Pharmacophore-based virtual screening and molecular docking simulation of terpenoid compounds as the inhibitor of sonic hedgehog protein for colorectal cancer therapy

Mega Maulina Ekawati1,a, Mochammad Arfin Fardiansyah Nasution1,b, Syafrida Siregar1,c, Ilmi Fadhilah Rizki1,d, Usman Sumo Friend Tambunan1,e

1 Faculty of Mathematics and Natural Sciences, Universitas Indonesia, Depok, West Java, Indonesia
Email: a megamaulinaekawati19@gmail.com; b marfin.f@sci.ui.ac.id; c syafrida.siregar@yahoo.com; d ilmi.fadhilah@sci.ui.ac.id; e usman@ui.ac.id

Abstract. Colorectal cancer remains as the global health burden, which accounts for roughly 1-2 million new cases and 600,000 deaths per year. Hedgehog (Hh) signalling pathway has an imperative role in the mechanism and formation of colorectal cancer. Sonic hedgehog (Shh) protein is the most studied Hh protein because it is expressed by several tissues and experiments with Shh protein are generally applicable to other Hh homologs. In the present study, about 56,336 terpenoid compounds were screened through various computational methods using pharmacophore-based virtual screening and molecular docking simulation to determine their inhibitory potency against Shh protein. From molecular docking simulation results, about ten ligands have been selected according to their Gibbs free binding energies (ΔG binding) and the molecular interactions that formed during the formation of the terpenoid compound-Shh complex. Three terpenoid compounds, namely arganine J, asiaticoside A, and clinoposide A, shown a very high binding affinity toward Shh protein due to their lower ΔG binding than robotnikinin, the standard ligand. Moreover, ADME-Tox, bioactivity, bioavailability, and pharmacology test results revealed that these compounds have better biological and pharmacological activity than the other terpenoid compounds. For further research, these terpenoid compounds can be used as a drug candidate for colorectal cancer therapy.

Keywords: Colorectal Cancer, Terpenoid, Hedgehog Signaling Pathway, Sonic Hedgehog (Shh) Protein, Molecular Docking.

1. Introduction
Cancer is a disease that occurs by the mutation in a gene that changes cell function. Cancer can cause vital gene dysfunction; this disorder is effective in the cell cycle and causes abnormal proliferation [1]. One of the most dangerous types of cancer is colorectal cancer. Every year, more than 1 to 2 million patients were diagnosed with colorectal cancer, and more than 600,000 people die because of this cancer. Colorectal cancer occurs when cells in the lining of the large intestine or rectum change and grow irrespressibly, forming a tumour. This cancer begins with the growth of polyps (adenomas) which can develop into colorectal cancer.

Nowadays, there are various methods developed for the treatment of colorectal cancer, one of which is by inhibiting the Hedgehog signalling pathway (Hh). In recent years, it has been known that the Hh pathway had a crucial role in tumorigenesis when the adult tissues were reactivated. In mammals, the Hh signalling pathway initiates binding to one of the following proteins: Sonic Hedgehog (Shh),...
Hedgehog Indian (Ihh) and Desert Hedgehog (Dhh). Shh is the most studied protein because Shh is expressed by various tissues and experiments with Shh protein, in general, can be applied with other Hh homologs [2]. However, the Hh signalling pathway shows a multifunctional formation in colorectal cancer. In addition, signalling pathways with Shh can also stimulate the function of the process of angiogenesis, cell proliferation, and metastasis [3].

Terpenoid compounds are natural compounds that have a variety of biological functions such as antimicrobial, anti-cancer, and anti-inflammatory. Thus, terpenoid group compounds can be used as drug candidates that can inhibit the Hedgehog signalling pathway.

In this study, the screening of terpenoid compounds has been carried out by the pharmacophore-based virtual screening and molecular docking simulation. We used pharmacophore-based virtual screening because they represent chemical features complimentary to the receptor in 3-dimensional space and molecular docking method was used to identify and optimize drug candidates by examining and modelling molecular interaction between ligands and target macromolecules [4]. Absorption, Distribution, Metabolism, and Excretion – Toxicity (ADMET) tests also have been carried out to found new targets compounds that have higher anti-cancer potential and pharmacological character testing. Simulation results and predictions will be compared with other tested Shh inhibitors. Thus, through this research, these terpenoid natural compounds can be used as inhibitors for Shh protein in the treatment of colorectal cancer treatment.

2. Research Methodology
The research methodology in this study was conducted according to the previous research with slight addition in the pharmacophore-based virtual screening phase [5, 6].

