Virtual screening of natural products as an inhibitor of DNA methyltransferase 1 enzyme for breast cancer disease

Ina Nur Istiqomah¹,a, Ahmad Husein Alkaft¹,b, Mutiara Saragih¹,c, Ade Hanna Natalia¹,d, Usman Sumo Friend Tambunan¹,e

¹Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Kampus UI Depok, West Java, Indonesia, 16424
Email: a inanuristiqomah@gmail.com, b ahmad.husein@sci.ui.ac.id, c mutiarasara89@gmail.com, d adehanna.natalias@gmail.com, e usman@ui.ac.id

Abstract. Breast cancer is the most prevalent cancer in women worldwide. It has the highest number of new cases which amounted to 40 per 100,000 cases per year, 12.9% of which leads to death. Epigenetic alteration plays a vital role in the process of cancer cell formation and propagation. DNA methylation is one of the most common types of epigenetic alteration which generally leads to breast cancer. The DNA Methylation, a transfer of methyl group from S-adenosyl-methionine (SAM) to cytosine in the CpG dinucleotide, is catalysed by a DNA Methyltransferase-1 (DNMT1) enzyme. In the present study, we performed a virtual screening of natural product compounds as an inhibitor of the DNMT1 enzyme. Virtual screening was conducted on 26,731 natural products obtained from the NCBI PubChem database. Three steps of rigid and one step of flexible molecular docking simulations were performed using MOE 2014.09. Through the simulations, 10 best ligands based on the Gibbs free binding energies (ΔGbinding) and the ligand-enzyme complex interactions were identified. The pharmacological test was conducted to observe the physicochemical, toxicity, carcinogenicity-mutagenicity, and bioactivity properties by employing DataWarrior 4.7.2., Toxtree, Molinspiration, admetSAR, and SWISSADME software. The results revealed that three best ligands from the phenolic group were selected due to their exceptional pharmacological characteristics as the drug candidate for breast cancer therapy.

Keywords: Breast Cancer; DNMT1 Enzyme; Epigenetic Regulation; Molecular Docking; Natural Products; Pharmacological Test.

1. Introduction
Cancer is a non-communicable disease which characterized by abnormal and uncontrolled cell growth that can damage surrounding tissues and can spread to other parts of the body. Cancer cells are malignant and come out or grow from any cell in the human body [1]. According to data from World Health Organization (WHO) in 2013, the incidence of cancer increased from 12.7 million cases in 2008 to 14.1 million cases in 2012 with the number of deaths increased from 7.6 million in 2008 to 8.2 million in 2012. Cancer is the second cause of death in the world by 13% after cardiovascular disease [2]. One of the most frightened types of cancer especially in women is breast cancer. According to data from Globocan, the International Agency for Research on Cancer (IARC) in 2012, breast cancer is cancer with the highest percentage of new cases in the world which is 43.3% or 40 per 100,000 women while the percentage of breast cancer deaths is 12.9%. Meanwhile, breast cancer in Indonesia ranks the second place after cervical cancer. About 26 females per 100,000 female population suffering from breast
cancer every year. Patients with the number of breast cancers in Indonesia have the highest number in 2013 with a total of 61,682 cases [3].

Epigenetic alteration plays an essential role in the process of the formation of cancer cells. DNA methylation is a type of epigenetic change that often occurs and specific to a particular tissue. It is one of the epigenetic control mechanisms which do not change the sequence of the nucleotides in DNA [4]. In normal cells, the actively expressed gene is a gene that does not experience methylation in the CpG region. In cancer cells, the hypermethylation prevents the gene to be transcribed. So changes in DNA methylation pattern are one of the main signs of all cancers in humans.

DNA methylation is mediated by DNA methyltransferases (DNMTs) which catalyse the transfer of methyl groups from S-adenosyl-methionine (SAM) to cytosine in CpG dinucleotides. Three active forms of DNMTs have been identified, namely DNMT1, DNMT3a, and DNMT3b [5]. In breast cancer, epigenetic changes occur due to an increase in DNMT1 activity. Breast cancer has a high expression of DNMT1. Hence, epigenetic therapy for the treatment of breast cancer is possible to be performed by inhibiting DNA methylation caused by DNMT1 [6, 7]. Therefore, inhibition of DNMT1 activity can be used as a way to prevent cancer cell formation.

