Discovery of novel drug candidates based on herbaric acid derivates as potential inhibitors of the hedgehog signaling pathway in cervical cancer therapeutics

M A F Nasution, A A Parikesit and U S F Tambunan

Bioinformatics Research Group, Department of Chemistry, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Kampus UI Depok, Depok 16424, Indonesia

Corresponding author’s e-mail: usman@ui.ac.id

Abstract. Cervical cancer ranks as the second most deadly cancer in women worldwide and as the most deadly in developing countries. However, there is currently no effective treatment for this disease. Therefore, it is necessary to find improved drugs for cervical cancer treatment. Cervical cancer is caused by human papillomavirus (HPV) infection, which has E6 and E7 proteins that may activate the Hedgehog (Hh) signaling pathway and regulate the proliferation, survival, and migration of cervical cancer cells. In this study, a novel series of herbaric acid derivates were designed and developed as potential inhibitor candidates of the Sonic hedgehog (Shh) signaling pathway. All of the potential inhibitors were analyzed and compared with Shh inhibitors, such as robotnikin, through molecular docking simulations. Molecular docking simulations of 6310 ligands were performed using the Accelrys Discovery Studio 2.5 software according to the LibDock method. After the analysis of the ligand–Shh protein complex interaction in the docking simulation, it was found that Sd32, Sa32, and Wc34 ligands were best at inhibiting the Sonic hedgehog protein.

1. Introduction
Cancer, or malignant neoplasm, can be defined as a group of diseases characterized by genomic abnormalities, uncontrolled cell growth, and gene expression deviations [1]. One of the deadliest types of cancer is cervical cancer, which occurs specifically in the cervical area, i.e., the cavity-shaped area between the vagina and the uterus (commonly known as the cervix). This cancer is caused by human papillomavirus (HPV) infection. The presence of HPV is detected in 99.7% of cervical cancer cases, of which 70% are caused by HPV infection types 16 and 18 [2–4]. Cervical cancer is ranked second as the most lethal cancer among women worldwide (most lethal in developing countries), with at least 500,000 new cases occurring each year and a mortality rate reaching one-third [5,6]. It is estimated that cervical cancer has become the second most prevalent cancer in women in Indonesia after breast cancer, with 20,928 new cases of cervical cancer diagnosed each year [7]. There is currently no effective treatment for cervical cancer. It is known that HPV can cause mutations in stem cells, which eventually become cancer stem cells [8]. In general, stem cells can be defined as cells that have the ability to live eternally by constant self-renewal and the generation of mature cells from a network of differentiation stages. These stem cells are also responsible for the development of new tissue cells. In the event of stem cell mutation, cancer stem cells may be produced. Cancer stem cells can cause the development of tumor cells in tissues and trigger the development of cancer in the human body [9].

A variety of methods have been studied and developed to effectively treat cancer, one of which is inhibition of the signaling pathway of cancer stem cells. One of the pathways involved in cancer stem cell development is the Hedgehog signaling pathway (Hh). This signaling pathway is crucial in embryonic development and differentiation, as well as in stem cell renewal [10]. Abnormal Hh activity is commonly associated with the development of prostate, lung, pancreas, stomach, and cervix cancers [5,11]. The Hh signaling pathway can affect the proliferation, survival, and migration of cancer stem
cells; therefore, intervention of this pathway is an option for the effective treatment of cervical cancer. The Hh signaling pathways consist of three proteins (Sonic Hedgehog (Shh), Indian Hedgehog (Ihh), and Desert Hedgehog (Dhh)), which play a crucial role in the early development of stem cells. The Shh protein is the most extensively studied protein in the Hh signaling pathway, particularly because of its importance in cell division from adult stem cells and involvement in the development of several types of cancer [12,13].