2.1. Preparation Standard Ligand and Terpenoid Compounds
Robotnikinin compounds which used as the standard ligands were obtained from the ChemSpider database (http://www.chemspider.com). The three-dimensional structure of these compounds was saved in .sdf format and then converted to .mdb to be identified by Molecular Operating Environment (MOE) 2014.09. Then, there were 56,336 terpenoid compounds have been obtained from the PubChem database. The ligand must be screened using DataWarrior v.4.7.2 software to eliminate the unwanted ligands with any toxicity properties (e.g., mutagenic, tumorigenic, irritant, and reproductive effect). The standard ligand and terpenoid compounds were optimized using MOE 2014.09 software with MMFF94x force field and Root Mean Square (RMS) gradient value of 0.001 kcal/mol Å.

2.2. Preparation of Sonic Hedgehog Protein
The three-dimensional Shh protein structure was obtained from the RSCB PDB database (PDB ID: 3HO5). The protein structure was saved in .pdb format. Then, Shh protein was optimized with Amber10: EHT force field in Gas Phase solvation. All solvent molecules (H2O), chain A, and chain B from protein sequences have been removed using MOE 2014.09 software. After the optimization stage, Shh protein structure was saved in .moe format.

2.3. Pharmacophore Generation and Validation
In this research, Shh protein was docked with robotnikinin compound using ‘Induced Fit’ protocol and Amber10: EHT force field. The active side of the Shh protein was determined by the Site Finder command. After completing the docking process of the Shh-robotnikinin protein, one ligand that has the lowest RMSD value has selected to the generation pharmacophore stage. Generating pharmacophore can be found using Pharmacophore Query Editor. The pharmacophore features that have been completed in the generation phase must be validated by using the Pharmacophore Editor menu. The compounds must be fit on the pharmacophore features as hit molecules.
2.4. Pharmacophore-based Molecular Docking Simulation
The molecular docking simulation based on pharmacophore using terpenoid compounds that have been validated in the pharmacophore hit molecule with Shh protein. The docking simulation was using Amber10: EHT force field, R-Field solvation, and the placement must be Pharmacophore. Molecular docking simulations were divided into three parts, which are virtual screening, rigid receptor, and induced fit protocol. The virtual screening docking was performed with 30 retains. Then, continue using rigid docking simulation with 30, and 100 retains. After that, continue using induced fit simulation with 100, and 300 retains. After all the simulations were carried out, the best terpenoid ligands would be chosen which had the lowest Gibbs free energy binding and the best ligand conformation.

2.5. Analysis of Docking Results
Analysis of docking results was done with the value of energy binding. Drug scan analysis was also done using some software, such as Toxtree v2.6.6, and SwissADME. After the analysis has been carried out, the best compound would be chosen as the drug candidate for colorectal cancer therapy.

3. Result and Discussions

3.1. Preparation Standard Ligand and Terpenoid Compounds
The standard ligand of this research is robotnikinin, and terpenoid compounds were used to inhibit Shh protein in Hedgehog signaling pathway. Optimization of terpenoid ligands and standard ligands were performed in the MOE database viewer. The initial stage of optimization was performed wash command. The purpose of wash is to apply to each molecule a set of cleaning rules such as removing extraneous salts or adjusting protonation states, in order to ensure that each structure is in a form suitable for subsequent modelling steps such as conformation enumeration and protein-ligand docking. After that, MMFF94x force field was used because it is suitable for small organic molecules. Then, Energy minimization was used for molecule building, determining low energy conformations, conformational search, and preparation for molecular dynamics simulations. The RMS gradient is determined at 0.001 kcal/molÅ, which means energy minimization was done with the RMS gradient value reaching that value. Preserve Existing Chirality was activated to maintain initial chirality before minimizing. If the minimization stage has been completed, the ligands ready to use for molecular docking simulation.

3.2. Preparation of Sonic Hedgehog Protein
Shh protein structure was obtained from the RCSB PDB database with PDB ID: 3HO5. Shh protein structure has three chains namely chain A and chain B which represent the Hedgehog-interacting protein (HHIP) and the H chain which represent the Shh protein. In addition, Shh protein also has two unique ligands, chain A, and chain H which have interactions with Zn2+ ions and the H chain which has an interaction with Ca2+ ions.

The Shh protein optimization stage was begun by choosing Amber10: EHT Gas Phase solvation. Forcefield Amber10: EHT was chosen because it is more validated for proteins and nucleic acids than Amber12: EHT force field. Then the Fix Hydrogen button was selected to add or remove hydrogen according to the specifications of the force field. The Fix Charge button was also selected to recalculate partial charges. In addition, there was a Sequence Editor (SEQ) menu that can be used to manage residues, chains, and group tags. Then, Shh protein has three main chains, namely A, B and H. But only the H chain was used, because this chain represents the Shh protein and also has interactions with typical ligands, so the A and B chains must be eliminated by the SEQ menu.