Natural products are known as an important source in the manufacture of drugs because they have good bioactivity, exceptional bioavailability and low toxicity [8]. Several studies have been conducted showing natural products such as epigallocatechin-3-gallate-3-gallate (EGCG), curcumin, and genistein can be used as an inhibitor of DNMT1 [8-10]. Therefore, virtual screening of natural products as an inhibitor of DNMT1 as a therapeutic agent for breast cancer treatment will be performed in this study.

2. Materials and Methods

2.1. Material and Methods
The research was performed through in silico to discover a drug candidate from natural products. The pharmacological test was done using Osiris DataWarrior 4.7.2, Toxtree, Molinspiration, AdmetSAR, and SwissADME. Molecular Operating Environment (MOE) 2014.09 software was employed to perform the rigid and flexible docking simulation. The natural products were acquired for NCBI PubChem.

2.2. Preparation of DNA Methyltransferase 1 Enzyme
The 3D structure of DNMT1 enzymes were obtained from the Protein Data Bank at the Collaboratory Research for Structural Bioinformatics (RCSB PDB) with PDB ID 4WXX. The geometry optimization and energy minimization of the DNMT1 Enzyme 3D structure was performed by removing water molecules. Then, protonation with protonate 3D protocol was carried out to change the macromolecular state to its ionization level. Then, optimization was done with hydrogen fix and partial charge. Energy minimization was carried out with the AMBER10: EHT force field with the R-field solvation and RMS Gradient was 0.001. Other parameters use the default and output files in the .moe format.

2.3. Preparation of Natural Products Ligand
The geometry optimization and minimization of 3D ligand structures were initiated by fanning all ligands into the MOE 2014.09. The purpose is to fix the ligand structure and the hydrogen atoms position in the ligand. Molecular docking simulations were performed using the 2014.09 MOE software. The placement method used in this step was the triangle matcher with the retain value of 100 and the London dG scoring function. The 100 best data were refined based on force field parameters. Other parameters were in accordance with the default of the MOE 2014.09 software, and the output file was saved in the .mdb format. The step of reviewing the enzyme preparation process and determining the active site of the enzyme in the sequence editor menu by selecting amino acid residues in the enzyme chain was followed by the molecular docking process which was done in the simulation dock menu. The molecular docking output was saved in .mdb format.
2.4. Pharmacological Properties ADMET
The potential ligands from molecular docking simulation underwent pharmacological properties analysis. Toxtree, SwissADME, Molinspiration, and AdmetSAR software were used to predict the pharmacological properties of ligands.

3. Results and Discussions

3.1. Preparation of Natural Product ligands
About 167,835 natural product compounds underwent initial screening process which was carried out using the DataWarrior 4.7.2 software. Druglikeness higher than 0, no mutagenic, no tumorigenic, no reproductive effective, no irritant, and no nasty function were the toxicity prediction parameters. Only 26,731 compounds fulfilled all the parameters in the screening test. Then, the compounds were saved in a mdb format for the next step.

3.2. 3D Protein Structuring and Visualization of DNA Methyltransferase 1 Enzyme
The protein which was used in this research, the DNMT1, is an enzyme that can trigger the growth of cancer cells in the breast. In this research, the protein was obtained from RCSB with PDB ID 4WXX. The water molecules must be removed because proteins are rigid and the effect of solvation is not involved in molecular docking simulations [11]. After that, the protonation process was carried out through the 3D protonate command which aimed to add hydrogen atoms to the protein crystal structure. The existence of hydrogen atoms is necessary for molecular mechanics and dynamics and all the electrostatic calculations involved [12]. Furthermore, an energy minimization process was performed to obtain enzyme conformation with the lowest energy using AMBER10: EHT force field. Visualization of the 3D structure of DNA Methyltransferase 1 Enzymes can be seen in Fig. 1.

![Figure 1](image1.png)

Figure 1. (a) 3D Structure of DNA Methyltransferase 1 Enzyme (DNMT1) 4WXX and (b) The active site of DNA Methyltransferase 1 Enzyme (DNMT1) 4WXX.

