6-Hydroxykaempferol 3,6,7-triglucoside M00005311		134.8	398.13

reconstruction results are summarized in Tables 2-4. An example of corrected result is shown in Figure 2.

DISCUSSION

First step of this research was reassessing the three-dimensional molecular structures from Indonesia herbal database which were created and published previously. The visualization performed in this study revealed that only 294 three-dimensional structures of compounds in the database matched their two-dimensional reference

structures. The remaining 1111 3DSs of compounds were experiencing a structural mismatch. The structural mismatches were categorized in three categories based on the degree of structural mismatch and their impact on chemical structure integrity.

The two-dimensional structures used as reference were downloaded based on the ID number. After the structures were revisualized and compared with those in priority I category (with two-dimensional reference structure that were downloaded based on the compound name), eight of the 178 structures were found to have a priority I

Table 3: Structures reconstructed using MarvinSketch visualization result.

No	File Format	Structural Mismatch	Quantity	Total
1.	*.mol & *.sdf	Substitution direction changes, stereochemical changes, and disappearance of double or triple bond(s)	3	6
		Electropositive substitution changes into one extra double bond	2	
2.	*.pdb	Substitution position changes to the opposite site	1	149
		Substitution direction changes, stereochemical changes	3	
		Substitution position changes to the opposite site	1	
		Disappearance of double or triple bond(s) on the chain/substitution	24	
		Disappearance of double or triple bond(s) on the cyclic ring(s)	2	
		Resonance on aromatic ring(s)	93	
		Resonance on aromatic ring(s) and carboxylic acid substitution	1	
Resonance on aromatic ring(s) and disappearance of double or triple bond(s) on the chain/substitution	25			

Table 4: Structures reconstructed using VegaZZ visualization result.

No	File Format	Structural Mismatch	Quantity	Total
1.	*.mol	Substitution direction changes, stereochemical changes	3	87
		Substitution direction changes, stereochemical changes, and substitution position changes to the opposite site	1	
		Resonance on aromatic ring(s)	65	
		Appearance of double or triple bond(s)	17	
		Disappearance of C atom and double bond(s)	1	
		Substitution direction changes, stereochemical changes	4	
2.	*.pdb	Substitution direction changes, stereochemical changes, and substitution position changes to the opposite site	3	152
		Substitution position changes to the opposite site	1	
		Disappearance of double or triple bond(s)	139	
		Disappearance of C atom and double bond(s)	4	