The gradient was used 0.05 RMS kcal / molÅ, which means energy minimization will be stopped when the RMS value gradient reaches the specified value. After all these steps have been carried out, Shh protein was ready for molecular docking simulation.
3.3. Pharmacophore Generation

Pharmacophore is a feature needed to ensure optimal supramolecular interactions with specific biological targets and to trigger their biological responses [7]. The generating pharmacophore was done by docking Shh protein with robotnikinin compounds using the Induced Fit and Amber10: EHT force field. The active side of Shh protein was also determined by the Site Finder command to determine the possibility of an active site in the receptor from the three-dimensional structure of atomic coordinates. From the ‘Site Finder’ command, there are 13 amino acid residues and one Zn\(^{2+}\) cofactor which binds to the active side of the Shh protein. After completing the docking process of the Shhrobotnikinin protein, one ligand that has the lowest RMSD value was selected to the generation pharmacophore stage. The lowest ligand with RMSD was chosen because the ligand is the most stable ligand among the others.

From the results of generating pharmacophore using the interaction between Shh protein and robotnikinin, there were three hydrophobic groups, one hydrogen bond donor and one hydrogen bond acceptor of pharmacophore features (Fig. 1)

![Figure 1. Pharmacophore features of Shh Protein-Terpenoid Ligand Complex.](image)

The pharmacophore features must be validated by using the Pharmacophore Editor menu. Pharmacophore validation was used to test whether the pharmacophore features is good enough to predict active compounds [8]. If a molecule can be placed inside a sphere that represents a query feature, it is considered a hit molecule [7]. In this research, robotnikinin compound was validated in the molecule that hits the pharmacophore features. Whereas in terpenoid compounds, from 24,125 compounds, only 663 compounds that include as molecules hit.

3.4. Pharmacophore-based Molecular Docking Simulation

The molecular docking simulation based on pharmacophore using terpenoid compounds that have been validated with Shh protein. In this docking simulation, Amber10: EHT was used to be the force field, and the selected solvation was R-Field. The placement must be Pharmacophore. The selection of the Placement method is important because it will result in a pose of the desired ligand conformation.

Molecular docking simulations were divided into three parts, which are virtual screening, rigid receptor, and induced fit protocol. In molecular docking using virtual screening protocols, there was no duplication with 30 retains. Retain is the number of poses of interaction between ligand and protein. Molecular docking virtual screening resulted in 2,510 compounds of 24,660 terpenoid ligand...
compounds. Then, molecular docking simulations were carried out using the rigid receptor protocol. The purpose of this protocol is to improve the prediction accuracy of poses. Rigid receptor refers to proteins and ligands that are fixed so that either the bond angle or the length of the bond cannot be changed [9]. After duplication with 30 and 100 retains, there were 40 ligand compounds were selected. Then, molecular docking simulations were performed using an induced fit protocol. The Induced Fit protocol allows ligands to be able to move on their conformations [9]. After duplication with 100 and 300 retains, there were 35 ligand compounds were selected. Of the 35 ligand compounds, ten ligand compounds were chosen which had the lowest Gibbs free energy binding ($\Delta G_{\text{binding}}$) and the best ligand conformation than the other ligands.

### 3.5. Analysis of Molecular Docking

Molecular docking simulations give the $\Delta G_{\text{binding}}$ value of the ligand-protein complex formed throughout the simulations. $\Delta G_{\text{binding}}$ is a function of intermolecular, electrostatic hydrogen bonds and Van der Waals interactions between ligands and protein pose [10]. The $\Delta G_{\text{binding}}$ value of standard and ten best ligands are shown in **Table 1**.