The 3D structure visualization of the enzyme should be accomplished to investigate the active site of the enzyme. The active site of the enzyme was determined from the literature or software such as MOE through the Site Finder menu. The active site is a location where the enzyme catalytic activity will decrease when there is an interaction between enzymes and inhibitors.

3.3. Molecular Docking Simulation
The first step is the virtual screening. At this step, the ligand moves to find out the conformation that matches to the protein binding site. The lowest value of bond energy values from the three standards at this step is the standard ligand S-Adenosyl-L-homocysteine (SAH) with the ΔGbinding value of -11.762 kcal/mol. About 6,227 compounds were successfully selected that will be continued to the next step.
The next step was rigid docking with retain 1, 30 and 100, respectively. The results of a decent molecular docking simulation were indicated by the Root Mean Square Deviation (RMSD) value lower than 2 Å [13]. The RMSD value states the ligand conformation difference with protein conformation. A total of 792 ligands were successfully selected through molecular docking simulations with the retain value of 1, 30, and 100 based on the RMSD value and Gibbs free bond energy value. The last step was the flexible docking. In this step, ligand and protein were both moving to find the best conformation. Flexible docking worked following the induced fit principle in enzyme work where the binding pocket of the protein will adjust its ligand shape to produce the best conformation. At this step, only 100 of the best ligands out of 792 selected ligands were used in the flexible docking step. The retain value of 100 was selected. The final result of flexible docking collected 72 ligands that have the lower value of ΔGbinding than the standard. Table 1 shows the Gibbs bond free energy value, RMSD value and pKi value from the database ligand.

### Table 1. ΔGbinding, pKi and rmsd value of 10 best ligands.

| Ligand | ΔGbinding (kcal/mol) | pKi   | RMSD (Å) |
|--------|----------------------|-------|----------|
| 4      | -17.1918             | 12.5240 | 1.2387  |
| 7      | -17.1918             | 12.5240 | 0.8469  |
| 9      | -16.2026             | 11.8033 | 1.6698  |
| 10     | -16.0662             | 11.7040 | 1.5867  |
| 12     | -15.8187             | 11.5237 | 1.8879  |
| 13     | -15.8084             | 11.5162 | 1.8715  |
| 14     | -15.7544             | 11.4768 | 1.2647  |
| 15     | -15.6963             | 11.4345 | 1.5990  |
| 16     | -15.6103             | 11.3719 | 0.8325  |
| 19     | -15.5655             | 11.3392 | 1.7339  |
| SAH*   | -11.7621             | 8.5685  | 2.1654  |

Note: *) = standard

3.4. Pharmacological Properties Analyses

After performing the molecular docking simulation, ten best ligands were selected based on the visualization of the best enzyme-ligand complex interactions and the lowest Gibbs free energy binding value. From the result, the three best ligands were chosen and then analysed for their pharmacological properties and toxicity using various software. The characteristics of the best physicochemical properties of three ligands is presented in table 2.

### Table 2. Physicochemical characteristics of three best ligands.

| Ligand | Molecular mass | H-Asp | H-Don | TPSA | cLogP | Druglikeness | Rotatable Bond |
|--------|----------------|-------|-------|------|-------|--------------|----------------|
| 12     | 624.59         | 15    | 7     | 223.29 | -0.5383 | 2.0396       | 8              |
| 14     | 610.56         | 15    | 8     | 234.29 | -0.8140 | 0.8012       | 7              |
| 16     | 610.56         | 15    | 8     | 234.29 | -0.8140 | 0.8012       | 7              |
| SAH*   | 384.82         | 11    | 4     | 212.38 | -3.7275 | -0.4433      | 7              |

Note: *) = standard

Benigni Bossa carcinogenicity-mutagenicity testing was accomplished using Toxtree 2.6 offline software. Carcinogenicity properties of Benigni Bossa mutagenicity were shown in table 3.
Table 3. Carcinogenicity-mutagenicity test of three best ligands by Toxtree 2.6.