Our previous work has found that herbaric acid (IUPAC name: 2-(4,6-dihydroxy-3-oxo-2,3-dihydro-1H-inden-1-yl)acetamide) may be a potential lead compound for cervical cancer therapy. It can bind Zn$^+$ at the histone deacetylase (HDAC) protein [14]. However, this compound has a low oral bioavailability and so requires binding site modification to overcome this weakness, as well as to increase its binding affinity to several receptors. Using this information, we designed a novel inhibitor based on modification of herbaric acid and observed its interactions with the Shh protein through molecular docking simulations. Through this research, we obtained the structure of a herbaric acid-based compound that targets the Shh protein and could be used as a drug candidate for cervical cancer therapy.

2. Materials and methods
The research methodology used in this study was developed from methods used in our previous research [14–16].

2.1. Determination and preparation of the Shh protein
The three-dimensional (3D) complex crystal structure of ShhN and Hedgehog-interacting protein (Hhip) Homo sapiens was downloaded from PDB ID: 3HO5. It was observed at a resolution of 3.01 Å and contains one Zn$^+$ ion and two Ca$^+$ ions on the active side of Shh [17,18]. This protein structure was saved in the .pdb format. The next step was removal of all the solvent molecules (H$_2$O), Hhip proteins, and any ligands it contained using Accelrys Discovery Studio 2.5 software. The Shh protein was prepared using the Prepare Protein feature followed by minimization of the protein structure using CHARMM force field and partial charge MMFF94. Minimization was performed using the Smart Minimizer algorithm within 1000 steps and with an RMS Gradient of 0.001. The Shh protein structure was saved in the .pdb or .dsv formats.

2.2. Design and preparation of the ligands
In this study, all ligand inhibitors were drawn using the ChemDraw Ultra 12.0 software. Herbaric acid was chosen as the main backbone of the ligand inhibitors. The carboxylic group (–COOH) in the herbaric acid compound was replaced by the amide group (–CONHR). The structures of the drawn compounds were classified into several groups according to the different types of heterocyclic fragments (such as benzene, cyclohexane, cyclopentene, pyridine, pyrimidine, pyridazine, pyrazine, pyrrole, furan, thiophene, pyrazole, oxazole, thiazole, isoxazole, isothiazole, oxadiazole, triazole, indole, benzo[b]thiophene, benzofuran, and quinoline) directly attached to the nitrogen atom of the amide group. These ligands were gathered with coded alphabetic capital letters (A through W), followed by an alphabetical lowercase a, b, c, or d depending on the location of –X on the heterocyclic functional groups, and ended with the provision that defined the shape of the –X group (such as halide group, alkyl, hydroxyl, amine, ether, carbonyl, carboxylate, esters, amide, hydroxylamine, sulphonate, or boronic acid). Thus, 6310 ligands were created. The full list of the –R and –X groups on the ligand modifications can be seen in the Supplementary Materials provided in the Appendices. All of these ligands were then stored in the .mol format. After the entire molecular structure had been constructed, all of the structural files were prepared, minimized, and optimized using the Accelrys Discovery Studio 2.5 software.

2.3. Molecular docking simulations and visualization analysis of the docking results
In this research, molecular docking simulations were conducted using the LibDock module in the Accelrys Discovery Studio 2.5 software. After all of the ligands and the Shh protein had been prepared and minimized, the Shh ligand and protein were subjected to the docking process after the cavity site of the receptor had been determined using Site Sphere (with the Number of Hotspot and Docking Tolerance set to 100 and 0.20, respectively). Next, the Gibbs free binding energy ($\Delta$G_{bind}) of the best 200 ligands in this step were determined using the Calculate Binding Energies module. This was done using the Steepest Descent minimization parameter with a maximum of 1000 steps (on In-situ Ligand
Minimization parameters) and 500 (on Ligand Conformational Entropy parameters), and Generalized Born with Molecular Volume solvent on the Solvent Model Implicit option. Finally, Molecular Operating Environment (MOE) 2008.10 was used to predict the molecular interaction between the selected ligand at the binding site of the Shh protein [19,20].