**Table 1. $\Delta G_{\text{binding}}$ value of standard ligand and ten best ligands.**

| Ligand                              | $\Delta G_{\text{binding}}$ (kcal/mol) |
|-------------------------------------|----------------------------------------|
| Terp24334 (Pachymic Acid)           | -13.3748                                |
| Terp3615 (KVAQLXUMUVEKGRUHFFFAOYSA-N) | -13.2817                               |
| Terp10703 (Arganine J)              | -13.0644                                |
| Terp11927                           | -12.9309                                |
| Terp13617                           | -12.9302                                |
| Terp4531                            | -12.8481                                |
| Terp5657 (IC9564, Betulinic Acid Derivates) | -12.1991                                |
| Terp1396                            | -12.1309                                |
| Terp1274 (Asiaticoside A)           | -11.9764                                |
| Terp13194 (Clinopodiside A)         | -11.1319                                |
| Robotnikinin                        | -10.9158                                |

### 3.6. Drug Scan Analysis

#### 3.6.1. Carcinogenicity and Mutagenicity Prediction Test.

Carcinogenicity and mutagenicity prediction of ligands was carried out using Toxtree v.2.6.13 software based on Benigni-Bossa rules that used to predict carcinogenicity and mutagenicity of ligands based on their structure. The main target is to find molecular functional or substructure groups that are known to be related to carcinogenic activity of chemicals, because if one or more functional groups are identified, the system will mark the carcinogenicity potential of the chemical [9]. The parameters to predict carcinogenicity and mutagenicity are genotoxic and non-genotoxic carcinogenicity, potential carcinogenicity based on QSAR and mutagenicity potential against *Salmonella typhimurium* bacteria. The bacteria were chosen because each strain in *Salmonella typhimurium* is especially sensitive to chemical carcinogens. Then, the other parameters like genotoxic carcinogenicity can cause direct damage to DNA, because many types of mutagen are recognized in this category, and non-genotoxic carcinogenicity parameters can cause indirect damage to DNA [11]. The test results can be seen in **Table 2**.
Table 2. The Carcinogenicity and mutagenicity prediction test by Toxtree v.2.6.13.

| Ligand       | Genotoxic Carcinogenicity | Non-genotoxic Carcinogenicity | Potential Mutagenicity of S. typhimurium | Potential Carcinogenicity Based on QSAR |
|--------------|---------------------------|-------------------------------|------------------------------------------|----------------------------------------|
| Terp24334    | Negative                  | Negative                      | Negative                                  | Negative                               |
| Terp3615     | Negative                  | Negative                      | Negative                                  | Negative                               |
| Terp10703    | Negative                  | Negative                      | Negative                                  | Negative                               |
| Terp11927    | Negative                  | Negative                      | Negative                                  | Negative                               |
| Terp13617    | Negative                  | Negative                      | Negative                                  | Negative                               |
| Terp4531     | Negative                  | Negative                      | Negative                                  | Negative                               |
| Terp5657     | Negative                  | Negative                      | Negative                                  | Negative                               |
| Terp1396     | Negative                  | Negative                      | Negative                                  | Negative                               |
| Terp1274     | Negative                  | Negative                      | Negative                                  | Negative                               |
| Terp13194    | Negative                  | Negative                      | Negative                                  | Negative                               |
| Robotnikinin | Negative                  | Positive                      | Negative                                  | Negative                               |

Based on the test results, all ligands have a negative result on the parameters. This indicates that the candidate ligands do not have carcinogenicity and mutagenicity properties. However, the standard robotnikinin compound has positive results for non-genotoxic carcinogenicity parameter. It might be due to the presence of aryl halide fragments in these compounds which have the potential to be carcinogenic compounds, in accordance with Benigni-Bossa rules.

3.6.2. Pharmacokinetic Properties Prediction. Drug development involves the process of adsorption, distribution, metabolism, and excretion (ADME). These properties can be predicted by using the SwissADME software [10]. The test results can be seen in Table 3.

Table 3. The pharmacokinetic properties prediction by SwissADME software.

| Ligand       | GI Absorption | Cytochrome Inhibitor | Bioavailability |
|--------------|---------------|----------------------|-----------------|
|              |               | 1* 2* 3* 4* 5*       |                 |
| Terp24334    | High          | No No No No Yes      | 0.56            |
| Terp3615     | High          | No No Yes No Yes     | 0.56            |
| Terp10703    | Low           | No No No No Yes      | 0.17            |
| Terp11927    | High          | No No No No Yes      | 0.56            |
| Terp13617    | High          | No No No No Yes      | 0.56            |
| Terp4531     | Low           | No No No No Yes      | 0.56            |
| Terp5657     | Low           | No No No No Yes      | 0.56            |
| Terp1396     | Low           | No No No No Yes      | 0.11            |
| Terp1274     | Low           | No No No No Yes      | 0.17            |
| Terp13194    | Low           | No No No No Yes      | 0.17            |
| Robotnikinin | High          | No No Yes No Yes     | 0.55            |

*1 = CYP1A2; 2 = CYP2C19; 3 = CYP2C9; 4 = CYP2D6; 5 = CYP3A4.
From the results of the tests in the parameters of gastrointestinal absorption, Terp24334, Terp3615, Terp11927, Terp13617, and robotnikinin ligands have high results, which showed that the ligands are suitable for oral use. Whereas ligands Terp10703, Terp4531, Terp5657, Terp1396, Terp1274, and Terp13194 have low results, which showed that these ligands are not well used orally, but it still can be used through other metabolic pathways such as injection.