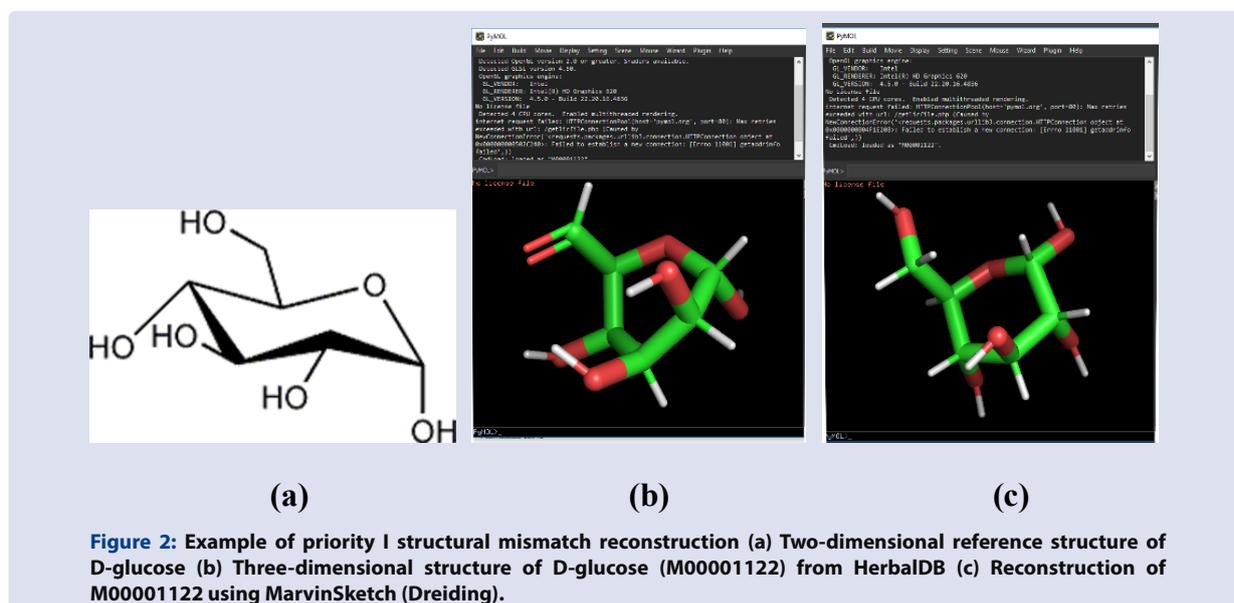


Figure 2: Example of priority I structural mismatch reconstruction (a) Two-dimensional reference structure of D-glucose (b) Three-dimensional structure of D-glucose (M00001122) from HerbalDB (c) Reconstruction of M00001122 using MarvinSketch (Dreiding).

structural mismatch because these structures were assigned a wrong ID number. These structures were then eliminated from the priority I category. The remaining 170 structures were then repaired using a few different parameters in the MarvinSketch and VegaZZ software. A total of 1700 three-dimensional structures were created. Next, these structures were then visualized again using PyMOL.

In this study, several software were evaluated in creating and repairing the three dimensional structure. Two softwares and four force fields were compared in this study. The forcefields are Dreiding,¹² MMFF94,¹³ Amber,¹⁴ and CHARMM.¹⁵ Amber and CHARMM were two force fields accounted for the most commonly used in simulations of biomolecules.^{16,17} Dreiding and MMFF94 were created after Amber and CHARMM and served also in several computations in drug design. However each force field has its particular strengths and weaknesses based on its parameterization therefore all force fields available were tried in this research.¹⁸

The structures that were constructed with MarvinSketch using two force fields; Dreiding and MMFF94, yielded structurally identical result. The difference between the two force fields arose because there are two structures that cannot be constructed using MMFF94, probably because of the program limiting the structure construction process. Most of the structures in the database are cyclic. Previous research also found that MMFF94 performed poorly in the set of compound with alicyclic and aromatic chemical structure.¹⁹ From our research, the Dreiding force field constructed structures with a relatively lower energy calculation for these structures compared to MMFF94. However, the chosen forcefield should be based on application not by its energy. Furthermore, in the case of Dreiding force field, the anomeric effect was not considered during development, so it is important to note this during creation of carbohydrate 3DS.²⁰

The structures that were constructed in *.mol and *.sdf format were also structurally identical. Of the 170 structures that were reconstructed, 164 structures in *.mol and *.sdf format were successfully repaired, whereas only 21 structures in the *.pdb format were successfully repaired. Structures with one or more aromatic ring(s) and carboxylic acid substitutions were constructed as resonance structures and structures with double or triple bond(s) were constructed as a single bond(s) structure owing to the limitation of *.pdb file. The *.pdb file format cannot defined a few types of bond. All covalent bonds in proteins and in nucleic acids are properly assigned, however, hydrogen bonds, sulfur and water bridges and coordinated bonds were not recognized. Covalent bonds in hetero groups were determined based on geometry position, two atoms are considered bonded if they are closer than the sum of their covalent radii plus a small tolerance.²¹ Therefore, it may induce misread of the atoms as unbinded.

The structures that were constructed with the VegaZZ software using two force fields; AMBER and CHARMM, also yielded structurally identical results. The two force fields cannot calculate the conformer's energy. Hence, there is no difference between the results when using either of the two force fields. The structures that were constructed in *.mol and *.sdf format were structurally identical. Of the 170 structures that were reconstructed, 83 structures in *.mol and *.sdf format were successfully repaired. Some of the structures with one or more aromatic ring(s) were constructed based on resonance structures. All of the structures with electropositive substitution were constructed as it resonance structure. This might happen because the program cannot define a 3DS with electropositive substitution. In addition, some of the structures with single bond(s) were constructed as a double or triple bond(s) structure.

Of the 170 structures that were reconstructed, 18 structures that were constructed in *.pdb format were successfully repaired. The remaining structure with double or triple bond(s) were constructed as a single bond(s) structures owing to the limitations of the *.pdb file format as discussed before.

CONCLUSIONS

Of the 1405 structures in HerbalDB, only 294 3DSs of compounds structurally match with their two-dimensional reference structures. Of the other 1111 structures, 170 structures were reconstructed using a few different parameters in the MarvinSketch and VegaZZ software. The reconstruction yielded a total of 1700 3DSs; 680 structures in *.mol format, 680 structures in *.pdb format and 340 structures in *.sdf format. From these 3DSs, 164 structures in *.mol and *.sdf format and 39 structures in *.pdb format were successfully reconstructed based on the two-dimensional reference structures. The relatively best method for constructing three-dimensional structures in HerbalDB is by using the MarvinSketch software in *.mol and *.sdf format and the Dreiding force field.