3. Results and discussions

Determination of Shh protein sequences began by accessing the NCBI (National Biotechnology Information Center) website, which identified the Shh Homo sapiens protein with 462 amino acid residues as the chosen target protein. The next step was analysis of the protein amino acid residues using the Basic Local Alignment Search Tools (BLAST) algorithm for Proteins (BLASTP). This was followed by accessing of the NCBI website to determine the location of the conserved region or conserved domain of the Shh protein. From this analysis, it was found that the Shh protein consists of two main components: the Hedgehog amino-terminal signaling domain (residues 39–184) and the Hedgehog Domain N-terminal region (residues 197–318). The zinc ion (Zn\(^{2+}\)) and two calcium ions (Ca\(^{2+}\)), which act as cofactors for later Shh proteins interacting with Hhips in the Hedgehog signaling pathways, are located at the Hedgehog amino-terminal signaling domain. The Zn\(^{2+}\) ion serves as an interactor, which plays a significant role in activating the hedgehog signaling pathway by regulating the formation of the Shh–Hhip complex. The Ca\(^{2+}\) ions prevent any electrostatic repulsion between the two proteins [21]. Therefore, the designed inhibitor ligand should use the Zn\(^{2+}\) ions and the two Ca\(^{2+}\) ions as the inhibition target of the ligand–protein complex. An illustration of the catalytic site of the Shh protein is shown in figure 1.

Active site determination of the protein was conducted prior to the molecular docking simulations. In this study, the active site of the Shh protein was determined using Find Sites in the Receptor Cavities feature of the Accelrys Discovery Studio 2.5 software after the protein minimization step had been carried out. At this stage, the cavities having the closest location between the Zn\(^{2+}\) ion and the two Ca\(^{2+}\) ions were located since the metal ion plays a crucial role in activating the Hedgehog signaling path. The charge relay system on the active site of the Shh protein is a significant inhibition target for preventing Shh protein binding to Hhip and for activating the Hedgehog signaling pathway so that cancer stem cells are not able to perform proliferation because of the inexistence of the transcription process in this signaling pathway. This step can be used as a strategy to cure cancer, including cervical cancer.

In this research, herbaric acid derivate compounds were selected as the main compounds for inhibiting the Shh protein. These compounds serve as a bioisostere of herbaric acid, in which the carboxylic group (–COOH) in the herbaric acid compound is replaced by the amide group (–CONHR). This was done to improve the stability of the ligand body, as well as to increase the druglikeness of the ligand. An example of the drawn ligand in this research can be seen in figure 2.

Finally, molecular docking simulations were conducted to obtain ligand conformation and orientation to the catalytic site of an enzyme/protein. The simulations used molecular mechanical approaches to minimize ligand energy at the binding site of the receptor to obtain the most stable conformation. The enzyme was rigid, whereas the ligand had the flexibility to find the most stable

Figure 1. 3D illustration of the catalytic site of the Sonic Hedgehog (Shh) protein. The left and right figures were generated through UCSF Chimera 1.12 and Accelrys Discovery Studio 2.5 software, respectively.
Table 1. Molecular docking simulation results from the Accelrys Discovery Studio 2.5 software using the LibDock and Calculate Binding Energies modules.

| No. | Ligand | ΔG_{binding} (kcal/mol) | pKi  |
|-----|--------|-------------------------|------|
| 1   | Wc34   | -134,0623               | 97,7040 |
| 2   | Sa32   | -130,5167               | 95,1199 |
| 3   | Da32   | -122,9780               | 89,6258 |
| 4   | Wa35   | -92,0798                | 67,1073 |
| 5   | Bc32   | -76,3784                | 55,6642 |
| 6   | Pd18   | -73,9746                | 53,9123 |
| 7   | Vc32   | -70,9440                | 51,7036 |
| 8   | Wc30   | -69,6659                | 50,7721 |
| 9   | Wc27   | -67,4936                | 49,1890 |
| 10  | Ja35   | -65,9240                | 48,0451 |
| 11  | Vh31   | -64,2880                | 46,8528 |
| 12  | Vh35   | -61,6646                | 44,9408 |
| 13  | Bh35   | -61,1512                | 44,9048 |
| 14  | Sh18   | -60,2017                | 43,8747 |
| 15  | Sd32   | -59,0560                | 43,0397 |
| 16  | Ma32   | -57,6948                | 42,0476 |
| S1  | Hit 1  | -54,9098                | 40,0180 |
| S2  | Herbaric Acid | -52,8208 | 38,4955 |
| S3  | Hit 3  | -51,5675                | 37,5822 |
| S4  | Hit 2  | -33,4938                | 24,4104 |
| S5  | Robotnikinin | -32,8220 | 23,9205 |