| Ligand | Structural Alert for Genotoxic Carcinogenicity | Structural Alert for Nongenotoxic Carcinogenicity | Potential S. typhimurium TA100 Mutagen | Potential Carcinogen based on QSAR |
|--------|-----------------------------------------------|-----------------------------------------------|--------------------------------------|----------------------------------|
| 12     | No                                            | No                                            | No                                   | No                               |
| 14     | No                                            | No                                            | No                                   | No                               |
| 16     | No                                            | No                                            | No                                   | No                               |
| SAH*   | Yes                                           | No                                            | No                                   | No                               |

Note: *) = standard

Bioactivity test of natural products ligand was done using the Molinspiration online software. Molinspiration was used to determine the characteristics and bioactivity of ligand compounds of natural products as the best drug candidates. The parameters presented in the software is the result of comparing the similarity of the ligand results from the experiments with ligands with high bioactivity properties, such as G protein-coupled receptor (GPCR) ligands, modulator channels, kinase inhibitors, protease inhibitors, and enzyme inhibitors in the experimental database [14]. The three best ligand test results using Molinspiration online software can be seen in the table 4.

Table 4. Bioactivity test by Molinspiration software.

| Ligand | Ion Channel Modulator | Nuclear Receptor Ligand | Enzyme Inhibitor |
|--------|-----------------------|-------------------------|------------------|
| 12     | -0,70                 | -0,29                   | -0,03            |
| 14     | -0,62                 | -0,16                   | 0,08             |
| 16     | -0,60                 | -0,12                   | 0,09             |
| SAH*   | 0,29                  | -1,25                   | 0,99             |

Note: *) = standard

The next test was admetSAR online software. The principle of this software is to compare the ligands tested with the standards in the database. The parameters tested included AMES toxicity test, biodegradation test, and drug metabolism analysis. Ames Test is a test conducted to determine the carcinogenic potential of a compound using Salmonella typhimurium bacteria [15, 16].

Metabolic analysis with admetSAR software was carried out to determine whether the ligand tested could be a substrate/inhibitor of the cytochrome P450 enzyme (CYP P450). As a result the Gibbs free energy value increases and the interaction on the active site become less optimal, so the enzyme activity decreases [17]. The results of CYP isoform substrate analysis with admetSAR can be seen in table 5.

Table 5. Isoform CYP substrate analysis.

| Ligand | CYP450 2C9 Substrate | CYP450 2D6 Substrate | CYP450 3A4 Substrate |
|--------|----------------------|----------------------|----------------------|
| 12     | No                   | No                   | Yes                  |
| 14     | No                   | No                   | Yes                  |
| 16     | No                   | No                   | Yes                  |
| SAH*   | No                   | No                   | No                   |

Note: *) = standard

The prediction of ADME ligand characteristics was performed by employing the SWISSADME online software. The test included PAINS test, Brenk test, Veber test, and Egan test. PAINS test and
Brenk test were conducted to find out if there were ligand fragments that could provide a false biological response. Veber and Egan tests were performed to determine the oral bioavailability of a ligand [18]. The results of the characteristics analysis of ADME ligand with SWISSADME can be seen in following table 6.

**Table 6. ADME test of three best ligands by SWISSADME**

| Ligand | PAINS | Brenk | Veber | Egan |
|--------|-------|-------|-------|------|
| 12     | 0     | 0     | Yes   | Yes  |
| 14     | 0     | 0     | Yes   | Yes  |
| 16     | 0     | 0     | Yes   | Yes  |
| SAH*   | 0     | 0     | Yes   | Yes  |

Note: *) = standard

Based on the results, three best ligands had been selected which lower Gibbs free energy binding and best pharmacological properties. The three selected ligands are ligan 12, 14, and 16. The visualization and the structure of three best ligands can be seen in Fig. 2. Based on the structure from those three selected ligands, it could be concluded that the selected ligands comes from the phenolic group.