ACKNOWLEDGMENTS

All authors acknowledge Universitas Indonesia for support and PITTA Research Grants 2018.

CONFLICTS OF INTEREST

Authors declare no conflicts of interest.

REFERENCES

1. Taft AC, da Silva VB, da Silva CHTdP. Current topics in computer-aided drug design. *Journal of Pharmaceutical Sciences*. 2008;97(3):1089-98.
2. Ramirez D. Computational methods applied to rational drug design. *Open Med Chem J*. 2016;10(7-20).
3. Vu C, Quyen L, Huong N. In silico drug design: Prospective for drug lead discovery. *International Journal of Engineering Science Invention*. 2015;4(10):2319-6734.
4. Katsila T, Georgios AS, Patrinos GP, Matsoukas MT. Computational approaches in target identification and drug discovery. *Computational and Structural Biotechnology Journal*. 2016;14:177-184.
5. Ferreira LG, dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. *Molecules*. 2015;20(7):13384-421.
6. Yanuar A, Mun'im A, Lagho ABA, Syahdi RR, Rahmat M, Suhartanto H. Medicinal plants database and three dimensional structure of the chemical compounds from medicinal plants in indonesia. *International Journal of Computer Science Issues*. 2011;8(5):180-83.
7. ChemAxon. MarvinSketch (Product Version 18.21.0) [Computer Software]. 2018. <https://chemaxon.com/products/marvin>.
8. Vega ZZ. Version 3.0. 3.11 [Computer Software]. Drug design laboratory, Department of pharmaceutical sciences, University of Milan. 2014. http://nova.disfarm.unimi.it/cms/index.php?Software_projects:VEGA_ZZ:Download.
9. DeLano WL (2002) Pymol: An open-source molecular graphics tool. <https://pymol.org/2/>.
10. Nakamura K, Shimura N, Otabe Y, Hirai-Morita A, Nakamura Y, Ono N, *et al*. KNApSack-3D: A three-dimensional structure database of plant metabolites. *Plant Cell Physiol*. 2013;54(2):e4.
11. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, *et al*. PubChem substance and compound databases. *Nucleic Acids Res*. 2016;44:D1202-13.
12. Mayo SL, Olafson BD, Goddard III WA. DREIDING: A generic force field for molecular simulations. *J Phys Chem* 1990;94:8897-09.
13. Halgren TA. Merck molecular force field VII. Characterization of MMFF94, MMFF94s and other widely available force fields for conformational energies and for intermolecular interaction energies and geometries. *J Comp Chem*. 1999;20:730-48.
14. Weiner SJ, Kollman PA, Nguyen DT, Case DA. An all atom force field for simulations of proteins and nucleic acid. *J Comput Chem*. 1986;7(2):230-52.
15. Ha SN, Giammona A, Field M, Brady JW. A revised potential-energy surface for molecular mechanics studies of carbohydrates. *Carbohydr Res*. 1988;180(2):207-21.

16. Smith MD, Rao JS, Segelken E, Cruz L. Force-field induced bias s in the structure of A β 21–30: A comparison of OPLS, AMBER, CHARMM and GROMOS force fields. *Journal of Chemical Information and Modelling*. 2015;55:2587-95.
17. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biol*. 2011;9:71.
18. Gonzalez MA. Force fields and molecular dynamics simulations. *Collection SFN*. 2011;12:169-200.
19. Ilana Y, Kanal JA, Hutchison GR. A sobering assessment of a small-molecule force field methods for low energy conformer predictions. 2017.
20. Foley BL, Tessier MB, Woods RJ. Carbohydrate force fields. *Wiley Interdiscip Rev Comput Mol Sci*. 2012;2(4):652-97.
21. Baber J, Hodgkin E. Automatic assignment of chemical connectivity to organic molecules in the cambridge structural database. *Journal of Chemical Information and Computer Science*. 1992;32(5):401-6.

Cite this article: Syahdi RR, Iqbal JT, Munim A, Yanuar A. HerbalDB 2.0: Optimization of Construction of Three-Dimensional Chemical Compound Structures to Update Indonesian Medicinal Plant Database. *Pharmacog J*. 2019;11(6):1189-94.