After molecular docking simulations had been performed, molecular interaction analysis between the ligand inhibitor and Shh protein through hydrogen bonding, metal coordination bonds, and van der Waals interactions was conducted. The MOE 2008.10 software was used to determine the molecular interaction of the selected ligand–Shh protein complex. The best ligand from the previous simulation, the Wc34 ligand, was observed first to visualize its molecular interaction in the cavity site of the Shh protein; this is shown in figure 3. From the visualization, the interaction involves not only the Zn$^{2+}$ that binds the –OH group to the ligand but also the Ca$^{2+}$, which forms a covalent coordination bond with the nitrogen atom in the hydroxylamine group (–NHOH). Additionally, the hydrogen atom attached to the
oxygen atoms of the hydroxylamine group interacted with the residues Glu90 and Glu126, and the hydrogen atom attached to the nitrogen atom of the amide group interacted with the residue of His133. So the oxygen atoms of the $-\text{OH}$ group binding to Zn$^{2+}$, which also interacts with the residues His140, Asp147, and His182.

The interaction between the Shh protein and the Sa32 ligand, i.e., the ligand with the second lowest Gibbs free energy and the second highest inhibition, can be observed in figure 4. Although the interactions were quite similar to those of the Shh-Wc34 complex (the Zn$^{2+}$ binds perfectly with the oxygen atom in the herbaric acid backbone), the Ca$^{2+}$ ion interacts with the oxygen atoms in the hydroxyamide group ($-\text{CONHOH}$). It should be noted that the Sa32 ligand also interacts with Glu90, Glu126, Asp129, Asp131, His140, Asp147, and His182 residues. Overall, the Sa32 inhibitor ligand shows significant molecular interaction with amino acid residues around the catalytic site of the Shh protein.

Finally, the interaction between the Shh protein and the Sd32 inhibitor ligand was also visualized and analyzed (figure 5). From the visualization results of the Shh–Sd32 complex, it appears that metal interaction of the three cofactors (Zn$^{2+}$ atom and two Ca$^{2+}$ atoms) occurred. Zn$^{2+}$ binds via a covalent
Figure 5. 2D illustration of the Sd32 ligand and the binding site of the Shh protein

coordination bond with an oxygen atom in the amide group of the ligand, whereas the two Ca²⁺ cofactors interact with oxygen atoms on the hydroxyamide functional group of the Sd32 ligand.

Specifically, one Ca²⁺ atom interacts with sp² oxygen atoms in the carbonyl (C=O) group, while the other binds the sp³ oxygen atom in the –NHOH group. The Sd32 ligand also interacts with other amino acid residues in the Shh protein, such as Glu53, Glu90, Glu126, Asp129, Asp131, and Ser138.

4. Conclusions

The screening of potential drug candidates using bioinformatics-based computational methods has progressed into an efficient, inexpensive approach in the drug design and development process. In this study, 6310 herbaric acid-based ligands were drawn and simulated through molecular docking simulations and compared with the Shh protein to examine the binding affinity and interaction in its cavity site. After a series of docking simulations, we concluded that the Sd32, Sa32, and Wc34 ligands were best at inhibiting the Shh protein, showing the potential to be developed as lead drugs to inhibit Hh signaling pathways in cervical cancer therapy. This study should be extended to a wet experiment by synthesizing these ligands and testing them through bioactivity and inhibition assays. Additionally, computational ADME-Tox should be performed to predict any toxicity risks that may occur, followed by molecular dynamics simulations to analyze the stability of the complexes in the human body.