**Figure 2.** Structure and visualization 3 best ligands (a) 2S)-5-Hydroxy-2-(3-hydroxy-4-methoxyphenyl)-4-oxo-3,4-dihydro-2H-chromen-7-yl2-O-(6-deoxy-α-L-mannopyranosyl)-β-D-glucopyranoside (b) (2S)-2-(3,4-Dimethoxyphenyl)-5-hydroxy-4-oxo-3,4-dihydro-2H-chromen-7-yl-6-O-(6-deoxy-α-L-mannopyranosyl)-β-D-glucopyranoside, and (c) (2S)-5-Hydroxy-2-(3-hydroxy-5-methoxyphenyl)-4-oxo-3,4-dihydro-2H-chromen-7-yl 2-O-(6-deoxy-α-L-mannopyranosyl)-β-D-glucopyranoside.
4. Conclusions
Natural products had been virtually screened through a molecular docking simulation process resulting in the three best ligands which have lower ΔGbinding and better interaction with DNMT1 compared to SAH. The result also showed the best physicochemical characteristic and pharmacological properties. In this research, was found that the respective three best ligands which the best pharmacological properties were known to be derived from natural products of phenolic groups.

Acknowledgment
This research was supported by Chemistry Department University of Indonesia and was funded by Hibah Publikasi Internasional Terindeks untuk Tugas Akhir Mahasiswa (PITTA). We thanks to all members of Bioinformatics Laboratory FMIPA UI for the discussion during the course of this study.

References
[1] Esteller M and Herman J G 2002 Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours J. Pathol. 196 1 1-7
[2] Siegel R L, Miller K D and Jemal A 2017 Cancer statistics, 2017 CA Cancer J. Clin. 67 1 7-30
[3] Stopec A T, Brown-Glaberman U, Wong H Y, Park B H, Barnato S E, Gradishar W J, Hudis C A and Rugo H S 2012 The role of targeted therapy and biomarkers in breast cancer treatment Clin. Exp. Metastasis 29 7 807-19
[4] Inbar-Feigenberg M, Choufani S, Butcher D T, Rosfman M and Weksberg R 2013 Basic concepts of epigenetics Fertil. Steril. 99 3 607-15
[5] Maldonado-Rojas W, Olivero-Verbel J and Marrero-Ponce Y 2015 Computational fishing of new DNA methyltransferase inhibitors from natural products J. Mol. Graph. Model. 60 43-54
[6] Fahy J, Jeltsch A and Arimondo P B 2012 DNA methyltransferase inhibitors in cancer: a chemical and therapeutic patent overview and selected clinical studies Expert. Opin. Ther. Pat. 22 12 1427-42
[7] Shin E, Lee Y and Koo J S 2016 Differential expression of the epigenetic methylation-related protein DNMT1 by breast cancer molecular subtype and stromal histology J. Transl. Med. 14 1 87
[8] Cragg G M, Newman D J and Snader K M 1997 Natural products in drug discovery and development J. Nat. Prod. 60 1 52-60
[9] Da Rocha A B, Lopes R M and Schwartsmann G 2001 Natural products in anticancer therapy Curr. Opin. Pharmacol. 1 4 364-9
[10] Nobili S, Lippi D, Witort E, Donnini M, Bausi L, Mini E and Capaccioli S 2009 Natural compounds for cancer treatment and prevention Pharmacol. Res. 59 6 365-78
[11] Kitchen D B, Decornez H, Furr J R and Bajorath J 2004 Docking and scoring in virtual screening for drug discovery: methods and applications Nat. Rev. Drug Discov. 3 11 935
[12] Hecht D and Fogel G B 2009 Computational intelligence methods for docking scores Curr. Comput. Aided Drug Des. 5 1 56-68
[13] Tambunan U S F, Fadilah F and Parikesit A 2010 Bioactive compounds screening from Zingiberaceae family as influenza a/swine flu virus neuraminidase inhibitor through docking approach Online J. Biol. Sci. 10 4 151-6
[14] Verma A 2012 Lead finding from Phyllanthus debelis with hepatoprotective potentials Asian Pac. J. Trop. Biomed. 2 3 S1735-S7
[15] Cheng F, Li W, Zhou Y, Shen J, Wu Z, Liu G, Lee P W and Tang Y 2012 admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties J. Chem. Inf. Model. 52 11 3099-105
[16] Maron D M and Ames B N 1983 Revised methods for the Salmonella mutagenicity test Mutation Research/Environmental Mutagenesis and Related Subjects 113 3-4 173-215
[17] Lin J H and Lu A Y 1998 Inhibition and induction of cytochrome P450 and the clinical implications Clin. Pharmacokinet. 35 5 361-90
[18] Daina A, Michielin O and Zoete V 2017 SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules Sci. Rep. 7 42717