In the cytochrome inhibitor parameter, the results of yes will show that the ligand has the potential as an inhibitor in the process of cytochrome metabolism, which can lead to toxic formation. While the results of no will indicate that the ligand does not have the potential as an inhibitor in the process of cytochrome metabolism. From the test results, there were only three ligands which were not cytochrome inhibitors, Terp10703, Terp1274, and Terp13194.

The other parameter, such as bioavailability values are used to predict the compound having at least 10% oral bioavailability [12]. The higher the bioavailability value of a drug, the better the drug can be used orally. From the test results, four ligands have low bioavailability values. These are Terp10703, Terp1396, Terp1274 and Terp13194.

4. Conclusion

In this research, pharmacophore-based virtual screening and molecular docking simulations have been carried out using terpenoid natural compounds to inhibit Shh protein. About 24,660 terpenoid ligands were carried out molecular docking simulations, and ten ligands were chosen which had lower binding free energy than the robotnikinin. The ten ligands were tested for mutagenicity/carcinogenicity, ADMETox properties, and pharmacokinetic properties. This research obtained the best three terpenoid compounds that have better drug characteristics compared to other ligands; there were Terp10703 (arganine J), Terp1274 (asiaticoside A), and Terp13194 (clinoposide A). Thus, for further research, the in vitro and in vivo tests should be performed to know the inhibition potential of Shh protein in the Hedgehog signalling pathway in the treatment of colorectal cancer.

Acknowledgement

This research is financially supported by the Directorate of Research and Community Engagement of Universitas Indonesia (DRPM UI) through Hibah Publikasi Internasional Terindeks Untuk Tugas Akhir Mahasiswa (PITTA) UI No: 2327/UN2.R3.1/HKP.05.00/2018. There is no conflict of interest stated in the manuscript.

References

[1] Hassanpour S H and Dehghani M 2017 Review of cancer from perspective of molecular J. Cancer Res. Pract. 4 4 127-9
[2] Fu X, Yang X, Li J, Tian X, Cai J and Zhang Y 2010 Opposite expression patterns of Sonic hedgehog and Indian hedgehog are associated with aberrant methylation status of their promoters in colorectal cancers Pathol. 42 6 553-9
[3] Wu C, Zhu X, Liu W, Ruan T and Tao K 2017 Hedgehog signaling pathway in colorectal cancer: function, mechanism, and therapy Oncotargets Ther. 10 3249
[4] Kapetanovic I 2008 Computer-aided drug discovery and development (CADD): in silico-chemico-biological approach Chem. Biol. Interact. 171 2 165-76
[5] Tambunan U S F, Nasution M A F, Azhima F, Parikesit A A, Toepak E P, Idrus S and Kerami D 2017 Modification of S-Adenosyl-L-Homocysteine as Inhibitor of Nonstructural Protein 5 Methyltransferase Dengue Virus Through Molecular Docking and Molecular Dynamics Simulation Drug Target Insights 11 1-14
[6] Tambunan U and Nasution M 2017 Identification of novel Ebola virus (EBOV) VP24 inhibitor from Indonesian natural products through in silico drug design approach AIP Conf. Proc. 1862 1 030091
[7] Qing X, Lee X Y, De Raeymaecker J, Tame J R, Zhang K Y, De Maeyer M and Voet A 2014 Pharmacophore modeling: advances, limitations, and current utility in drug discovery Journal ...
of Receptor, Ligand and Channel Research 7 81-92

[8] Fei J, Zhou L, Liu T and Tang X-Y 2013 Pharmacophore modeling, virtual screening, and molecular docking studies for discovery of novel Akt2 inhibitors Int. J. Med. Sci. 10 3 265

[9] Chen Y-C 2015 Beware of docking! Trends Pharmacol. Sci. 36 2 78-95

[10] Hecht D and Fogel G B 2009 Computational intelligence methods for docking scores Curr. Comput. Aided Drug Des. 5 1 56-68

[11] Benignia R, Bossaa C, Jeliazkovab N, Netzevac T and Worth A 2008 The Benigni/Bossa rulebase for mutagenicity and carcinogenicity—a module of Toxtree. In: JRC Scientific and Technical Reports: European Commission, Joint Research Centre, Institute for Health and Consumer Protection) pp 1-78

[12] 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