Acknowledgements

The authors would like to thank Universitas Indonesia for the research facilities, Enago (www.enago.com) for the English language review, and specifically are grateful to Qivi Azizah and Ahmad Husein Alkaff for constantly supporting the research through their critical reviews and feedbacks.

Appendices

The supplementary materials in this research can be accessed through the author's ResearchGate at https://www.researchgate.net/publication/319058903_Supplementary_materials_of_Discovery_of_Novel_Drug_Candidate_Based_on_Herbaric_Acid_Derivates_as_Potential_Inhibitors_of_the_Hedgehog_Signaling_Pathway_in_Cervical_Cancer_Therapeutics

References

[1] Desany B and Zhang Z 2004 Drug Discov. Today 9 795–802
[2] Ganguly N and Parihar S P 2009 J. Biosci. 34 113–23
[3] Division of STD Prevention US Department of Health and Human Services 1999 Prevention of Genital HPV Infection and Sequelae: Report of an External Consultants’ Meeting (Atlanta:
US Department of Health and Human Services) p 111

[4] Winer R L, Hughes J P, Feng Q, O’Reilly S, Kiviat N B, Holmes K K and Koutsky L A 2006 N. Engl. J. Med. 354 2645–54

[5] Samarzija I and Beard P 2012 Biochem. Biophys. Res. Commun. 425 64–9

[6] Yugawa T and Kiyono T 2009 Rev. Med. Virol. 19 97–113

[7] Badan Litbangkes Kementerian Kesehatan RI 2013 Riset Kesehatan Dasar (Jakarta:Kemenkes).

[8] Martens J E, Arends J, Van der Linden P J Q, De Boer B a G and Helmerhorst T J M 2004 Anticancer Res. 24 771–5

[9] Weissman I L, Anderson D J and Gage F 2001 Annu. Rev. Cell Dev. Biol. 17 387–403

[10] Shahi M H, Afzal M, Sinha S, Eberhart C G, Rey J A, Fan X and Castresana J S 2010 BMC Cancer 10 614

[11] Karhadkar S S, Bova G S, Abdallah N, Dhara S, Gardner D, Maitra A, Isaacs J T, Berman D M and Beachy P A 2004 Nature 431 707–12

[12] Peukert S and Miller-Moslin K 2010 Chem. Med. Chem. 5 500–12

[13] Kelleher F C 2011 Carcinogenesis 32 445–51

[14] Tambunan U S F, Parikesit A A, Nasution M A F, Hapsari A and Kerami D 2017 Iran. J. Pharm. Res. 16 1115–29

[15] Parikesit A A, Kinanty K and Tambunan U S F 2013 Pakistan J. Biol. Sci. 16 1836–48

[16] Tambunan U S F, Parikesit A A, Adam V C, Nasution M A F, Puspitasari R D and Kerami D 2017 Int. J. Geomat. 13 60–68

[17] Bosanac I, Maun H R, Scales S J, Wen X, Lingel A, Bazan J F, de Sauvage F J, Hymowitz S G and Lazarus R A 2009 Nat. Struct. Mol. Biol. 16 691–7

[18] Hwang S, Thangapandian S, Lee Y, Sakkiah S, John S and Lee K W 2011 J. Bioinform. Comput. Biol. 9 15–35

[19] Molecular Operating Environment (MOE) 2008.10

[20] Vilar S, Cozza G and Moro S 2008 Curr. Top. Med. Chem. 8 1555–72

[21] Bishop B, Aricescu A R, Harlos K, O’Callaghan C A, Jones E Y and Siebold C 2009 Nat. Struct. Mol. Biol. 16 698–703

[22] Halperin I, Ma B, Wolfson H and Nussinov R 2002 Struct. Funct. Genet. 443 409–43

[23] Kitchen D B, Decornez H, Furr J R and Bajorath J 2004 Nat. Rev. Drug Discov. 3 935–49

[24] Hwang S, Thangapandian S and Lee K W 2013 PLoS One 8 e68271